Hermann Winner
Stephan Hakuli
Felix Lotz
Christina Singer
Editors

Handbook of Driver Assistance Systems

Basic Information, Components and Systems for Active Safety and Comfort

Springer Reference

Handbook of Driver Assistance Systems

Hermann Winner • Stephan Hakuli
Felix Lotz • Christina Singer
Editors

Handbook of Driver Assistance Systems

Basic Information, Components and Systems for Active Safety and Comfort

With 737 Figures and 53 Tables

Springer Reference

Editors
Hermann Winner
Institute of Automotive Engineering
Technische Universität Darmstadt
Darmstadt, Germany

Stephan Hakuli
Continental Engineering Services GmbH
Frankfurt am Main, Germany

Felix Lotz
Continental AG
Frankfurt am Main, Germany

Christina Singer
Institute of Automotive Engineering
Technische Universität Darmstadt
Darmstadt, Germany

ISBN 978-3-319-12351-6 ISBN 978-3-319-12352-3 (eBook)
ISBN 978-3-319-12353-0 (print and electronic bundle)
DOI 10.1007/978-3-319-12352-3

Library of Congress Control Number: 2015955729

Printed on acid-free paper

This Springer imprint is published by SpringerNature
The registered company is Springer International Publishing AG Switzerland.

Preface

A Handbook of Driver Assistance Systems is challenged by the requirement to compile all relevant activities on advanced driver assistance systems (ADAS) into a comprehensive, understandable, and structured way. This book not only explains components, features, and well-known standard systems, but it aims at giving a complete picture, always focusing on the entire system. It is designed as a standard work for ADAS developers, researchers, and decision-makers. The first edition of this book was published in German language in 2009 and had the objective to close the gap of standard ADAS books available at that time. It was well received by the market, and in 2012 the second, corrected version was published. However, the field of ADAS is moving fast, so in 2015 the third edition – still in German language – was released. Its contents were completely revised, and numerous additional topics were added. This first edition in English language is the result of an increasing demand to make the German version of this book accessible for international readers.

The resulting scope of this handbook starts with the fundamentals of the development of ADAS, including human factors, ergonomics, and social and legal aspects. Simulation and virtual system integration gain importance in modern automotive development processes and are covered in this edition, together with established real-world-based processes for system verification and validation. Environment sensor systems play a key role in every system architecture. Therefore, current sensor principles and technologies are discussed in detail alongside the state-of-the-art actuators for steering and braking. Any ADAS with driver interaction demands an appropriate human-machine interface (HMI). Different multimodal HMI concepts are explained together with their individual requirements for a user-friendly design. ADAS are everywhere in today's passenger cars, commercial vehicles, motorcycles, tractors, and agricultural machinery. This book gives a comprehensive overview of state-of-the-art systems, including their functionality and particular requirements. Finally, the book closes with an outlook toward upcoming ADAS in research and development and concludes with the question "Quo vadis, ADAS?"

The editors of this book thank all the authors for their valuable contribution and a great collaboration. Thanks to the publisher for agreeing to create an international standard work on ADAS and especially to Daniela Graf and Andreas Maisch from

Springer for the reliable pleasant cooperation. Thanks to all lectors for ensuring high quality and a good reading experience. Thanks to all supporters, especially from the Institute of Automotive Engineering at Technische Universität Darmstadt, and to our student assistant Yannick Ryma for his reliable and accurate work. And last but not least, thank you for deciding to work with this book and providing us with your valuable feedback.

List of Companies and Universities

Companies

BMW Group	Guy Berg
	Edmund Donges (formerly)
	Simon Fürst
	Felix Klanner
	Martin Liebner
	Christian Ruhhammer
	Thomas Schaller
	Moritz Werling
Bosch Engineering GmbH	Matthias Mörbe
Bosch SoftTec GmbH	Andreas Engelsberg
	Thorsten Mlasko
Bundesanstalt für Straßenwesen	Tom Michael Gasser
	Andre Seeck
	Patrick Seiniger
Carl Zeiss AG	Christian Wojek
Continental AG	Alexander Bachmann
	Bernward Bayer
	Stefan Bunzel
	Axel Büse
	Stephan Cieler
	Oliver Fochler
	Steffen Gruber
	Stephan Hakuli
	Jens Hoffmann
	Andree Hohm
	Maximilian Höpfl
	Paul Linhoff
	Felix Lotz
	Stefan Lüke
	Mark Mages
	Roman Mannale
	Stefan Menzel
	Norbert Ocvirk
	Bernd Piller
	Martin Punke
	Thomas Raste
	James Remfrey

	Peter E. Rieth
	Ken Schmitt
	Stefan Schmitt
	Bernhard Schmittner
	Nicolaj Stache
	Jurgen Völkel
	Boris Werthessen
	Chen Zhang
Daimler AG	Jörg Breuer
	Uwe Franke
	Stefan Gehrig
	Sebastian Geyer
	Wolfgang Hurich
	Bernhard Morys
	Stephan Mücke
	Uwe Regensburger
	Michael Schopper
	Hans-Peter Schöner
	Simon Tattersall
	Christoph von Hugo
John Deere GmbH & Co. KG	Georg Kormann
	Marco Reinards
	Udo Scheff
Knorr-Bremse SfN GmbH	Falk Hecker
MAN Truck & Bus AG	Karlheinz Dörner
	Eberhard Hipp
	Walter Schwertberger
Novartis Pharma AG	Muriel Didier
Omron Electronics GmbH	Georg Otto Geduld (formerly)
Robert Bosch Car Multimedia GmbH	Ralph Behrens
	Andreas Engelsberg
	Thomas Kleine-Besten
	Werner Pochmuller
	Heiner Schepers
Robert Bosch GmbH	Susanne Ebel
	Hendrik Fuchs
	Frank Hofmann
	Ulrich Kersken
	Winfried König (formerly)
	Friedrich Kost
	Hans Löhr
	Martin Noll
	Peter Rapps
	Gunther Schaaf
	Anton van Zanten (formerly)
	Ulf Wilhelm
Unfallforschung der Versicherer	Matthias Kuhn
Valeo Schalter und Sensoren	Heinrich Gotzig
Volkswagen AG	Arne Bartels
	Michael Darms
	Sebastian Hamel
	Marc-Michael Meinecke
	Michael Rohlfs

	Falko Saust
	Simon Steinmeyer
ZF Lenksysteme GmbH	Peter Brenner
	Hendrik Büring
	Gerd Reimann

Universities

Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE	Marcel Baltzer Matthias Heesen
Hochschule München	Markus Krug
Karlsruher Institut für Technologie (KIT)	Peter M. Knoll
	Christoph Stiller
Max-Planck-Institut für Informatik	Bernt Schiele
Max-Planck-Institut für Intelligente Systeme	Andreas Geiger
RWTH Aachen	Eugen Altendorf
	Frank Flemisch
	Sonja Meier
Technische Universität Braunschweig	Markus Maurer
	Andreas Reschka
	Jens Rieken
Technische Universität Darmstadt	Bettina Abendroth
	Jurgen Adamy
	Eric Bauer
	Ralph Bruder
	Norbert Fecher
	Benjamin Franz
	Rolf Isermann
	Michaela Kauer
	Tran Quoc Khanh
	Ulrich Konigorski
	Ingmar Langer
	Stefan Leinen
	Matthias Pfromm
	Raphael Pleß
	Bernt Schiele
	Matthias Schreier
	Kai Schröter
	Christina Singer
	Nico Steinhardt
	Alexander Stoff
	Alexander Weitzel
	Hermann Winner
Technische Universität Dresden	Lars Hannawald
	Bernhard Schlag
	Gert Weller
Technische Universität München	Klaus Bengler
	Martin Kienle
	Thomas Weißgerber
Universität der Bundeswehr München	Berthold Färber
	Verena Nitsch

Universität Ulm	Klaus C. J. Dietmayer Dominik Nuss Stephan Reuter
University of South Carolina	Bryant Walker Smith

Contents

Volume 1

Volume 2

About the Editors

From left to right: Stephan Hakuli, Hermann Winner, Christina Singer, Felix Lotz

Hermann Winner began working at Robert Bosch GmbH in 1987 after receiving his Ph.D. in physics, focusing on the predevelopment of "by-wire" technology and adaptive cruise control (ACC). Beginning in 1995, he led the series development of ACC up to the start of production. Since 2002, he has been pursuing the research of driver assistance systems engineering topics as professor of Automotive Engineering at the Technische Universität Darmstadt. In 2012, he got the IEEE-ITS Institutional Award for Institutional Leadership in Research and System Engineering for Advanced Driver Assistance and Safety.

After finishing his studies in physics, Stephan Hakuli developed driving functions for highly automated vehicles as a research associate at the Institute of Automotive Engineering (FZD) at the Technische Universität Darmstadt. Between 2011 and 2015 he has worked as product manager and subject specialist for driver assistance systems at IPG Automotive GmbH. Now, he works at Continental Engineering Services GmbH in the Chassis & Safety segment as head of business development.

Felix Lotz studied Mechanical Engineering at Technische Universität Darmstadt and Virginia Polytechnic Institute and State University. Between 2011 and 2014, he has worked as a scientific assistant at the Institute of Automotive Engineering of Technische Universität Darmstadt and focused on research in the fields of system architectures and behavior planning of automated vehicles. Now, he works as a project manager in the field of automated driving at Continental AG.

Christina Singer studied Mechanical Engineering at Fachhochschule Südwestfalen, Technische Universität Darmstadt, and Virginia Polytechnic Institute and State University. Since 2011, she has been working as scientific assistant at the Institute of Automotive Engineering at Technische Universität Darmstadt, where her research is focused on effort-reduced application and release concepts for brake system controllers.

Contributors

Bettina Abendroth Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany

Jürgen Adamy Control Methods and Robotics Lab, Technische Universität Darmstadt, Darmstadt, Germany

Eugen Altendorf Institut für Arbeitswissenschaft (IAW), RWTH Aachen, Aachen, Germany

Alexander Bachmann ADC Automotive Distance Control Systems GmbH, Lindau, Germany

Marcel Baltzer Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany

Arne Bartels Fahrerassistenz und integrierte Sicherheit, Volkswagen AG – Konzernforschung, Wolfsburg, Germany

Eric Bauer Control Systems and Mechatronics Lab, Technische Universität Darmstadt, Darmstadt, Germany

Bernward Bayer Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Ralph Behrens Robert Bosch Car Multimedia GmbH, Hildesheim, Germany

Klaus Bengler Fakultät für Maschinenwesen, Technische Universität München, Garching, Germany

Guy Berg BMW Group, München, Germany

Peter Brenner Robert Bosch Automotive Steering GmbH, Schwäbisch Gmünd, Germany

Jörg Breuer RD/FZ, Daimler AG, Sindelfingen, Germany

Ralph Bruder Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany

Stefan Bunzel Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Hendrik Büring Robert Bosch Automotive Steering GmbH, Schwäbisch Gmünd, Germany

Axel Büse Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Stephan Cieler Interior Electronics Solutions, Continental Automotive GmbH, Babenhausen, Germany

Michael Darms Continental AG, Lindau, Bavaria, Germany

Muriel Didier Novartis Pharma AG, Basel, Switzerland

Frank Dierkes Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Klaus C. J. Dietmayer Institut für Mess-, Regel- und Mikrotechnik, Universität Ulm, Ulm, Germany

Edmund Donges Fürstenfeldbruck, Germany

Karlheinz Dörner Engineering Research Electronics (ERE), MAN Truck & Bus AG, München, Germany

Susanne Ebel Robert Bosch GmbH, Chassis Systems Control, Driver Assistance Systems, Leonberg, Germany

Andreas Engelsberg Bosch SoftTec GmbH, Hildesheim, Germany

Berthold Färber Institut für Arbeitswissenschaft, Universität der Bundeswehr München, Neubiberg, Germany

Norbert Fecher Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany

Frank Flemisch Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany

Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany

Oliver Fochler Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Uwe Franke Group Research and Advanced Engineering, Driver Assistance and Chassis Systems, Daimler AG, Sindelfingen, Germany

Benjamin Franz Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany

Hendrik Fuchs Corporate Research, Digital Communication and C2XC Systems, Robert Bosch GmbH, Hildesheim, Germany

Simon Fürst BMW Group, München, Germany

Tom Michael Gasser Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany

Georg Geduld Eichberg, Switzerland

Stefan Gehrig Group Research and Advanced Engineering, Driver Assistance and Chassis Systems, Daimler AG, Sindelfingen, Germany

Andreas Geiger MPI for Intelligent Systems, Tübingen, Germany

Sebastian Geyer Mercedes-Benz Cars, Daimler AG, Sindelfingen, Germany

Heinrich Gotzig Valeo Schalter und Sensoren GmbH, Driving Assistance Product Group, Bietigheim-Bissingen, Germany

Steffen Gruber Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Stephan Hakuli Continental Engineering Services GmbH, Frankfurt am Main, Germany

Sebastian Hamel Volkswagen AG, Wolfsburg, Germany

Lars Hannawald Verkehrsunfallforschung an der TU Dresden GmbH, Technische Universität Dresden, Dresden, Germany

Falk Hecker Knorr-Bremse AG, Schwieberdingen, Germany

Matthias Heesen Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany

Eberhard Hipp MAN Truck & Bus AG, München, Germany

Jens Hoffmann Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Frank Hofmann Corporate Research, Digital Communication and C2XC Systems, Robert Bosch GmbH, Hildesheim, Germany

Andree Hohm Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Maximilian Höpfl Continental AG/Continental Automotive GmbH, Frankfurt am Main, Germany

Wolfgang Hurich RD/FFS, 059-X820, Daimler AG, Sindelfingen, Germany

Rolf Isermann Institut für Automatisierungstechnik, Technische Universität Darmstadt, Darmstadt, Germany

Michaela Kauer Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany

Ulrich Kersken Robert Bosch GmbH, Hildesheim, Germany

Tran Quoc Khanh Fachgebiet Lichttechnik, Technische Universität Darmstadt, Darmstadt, Germany

Martin Kienle Lehrstuhl für Ergonomie, Technische Universität München, Garching, Germany

Felix Klanner BMW Forschung und Technik, München, Germany

Lars Kristian Klauske Carmeq GmbH, Berlin, Germany

Thomas Kleine-Besten Robert Bosch Car Multimedia GmbH, Hildesheim, Germany

Peter M. Knoll Institute for Theoretical Electrical Engineering and System Optimization, KIT (Karlsruhe Institute of Technology), Karlsruhe, Germany

Winfried König Offenburg, Germany

Ulrich Konigorski Control Systems and Mechatronics Lab, Technische Universität Darmstadt, Darmstadt, Germany

Georg Kormann Advanced Engineering, John Deere GmbH & Co. KG, Intelligent Solutions Group, Kaiserslautern, Germany

Friedrich Kost Robert Bosch GmbH, Abstatt, Germany

Markus Krug Department of Automotive Engineering, Hochschule München, München, Germany

Matthias Kühn Unfallforschung der Versicherer, Gesamtverband der Deutschen Versicherungswirtschaft e.V., Berlin, Germany

Ingmar Langer Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany

Stefan Leinen Institut für Geodäsie, Fachgebiet Physikalische Geodäsie und Satellitengeodäsie, Technische Universität Darmstadt, Darmstadt, Germany

Martin Liebner Forschung und Technik, BMW Group, München, Germany

Paul Linhoff Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Hans Löhr Corporate Research, Security and IT Systems, Robert Bosch GmbH, Renningen, Germany

Felix Lotz Continental AG, Frankfurt am Main, Germany

Stefan Lüke Continental Division Chassis & Safety, Advanced Engineering, Frankfurt am Main, Germany

Mark Mages Chassis & Safety, Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Roman Mannale Continental Automotive GmbH, Schwalbach, Germany

Richard Matthaei Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Markus Maurer Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Marc-Michael Meinecke Fahrerassistenz und integrierte Sicherheit, Volkswagen AG – Konzernforschung, Wolfsburg, Germany

Stefan Menzel Continental AG/ADC Automotive Distance Control Systems GmbH, Lindau, Germany

Thorsten Mlasko Bosch SoftTec GmbH, Hildesheim, Germany

Matthias Mörbe Bosch Engineering GmbH (BEG-SC/PJ-T), Heilbronn, Germany

Bernhard Morys RD/FFS, 059-X820, Daimler AG, Sindelfingen, Germany

Stephan Mücke RD/FAA, Daimler AG, Sindelfingen, Germany

Verena Nitsch Institut für Arbeitswissenschaft, Universität der Bundeswehr München, Neubiberg, Germany

Martin Noll CC-DA/EAU3, Robert Bosch GmbH, Leonberg, Germany

Dominik Nuss Institut für Mess-, Regel- und Mikrotechnik, Universität Ulm, Ulm, Germany

Norbert Ocvirk Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Matthias Pfromm Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany

Bernd Piller Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Raphael Pleß Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany

Werner Pöchmüller Robert Bosch Car Multimedia GmbH, Hildesheim, Germany

Martin Punke Continental AG/ADC Automotive Distance Control Systems GmbH, Lindau, Germany

Peter Rapps Karlsruhe, Germany

Thomas Raste Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Uwe Regensburger Daimler AG, Sindelfingen, Germany

Gerd Reimann Robert Bosch Automotive Steering GmbH, Schwäbisch Gmünd, Germany

Marco Reinards John Deere Werk Mannheim, John Deere GmbH & Co. KG, Mannheim, Germany

James Remfrey Continental Corporation, Frankfurt am Main, Germany

Andreas Reschka Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Stephan Reuter Institut für Mess-, Regel- und Mikrotechnik, Universität Ulm, Ulm, Germany

Jens Rieken Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Peter E. Rieth Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Michael Rohlfs Volkswagen AG, Wolfsburg, Germany

Christian Ruhhammer BMW Group, München, Germany

Falko Saust Volkswagen AG, Wolfsburg, Germany

Gunther Schaaf Corporate Research, Driver Assistance Systems and Automated Driving, Robert Bosch GmbH, Renningen, Germany

Thomas Schaller BMW Group, München, Germany

Udo Scheff John Deere Werk Mannheim, John Deere GmbH & Co. KG, Mannheim, Germany

Heiner Schepers Robert Bosch Car Multimedia GmbH, Hildesheim, Germany

Bernt Schiele Computer Vision and Multimodal Computing, Max Planck Institute for Informatics, Saarbrücken, Germany

Bernhard Schlag Technische Universität Dresden, Dresden, Germany

Ken Schmitt Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Stefan Schmitt Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Bernhard Schmittner Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Hans-Peter Schöner RD/FFS, 059-X820, Daimler AG, Sindelfingen, Germany

Michael Schopper Daimler AG, Sindelfingen, Germany

Matthias Schreier Control Methods and Robotics Lab, Technische Universität Darmstadt, Darmstadt, Germany

Kai Schröter Formerly Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany

Walter Schwertberger MAN Truck & Bus AG, München, Germany

Andre Seeck Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany

Patrick Seiniger Referat F1, Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany

Bryant Walker Smith Stanford Law School, Center for Internet and Society, Stanford, CA, USA

Nicolaj Stache Continental AG/Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Nico Steinhardt Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany

Simon Steinmeyer Volkswagen AG, Wolfsburg, Germany

Christoph Stiller Institut für Mess- und Regelungstechnik, Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany

Alexander Stoff Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany

Simon Tattersall RD/FAA, Daimler AG, Sindelfingen, Germany

Simon Ulbrich Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany

Anton van Zanten Ditzingen, Germany

Jürgen Völkel Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Christoph von Hugo RD/FAA, Daimler AG, Sindelfingen, Germany

Thomas Weißgerber Lehrstuhl für Ergonomie, Technische Universität München, Garching, Germany

Alexander Weitzel Formerly Technische Universität Darmstadt, Darmstadt, Germany

Gert Weller Technische Universität Dresden, Dresden, Germany

Moritz Werling BMW Group, München, Germany

Boris Werthessen Continental AG/ADC Automotive Distance Control Systems GmbH, Lindau, Germany

Ulf Wilhelm Robert Bosch GmbH, Ditzingen, Germany

Thomas Winkle Daimler and Benz Foundation, Ladenburg, Germany

Hermann Winner Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany

Christian Wojek Carl Zeiss AG, Oberkochen, Germany

Chen Zhang Continental Teves AG & Co. oHG, Frankfurt am Main, Germany

Part I

Fundamentals of Driver Assistance Development

Capabilities of Humans for Vehicle Guidance

1

Bettina Abendroth and Ralph Bruder

Contents

Abstract

Human information processing systems, as well as the individual driver characteristics interacting reciprocally with them, are particularly significant for the task of vehicle guidance. This chapter will describe these connections between driver, vehicle and environment using a simple system model. The driver's intake, processing and output of information will be delineated. The relevant driver characteristics, capacities and skills for vehicle guidance will be described. Based on this understanding, requirements for vehicle guidance with regards to the driver will be systematized by considering subtasks and evaluated with respect to the limits of human capacity.

B. Abendroth (✉)
Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: abendroth@iad.tu-darmstadt.de

R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: bruder@iad.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_1

Vehicle guidance is a predominantly informational activity in which work content and information are converted into reactions. Normally, the driver performs the action of steering while continually processing information.

Accordingly, the most significant factors associated with vehicle guidance are the systems of information processing and the individual driver characteristics which interacts reciprocally with them.

In order to describe the connections between driver, vehicle and environment, a simple system model will be employed (cf. Abendroth 2001). It consists of two elements: driver and vehicle. The input variable of vehicle guidance, which is also influenced by environmental factors, impacts both of these elements. Above and beyond this, disruptive variables such as distractions caused by passengers may arise. This system's output variable can be described by the system functions of mobility, safety and comfort.

1 Human Information Processing

In order to describe human information processing there are a variety of models which specify general intake, which transforms a signal entering a receptor (stimulus) into a cognitive representation and a human reaction (response). Some of the most well-known models in the field of engineering include the sequential and resource models. Sequential models allege that the transformation from stimulus to response occurs in a strictly sequential manner, meaning that the next step can only be performed once the previous one is completed. Resource models are based on the assumption that the capacity available for various activities is limited and must be shared between all simultaneously performed tasks. The theory of multiple resources extends this view; according to this theory, the degree of interference between two tasks depends on whether they demand the same resources (Wickens 1992). In this model, simultaneous processing of visual, spatial image information (e.g., navigation displays) and auditory, verbal information (telephone calls, news on the radio) would be free from interference, since they use different sensory channels and different regions of the working memory. However, experimental studies have shown that this freedom from interference is not absolute.

Human information processing can be explained through a combined sequential and resource model (see Fig. 1). This is based on the following processing steps: information intake (perception), information processing in the narrower sense (cognition) and information output (motor function) (Schlick et al. 2010). Additionally, it must be remembered that the available capacity of resources is limited.

The efficiency of these three levels of the information processing system is influenced by available processing resources and requires the application of attention. This leads to targeted selection of information which is intended to be the content of conscious processing. The constant oversupply of information exceeds human processing capacity, so that it is impossible for a human being to consciously perceive everything which reaches the level of the sensory receptors.

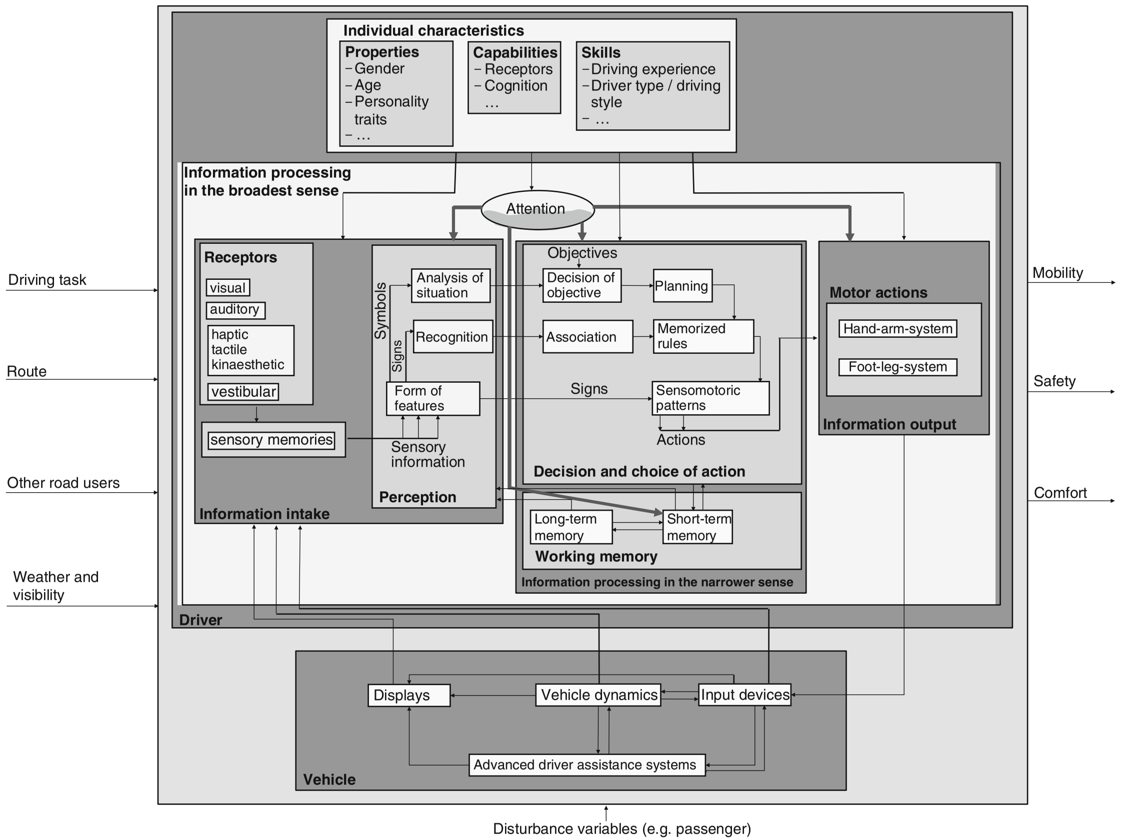


Fig. 1 Driver-vehicle-environment system model (cf. Abendroth 2001)

Human beings can distribute their entire attention to varying degrees among the three levels of the information processing system in order to select relevant sources of information and further process this information. For each operational task it is possible to learn an efficient distribution of attention, while in extreme cases a poor distribution of attention can cause human error.

On a theoretical level, varying forms of attention can be divided into the dimensions of selectivity and intensity. Selective use of attention describes the fact that human beings must decide between different competing sources of information. Within this divided attention, a human being must simultaneously perceive various stimuli, while a shift in attention means ignoring one stimulus in order to subsequently engage with another. Intensity of attention concerns the level of activation; relevant factors here include reduced vigilance (low proportion of relevant stimuli) and sustained attention (high proportion of relevant stimuli).

1.1 Information Intake

Information intake includes all processes relating to the discovery and recognition of information. The procedure of internally representing the environment will be designated as perception. This internal image of the environment is influenced by the situation in which a person finds themselves and the experiences at their disposal. Information intake occurs with the sensory organs. A human being can take in a variety of simultaneously transmitted information through all sensory channels in parallel; however, simultaneously processing a variety of information can diminish performance. The specific performance range of the sensory organs influences the quantity and quality of information absorbed, and thus every subsequent step of information processing. During driving a vehicle, visual, acoustic, tactile and vestibular perception are of the highest significance. The sensory memory register (also called ultra-short term memory) is also classified under information intake. The sensory memory register exclusively stores physically-coded information. Visual information is stored in the iconic memory register and acoustic information in the echoic register for a period of between 0.2 s and 1.5 s respectively (Schlick et al. 2010).

During the intake of visual information, the eye has the following three basic tasks: adaptation (adjusting the eye's sensitivity to the current light density), accommodation (alignment to varying visual distances) and fixation (directing the eyes onto the visual object so that both optical axes converge). The eye aids colour and object recognition as well as the perception of motion, spatial depth and size.

The ear fulfils three basic functions during the intake of auditory information: adaptation (increasing the auditory threshold necessary for distinguishing the hearing process), auditory pattern recognition (necessary for language and noise identification) and acoustic spatial orientation, which is accomplished by binaural (two-ear) hearing.

Tactile information intake involves the haptic or kinaesthetic perception channel. The tactile perception system enables recognition of distortions to the skin. Receptors (lamellar corpuscles and Meissner's cutaneous receptors) impart sensations of pressure, touch and vibration in and beneath the skin. The kinaesthetic perception system recognizes the stretching of muscles and the movement of joints. Various types of receptors located on the muscle spindles, joint regions and tendons enable the perception of body movements and the relative positions of parts of the body with respect to one another.

Spatial orientation is accomplished in human beings by the vestibular perception system. The vestibular apparatus, located in the inner ear, functions as a receptor. This apparatus also has the task of relaying information about maintaining balance and triggering positional reflexes for normal head and eye posture. During vehicle guidance, the vestibular sensory channel contributes to the perception of speed and acceleration of the vehicle.

Most of the information relevant to driving an automobile is taken in visually (usually around 80–90 % (Rockwell 1972)). The basis for a driver's correct choice of action is an optimally complete internal representation of the relevant transport space. It is also important for the driver to take in information relevant to vehicle guidance from a great distance, leaving sufficient time to react to this information. This can only be guaranteed by the eye, since the eye is the only far-reaching human receptor system which can be focused selectively (Cohen et al. 1991).

During tasks encompassing human behaviour in traffic, information intake is strongly dominated by the limits of eye movement. The area from which a driver can take in visual information is determined by the driver's visual field and range of vision while looking forward or turning around. Depending on the object's mapping location on the retina, there is a differentiation between foveal and peripheral vision: with foveal vision, the object is mapped onto the central depression of the retina (fovea); in this region, only objects within an aperture angle of 2° can be seen distinctly. The further the image is from the fovea, the less distinct it will appear. In the peripheral field of vision, movements and changes of brightness can be perceived. In the literature, various views can be found on the role of foveal and peripheral vision and their contribution to information intake during driving. For example, Cohen et al. (1991) assert that foveal information intake during driving plays a significant role under heavy stress, that is, under a high density of information and therefore high demands on information processing.

The size of the usable field of vision varies between good and bad drivers. While good drivers possess a usable field of vision of 9–10°, bad drivers can only manage 6–7° (Färber 1987). The usable field of vision is defined as a variable spatial expansion around the fovea which determines the region within which a person can identify the information necessary for performing a particular defined task.

The quality of a person's visual information intake is influenced by the type of signals and the frequency of their occurrence. Thus Schmidtke (1993) distinguishes between critical, neutral and non-critical signals, as well as critical and non-critical additional signals. With respect to the frequency of information occurrence, the studies of several authors (Schmidtke 1993 gives an overview) have shown that observational performance improves under an increased occurrence of signals requiring reaction over a unit of time. This rule applies up to an optimal signal frequency of approximately 120–300 signals per hour. If this signal frequency is significantly exceeded, the observer enters a situation of excessive demand, with the result that more and more signals remain unanswered. In their “Theory of Pathway Inhibition”, Galinsky et al. (1990) assert that similar stimuli interfere with one another, and therefore heterogeneous stimuli result in better attention performance.

1.2 Information Processing

Signals from the environment (e.g., road conditions, other vehicles, weather and visual conditions) and the vehicle (e.g., displays, input devices and vehicle dynamics) are taken in by human receptors, adapted and then further processed at the level of information processing in the narrower sense (cognition). Here a decision is made about whether information should lead to an action (active case) or is merely endured (passive case). This decision is principally influenced by the driver’s individual characteristics. The range of choices and actions can be explained by three levels of behaviour which build upon one another, which Rasmussen (1983) explains as skill-based, rule-based and knowledge-based (see ▶ Chap. 2, “Driver Behavior Models,” Sect. 3). The behaviour level on which information processing occurs depends on the type of the task to be performed as well as the driver’s individual characteristics, in particular the driver’s experience with the demands.

Memory plays a central role in the cognitive processing of information. With the help of memory, sensory impressions are compared with learned and stored structures of thought and judgment. According to the classic three-level storage model, memory consists of the sensory register (ultra-short term memory), short-term memory and long-term memory. Information is actively processed in short and long-term memory. In a continuous process, stored information is summoned from long and short-term memory and compared with those characteristic traits absorbed by the senses.

A driver’s risk of accident is influenced by individual tolerance and inaccurate perceptions of traffic risks. One significant aspect of the decision procedure within information processing is the fact the selected action promises the greatest utility under varying external circumstances with regard to associated risk. The concept of risk is defined in various ways. Often it is interpreted as the probability that an undesirable outcome will occur. Thus, for example, Brown and Groeger (1988) define risk as the relation between variables which describe the negative consequences of outcomes and variables which characterize the probability of those conditions arising under which the consequences represent a possibility. This view, however, does not include consciousness of risk. Wagenaar (1992) sees risk as a multidimensional characterization of a negative expectation resulting from a probabilistic decision process.

Countless models have been developed to explain drivers’ risk perception. Among the most well-known are the “zero risk” model (Näätänen and Summala 1976) and the “risk homoeostasis” model (Wilde 1982). According to the zero risk model, humans behave such that their subjective risk amounts to zero; this model is based on the individual motivations influencing driving behaviour and adaptation to risks perceived in road traffic. The theory of risk homoeostasis asserts that, upon reduction of objective risk (e.g., through technical measures), humans alter their behaviour so far in the direction of "more dangerous" that their subjective appraisal of risk takes on the same distance to their personally accepted risk as before the introduction of the measure (Wilde 1982).

The model of subjective and objective safety contrasts subjectively experienced safety with those forms of safety which are physically measurable (von Klebelsberg 1982). The threat-avoidance model (Fuller 1984) proceeds from the assumption that a driver's behaviour when perceiving a potentially dangerous event is primarily selected by weighing the utility and cost of all the alternatives.

The chief components of risk perception, according to Brown and Groeger (1988), are information about potential dangers in the traffic environment and information about the capability of the driver-vehicle system, each of which mitigate the potential danger of an accident.

1.3 Information Output

On the third level of the information processing system, decisions reached at the level of information processing in the narrow sense are transformed into actions. For driving, these actions encompass motor movements of the hand-arm system as well as the foot-leg system. The physical stress of a physiologically performed task is low compared with the stress associated with information intake and processing and is further reduced by technical support systems in the vehicle (e.g., power steering).

2 Driver Characteristics and the Limits of Human Performance Capacity

Human performance is generally characterized by work output and the demands on the individual carrying out the task. Both work output and demands are subject to inter- and intra-individual variation: not all individuals fulfil the same task equally well, but even individuals can demonstrate performance variability if performance of the same task is measured at different points in time. This variability can be attributed to individual human characteristics and thus also to varying performance conditions. The following section will delineate the relevant human performance conditions for vehicle guidance and their impact on driving performance and safety. To this end, a systematic differentiation between characteristics, capacities and skills will be carried out.

Properties Properties will be defined as intra-individual influence variables which are predominantly independent with regards to time (or which only change over very large periods of time). The most relevant traits for driving are often identified as gender, age and personality traits.

While some studies have identified gender-specific differences in driving behaviour, other studies were unable to confirm differences with respect to risk behaviour or speed. Nevertheless, differences in perception of the risk of an accident have been determined between men and women: men estimate their driving capabilities better than women, though women tend to underestimate their capacity while men tend to overestimate theirs. Furthermore, male drivers judge certain behaviours to be less dangerous and less likely to result in an accident than female drivers.

The human capacity to orient oneself with the senses, process information received and carry out motor actions alters with age, during which human organs undergo changes. Increasing functional deficits can be at least partially compensated for by the fact that older drivers generally have more driving experience. There are various approaches to defining the term "older". Often calendar or chronological age is used as an orientation point, according to which people are considered older after their 60th or 65th year of life, even though the functional changes associated with the aging process are subject to significant inter-individual variations.

Various personality traits can also influence a driver's behaviour. Thus, correlations have been identified between a driver's risk tolerance and their driving speed and use of traction. Drivers who are emotionally unstable, impulsive and incapable of teamwork have a higher accident risk than people who are adaptable and emotionally stable. Furthermore, selectiveness, perception style and reaction time are identified as individual traits which act as indicators of accident involvement.

Capacities Capacities are defined as accessible, intra-individual, time-dependent short or long-term changes; they affect the physiological function of organs, or the so-called basic human functions.

A driver's actions are influenced by those mental capacities defined as intelligence, particularly on a knowledge-based level. The concept of intelligence is debated in the literature. According to a broad definition, intelligence is understood to be the hierarchically structured entirety of those general mental abilities which determine the level and quality of a person's thought process. With the help of these abilities, the characteristics significant to action within a problem situation can be recognized in context, allowing for alteration of perceived objectives according to the situation.

However, human cognitive and sensomotoric abilities such as reaction capacity also indirectly influence driving through the impact of these traits on information processing.

With increasing age, the ability of receptors diminishes, which leads to overall limitation in information intake.

Parts of the eye change with age due to a loss of tissue fluid. The resulting effects on visual capacity are summarized in Table 1.

Progressive limitation of the visual field during aging aggravates motion perception issues while driving, since the motion of relevant objects is first observable in the peripheral field of vision.

Table 1 Changes to the visual system which occur with age (↑ increase; ↓ decrease)

Impact		Cause and influence variables	
↓	Range of accommodation	↓	Fluid in tissue
↓	Static visual acuity		Lighting conditions
↓	Dynamic visual acuity	↓	Speed of accommodation
		↑	Dullness of sensory cells
↑	Sensitivity to bright light	↑	Functional disruptions in the retina
		↑	Adaptation time
↓	Contrast sensitivity		
↑	Necessary light density	↑	Deterioration of cornea, lens and vitreous body
↑	Limitation of the visual field		

Changes in hearing during aging include reduction in auditory amplitude, particularly at higher frequencies. Difficulties in discriminating the frequency and intensity of tones, as well as in recognizing complex noises such as language under difficult perceptive conditions (e.g., background noise, distortion) and partially impeded directional hearing are further changes.

As age increases there is also a loss in sensitivity of tactile perception.

The sense of equilibrium is best formed in 20–30-year-olds and decreases strongly after the 40th year of life, reducing by half in the ages between 60 and 70.

Sensory storage works less efficiently as age increases. Acoustic signals demonstrate a higher speed of decay in echoic storage, while visual signals remain longer in iconic storage. During the preparation of information relevant to driving, this leads to acoustic information only being available for processing for a temporally shortened span, while visual stimuli can only be received to a limited extent due to blocking of the iconic storage register.

Among the individual areas of attention, older people exhibit reduced performance, which in their additive effect results in an overall decrease in attention capacity. Consequently, older drivers must decide on their actions based on a relatively smaller pool of environmental information than younger drivers, since they do not have access to all potentially important information.

Overall it has been demonstrated that older drivers can encounter difficulties, above all in complex and novel situations which require fast action. Additional impediments occur in the form of limitations to information intake, which result in partially delayed sensorial preparation of relevant information, leaving less time for older drivers to process information relevant to driving and act accordingly.

Skills Skills are understood as human working functions which are determined by basic human functions and the concrete formal state of the task and work environment. In the context of driving, driving experience, driving style (classified by vehicle dynamic parameters chosen by the driver) and driver type (classified by observed driving behaviour) are particularly significant. Driving experience can have varying effects on the risk of accidents. With increased driving experience, driving skills are improved, along with recognition and judgment of risks. Improvement in driving skills can be attributed to the fact that the number of varying driving situations experienced grows with an increase in distance driven, thus enabling the formation of routine actions. While vehicle control becomes better with increased driving experience, experience in other areas leads to the formation of errors and bad habits, for example not looking in the mirror, braking late and tailgating. When it comes to skills which reflect their control over the vehicle, beginners have been shown to perform more poorly than experienced drivers. This is evident in late acceleration, poor and inconsistent steering motions and slow gear shifting. Inexperienced drivers make more steering motions than experienced drivers. Inexperienced drivers' eye behaviour is often described as less efficient, since they fixate too frequently on points in close proximity. Thus young, inexperienced drivers recognize distant accident risks relatively poorly compared to experienced drivers; however, there is no difference between the two groups when it comes to recognizing nearby dangers. With increasing experience drivers learn to recognize dangerous objects and events based on certain parts of the traffic system. This also corresponds with the fact that visual fixation and search patterns differ between inexperienced and experienced drivers. Various rates of speed are demonstrated when driving around curves depending on driving experience. Experienced drivers drive faster into curves and slow down more within the curve than inexperienced drivers.

Driving style is influenced both by driving experience and the personality of the driver. Various driving styles have been identified. These can be used to describe drivers of commercial vehicles as "weak and lax," "jerky and abrupt" or "swift and lively." Drivers of passenger vehicles are identified by parameters for speed, longitudinal acceleration and distance to the car in front as "slower and more comfort-conscious," "average, with high safety consciousness" and "fast and sporty." On the basis of behaviour observation, similar driver types were found, which were described as "inattentive average driver," "less habitual, indecisive driver," "sporty, ambitious driver" and "risk-seeking, aggressive driver."

3 Demands on Vehicle Operators in the Driver-Vehicle-Environment System

The demands on the driver result from the task of vehicle guidance, which is jointly determined by environmental factors. Here, the complexity of the situation which the driver must manage is paramount. This results from the characteristics of the route and the dynamic behaviour of other road users. How the driver manages these challenges depends on both the driver's individual characteristics and the driver support offered by the vehicle (advanced driver assistance systems). Depending on the volume and duration of stress, bottlenecks occur in the driver's information processing system, which can lead to deviation from so-called "normal behaviour" all the way to critical traffic situations and even accidents, based on the continuum of traffic behaviour as outlined by von Klebelsberg (1982). In order to identify these bottlenecks, the following section will compile the subtasks involved in vehicle guidance, and the demands resulting from these tasks.

Subtasks Involved in Vehicle Guidance Approaches to describing the task of vehicle guidance by means of subtasks exist at various levels of detail; some were derived for special explanatory purposes or singular aspects of vehicle guidance. In the following section, only two frequently mentioned classifications will be presented.

An arrangement of driver tasks according to their significance to fulfilling the purpose of a journey has been suggested by Bubb (2003). Primary activities consist of those activities which are absolutely necessary for the completion of the journey, such as steering and pressing on the accelerator, and which are predominantly determined by the course of the road, other road users and environmental conditions. Secondary activities are characterized by an output of information to the environment (e.g., honking the horn or using the indicators) as well as reactions to the current situation, such as turning on the windscreen wipers or turning on headlights. Tertiary actions are not directly connected with the actual operation of the vehicle, but rather serve to increase the comfort of a journey (e.g., controlling ventilation and air conditioning or operating the radio).

The 3-level model proposed by Donges (1982) (see ▶ Chap. 2, “Driver Behavior Models”, Sect. 4) describes a hierarchy of primary driver tasks at the highest level using the following activities: navigation (choosing the route for a journey), guidance (determining target lane and target speed) and stabilization (adjusting vehicle movements to the designated driving variables).

This hierarchy also reflects the temporal margin available for the execution of a particular task, along with the tolerance for error. While a delayed decision or error at the navigation level generally does not lead to a critical situation, thoroughly critical driving situations or even accidents can arise at the stabilization level.

Requirements for Vehicle Guidance Generally the requirements involved in an activity stem from the tasks to be performed. With regard to task non-specific, situational operating conditions, stressors arise which can be described objectively. Among these situational factors, the duration and temporal composition of these requirements and influences from the operational environment are significant.

In order to investigate the requirements for operational tasks, various procedures of activity analysis have been developed. In order to analyse the requirements for driving, Fastenmeier (1995) created a version of the “Fragebogen zur Arbeitsanalyse” (work analysis questionnaire, FAA, Frieling and Graf Hoyos 1978), modified for road transport. This modified version takes into account information processing and vehicle control, while the former is further subdivided into information sources, sensory and perception processes, acts of judgement and thought and decision processes. A total of 32 operational elements were defined for information processing and seven operational elements for vehicle manipulation.

On the basis of the FAA modified by Fastenmeier (1995) and with the help of the section on the cognitive performance of the “Tätigkeitsbewertungssystem” (activity evaluation system, TBS, Hacker et al. 1983), the requirements associated with vehicle guidance can be derived.

In the list of requirements presented below, the domain of information sources, sensory and perception processes encompasses acts of orientation to the environment. Additionally, perceived circumstances are registered and processed as signals. Signals are stimuli which are identified and differentiated, which have a particular meaning for the activity of operation when appearing in a particular manifestation and which mark a specific action as necessary. Evaluation is rendered by the deduction of diagnoses about circumstances used to find appropriate measures. To this end, stimuli are set apart and compared and signal manifestations are combined. Requirements on thought and decision-making can consist of diagnostic efforts which encompass the investigation of possible variants or prognostic efforts which assist the selection of functional variants. Vehicle control takes place in the context of processing efforts.

I Information Sources, Sensory and Perception Processes

- Visual displays in the vehicle
  - Instruments (e.g., speedometer), positioning of input devices (e.g., heated rear window), information from on-board computer (e.g., exterior temperature)
- Acoustic information
  - Speech output from the navigation system, sirens on emergency and rescue vehicles
- Secondary acoustic information
  - Radio, conversations with passengers or on the telephone
- Other road users
  - Vehicles, pedestrians
- Characteristics of the route
  - Transversal and longitudinal course of the route, junctions, road width, number of lanes
- Traffic signs
  - Speed limits, right of way and direction signs
- Condition of the road surface, weather and visual conditions
  - Moisture, dirt, snow, black ice; light in the eyes, rain or snowfall, fog

E Evaluation

- Longitudinal distances from or between other road users or objects
  - From the vehicle in front, between two vehicles in adjacent lanes, from pedestrians, cyclists and obstacles in the same lane.
- Horizontal distances from or between other road users or objects
  - Vehicles at the “same point” as vehicles by the side of the road
- Speed of the driver’s vehicle and other vehicles or road users
- Anticipation of critical traffic situations
  - Sudden vehicle merging, disregard of right of way by others, a child running into the street

D Decision Making and Thought Processes

- Selecting suitable actions for navigating the vehicle
  - Deciding which route to choose, which direction to take at intersections
- Selecting suitable actions for vehicle guidance
  - Deciding what speed to drive at and what distance to maintain from other cars, passing manoeuvres, choosing a lane and a lateral position therein

V Vehicle Manipulation

- Controlling the vehicle's longitudinal motion in order to stabilize the vehicle
  - Accelerating, braking, shifting gears
- Controlling the vehicle's horizontal motion in order to stabilize the vehicle
  - Steering
- Controlling further functions
  - Lights, windscreen wipers, radio

4 Evaluating the Requirements for Driving with Respect to Human Performance Capacity

In conclusion, the demand areas delineated above will be evaluated with respect to human performance capacity in order to demonstrate meaningful areas for technical driver support.

Information Sources, Sensory and Perception Processes Perception of those information sources relevant to fulfilling the task of vehicle guidance is of the utmost importance for the driver. With the help of this information, the driver creates an internal image of the current state of the vehicle and its environment which serves as a basis for the driver's decisions and actions.

From this, the following requirement can be concluded: namely, that relevant situation-dependent information about the vehicle and its environment must also be perceivable by the driver. This impacts information newly accrued for the driver through the use of driver assistance systems and the need for systems which attempt to compensate for the driver's information deficits from the environment.

The process of human perception is limited by perception thresholds and the necessary application of attention. Perception thresholds vary from individual to individual, so for example age is a significant influencing factor, and they are also dependent on the environment. Since driving occurs in radically differing environments, it is important to ensure that information presented in the vehicle falls above the perception threshold and that relevant information from the environment, in case it cannot be perceived under certain conditions, is supported by technology (e.g., night vision systems which mark relevant information such as pedestrians). During driving, visual, acoustic and tactile information play a particularly large role and must be configured according to their environment. Light conditions can vary between full brightness or strong glare all the way to full darkness. Differences in acoustic environment can be equally great. There are vehicle situations without background noise or exterior sounds entering the vehicle, all the way up to conversations or loud music in the vehicle. Tactile information within the vehicle must also be adapted to possible vibrations transmitted by the vehicle or road. Particularly when configuring visual information within the vehicle, it is important to note that human beings can only sharply distinguish objects when they are mapped on the fovea up to an aperture angle of 2°. Hence it is necessary for the driver's vision to navigate far from the vehicle's external environment in order to take in complex information within the vehicle. This goes beyond simply coded signals, leading the driver to be visually distracted from the actual task of vehicle guidance.

Whether the driver perceives relevant information or not depends largely on whether the driver pays attention to the information. This application of attention is strongly influenced by the overall driver-vehicle-environment. Here the amount and type of mutually competing information from the vehicle and its environment, the mental and/or emotional preoccupation of the driver with concerns irrelevant to driving as well as the driver's personal experience all play a role. In general it has been shown that drivers perform better in paying attention to nearer objects and that changes in application of attention occur more quickly and efficiently from "far to near" than the reverse.

Acts of Judgement Acts of judgement are required in order for the driver to evaluate distances, speeds and potential critical situations.

Since judging absolute distances is difficult for humans, the driver uses various information as variables for judging longitudinal distance. Perspective speed, which is calculated by the size of the vehicle in front as well as the difference in speed and absolute distance to that vehicle, sends the driver a message about how quickly the distance to the vehicle in front is changing. Similarly, the time to collision (TTC), which takes into account absolute distance to the vehicle in front as well as difference in speed, is often identified as a relevant judgement variable for the driver. It can be assumed that TTC determines a driver's actions (Färber 1986).

The human threshold for perceiving motion while driving under ideal visual conditions lies between 3 and $10 \cdot 10^{-4}$ rad/s. However, length of observation also influences the threshold for perceiving distances and differences in speed from vehicles in front (Todosiev 1963). With a decreasing difference in speed and decreasing length of observation, the distance from which a difference in speed can be registered also decreases. In general, it has been shown that drivers tend to leave a larger safety margin than necessary at lower speeds, while they fall short of this at higher speeds.

Acoustic information can also contribute to judging distance from other vehicles; however, subjective evaluation errors can result, for example overestimating the distance of a very quiet truck or underestimating the distance of a very loud passenger car.

Anticipation of critical situations is influenced by the driver's experience with a given potential critical situation. Depending on which situations a driver has already experienced and committed to long-term memory, the driver will classify a critical situation as such based on the situation's distinguishing characteristics and react accordingly.

Decision-making and Thought Processes While fulfilling the tasks of navigation and guidance, the driver must choose an action appropriate to a given situation on the basis of decision-making and thought processes. Under the condition that humans are given sufficient time to make a necessary decision based on an external traffic situation, humans will choose more successfully than technological systems. This can be attributed to the fact that a driver has access to a more complete representation of the driving environment, although this will be less precise in particular aspects, and that with increased driving performance a driver can call on more and more experiences with identical and similar situations.

Driver reaction times lie at around 0.7 s in expected situations (e.g., approaching a vehicle), at around 1.25 s in unexpected but typical situations (e.g., braking of the car in front) and up to 1.5 s in surprising situations (Green 2000). The more critical a situation is, the faster the driver reacts. Slowness and human reaction time vary depending on driving situations and attention. Drivers react faster in heavy traffic and choose smaller distances.

Vehicle Manipulation Manipulating a vehicle in order to fulfil primary and secondary driving tasks normally does not present a problem for the driver. Control of longitudinal and horizontal movements occurs on a skill-based level for the driver, which means they are automatic processes that scarcely require any attention. This allows a driver to react quickly and flexibly to situational changes. The case is similar with secondary activities as long as they occur frequently and are accordingly well-practiced by the driver.

However, it is possible for the driver to meet excessive demand on the level of tertiary driving tasks, particularly if those functions are only seldom used, if complex control menu structures must be navigated or if the driver is confronted with infrequently occurring warning signals.

References

Abendroth B (2001) Gestaltungspotentiale für ein PKW-Abstandsregelsystem unter Berücksichtigung verschiedener Fahrertypen (Design potential for a vehicle-distance-regulation system taking into account different types of drivers). Ergonomia, Stuttgart

Brown ID, Groeger JA (1988) Risk perception and decision taking during the transition between novice and experienced driver status. Ergonomics 31(4):585–597

Bubb H (2003) Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit? (Driver assistance primarily to contribute to the comfort or safety?) VDI-Bericht 1768. VDI, Düsseldorf, pp 257–268

Cohen AS, Hirsig R et al (1991) The role of foveal vision in the process of information input. In: Gale AG (ed) Vision in vehicles – III. Elsevier, Amsterdam

Donges E (1982) Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen (Aspects of active safety conducting of passenger cars). Automob Ind 27(2):183–190

Färber B (1986) Abstandswahrnehmung und Bremsverhalten von Kraftfahrern im fließenden Verkehr (Distance perception and braking behavior of drivers in moving traffic). Z Verkehrssicherheit 32(1):9–13

Färber B (1987) Geteilte Aufmerksamkeit: Grundlagen und Anwendung im motorisierten Straßenverkehr (Divided attention: fundamentals and applications in motorized traffic). TÜV Rheinland, Köln

Fastenmeier W (1995) Die Verkehrssituation als Analyseeinheit im Verkehrssystem (The traffic situation as unit of analysis in the transport system). In: Fastenmeier W (ed) Autofahrer und Verkehrssituation. Neue Wege zur Bewertung von Sicherheit und Zuverlässigkeit moderner Straßenverkehrssysteme. TÜV Rheinland, Köln

Frieling E, Graf Hoyos C (1978) Fragebogen zur Arbeitsanalyse – FAA (work analysis questionnaire). Huber, Bern

Fuller R (1984) A conceptualization of driving behaviour as threat avoidance. Ergonomics 27(11): 1139–1155

Galinsky T, Warm J, Dember W, Weiler E, Scerbo M (1990) Sensory alternation and vigilance performance: the role of pathway inhibition. Hum Factors 32(6):717–728

Green M (2000) "How long does it take to stop?" Methodological analysis of driver perception-brake times. Transport Hum Factors 2(3):195–216

Hacker W, Iwanowa A, Richter P (1983) Tätigkeitsbewertungssystem – TBS. Handanweisung (Job evaluation system – TBS. hand instruction). Psychodiagnostisches Zentrum, Berlin

Näätänen R, Summala H (1976) Road-user behaviour and traffic accidents. North-Holland Publishing, Amsterdam/Oxford

Rasmussen J (1983) Skills, rules, and knowledge: signals, signs, and symbols, and other distinctions in human performance models. IEE Trans Syst Man Cybern SMC-13(3):257–266

Rockwell T (1972) Skills, judgment and information acquisition in driving. In: Forbes TW (ed) Human factors in highway traffic safety research. Wiley, New York

Schlick C, Bruder R, Luczak H (2010) Arbeitswissenschaft (Ergonomics and human factors). Springer, Berlin

Schmidtke H (1993) Wachsamkeitsprobleme (vigilance problems). In: Schmidtke H (ed) Ergonomie. Hanser, München/Wien

Todosiev EP (1963) The action point model of the driver-vehicle- system. PhD dissertation, Ohio State University

von Klebelsberg D (1982) Verkehrspsychologie (psychology in traffic). Springer, Berlin

Wagenaar WA (1992) Risk taking and accident causation. In: Yates JF (ed) Risk-taking behaviour. Wiley, Chichester

Wickens CD (1992) Engineering psychology and human performance. HarperCollins, New York

Wilde GJS (1982) The theory of risk homeostasis: implications for safety and health. Risk Anal 2:209–225

Driver Behavior Models

2

Edmund Donges

Contents

Abstract

The attractive power of driving motor vehicles results from the expansion of human mobility far beyond the natural physiological capabilities of people such as radius of action, power, speed, transport capacity, independence of timing and choosing routes, etc. Unfortunately, the enormous kinetic energy involved in the dynamic process of driving is coupled with an inherent danger of loss of control.

Both human factors and technical features of the road traffic system combined with their interactive compatibility decisively influence primary safety, i.e., accident prevention potential.

Science established a successful method of gaining and comprehensively representing knowledge by derivation of empirical models in terms of qualitative (descriptive) and particularly quantitative (mathematical) models. Such modeling approaches for human driver behavior in road traffic with scope on

---

E. Donges (✉)
Fürstenfeldbruck, Germany
e-mail: edmund.donges@t-online.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_2

specific capabilities and limitations are reported on, aiming at their challenges for driver assistance systems.

As focal points of the following discussion, the predictability principle (in control systems theory terminology, this could be named "anticipatory observability criterion"), the acquisition of driving skills, and the quantification of collective and individual human driving competence will be highlighted. The gaps between intrinsic human performance limitations and physico-technical constraints of vehicle, road, and traffic-environment characteristics open up the most promising realms for the development of driver assistance systems.

1 Introduction

Actively participating in road traffic as a driver of an automobile is a complex planning, monitoring, and control task. Considering today's legal status, the driver is ultimately responsible for succeeding or failing. In order to provide optimal preconditions for the driver, the engineering community has to recognize on the one hand the excellent human capabilities but on the other hand his inherent performance limitations aiming at a suitable adaptation of the technical state of the art. This is comprehensively valid especially for driver assistance systems.

The idea of summarizing knowledge about the behavior of drivers in the form of theoretical models began in the second half of the twentieth-century (Kondo 1953). Fiala (1966) was the forerunner of corresponding research in German-speaking countries. In-depth overviews of these approaches are to be found, for example, in Jürgensohn (1997) and Johannsen (2006). The following sections describe two approaches from different disciplines that have received attention in the last three decades and triggered a series of subsequent developments.

2 Three-Level Model for Target-Oriented Human Activities According to Rasmussen (1983)

First a qualitative model generally applicable for any kind of human work is explained. It originates from engineering psychology and was introduced in Rasmussen (1983). It distinguishes three categories of cognitive loads for working people. Those demand levels stretch from everyday routine situations over unexpected challenges to rare critical incidents. This three-level structure is shown in Fig. 1 (left). Initially deduced from observation of experienced personnel, it later on also proved suitable for the description of different phases of human learning behavior.

Automobile driving in literal sense belongs to the target-oriented sensorimotor human activities: moving the vehicle with passengers and cargo from a starting point to its destination by using available sensory information and directing the vehicle by motoric action on the vehicle's controls.

Complex demand situations, for which the human operator is not prepared and which require of him a previously unpracticed response, lead him to a level of

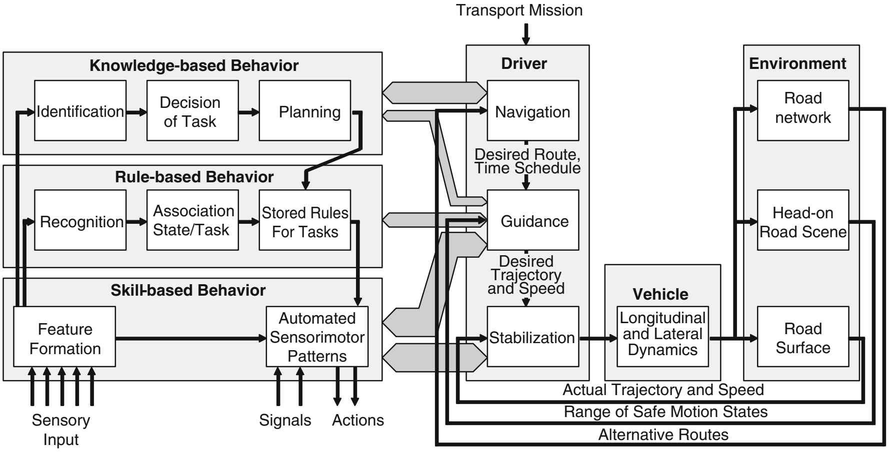


Fig. 1 Categories of human target-oriented behavior and three-level hierarchy of driving task (Donges 1992)

knowledge-based behavior. This form of behavior is essentially characterized by a mental process in which the operator searches for problem-solving action alternatives based on knowledge already present or yet-to-be acquired. If discovered, he mentally checks their suitability for managing the situation and applies the motoric response deemed most effective. Otherwise – often because of lack of time – there will be no action. Successful solutions may be stored as rules for future challenges.

The next level of **rule-based behavior** differs from knowledge-based behavior by the fact that the associated situational conditions have already occurred frequently during earlier occasions and that the human operator already possesses a repertoire of stored behavioral patterns (rules) whose probably most effective variant is chosen according to subjective experience.

The third level is named **skill-based behavior**. It is characterized by reflex-like stimulus – response mechanisms. Those have been rehearsed in a more or less long-lasting learning process and then go by in a quasi-autonomous, steady flow. Such well-established skills are the most effective forms of human behavior in terms of time and strain because both the sensory and motoric apparatus of the human operator are coordinated subcortically. Skill-based behavior is typical for routinely recurring work sequences and even leaves some leeway for secondary activities which may not necessarily be task related.

3 Three-Level Hierarchy of the Driving Task According to Donges (1982)

This general classification scheme of human target-oriented behavior deduced from a psychological approach is compared to a three-level hierarchy of the vehicle-driving task Fig. 1 (right) originating from an engineering point of view

(Donges 1982), the combination of both three-level structures first published in Donges (1992).

The **navigation** task includes the selection of a suitable travel route from the available road network, as well as an estimate of traveling time. If information about current interferences occurs such as accidents, road works, or traffic congestion, a modified route planning may be necessary. In a previously unknown traffic environment, the navigation task requires a conscious planning process and must therefore be assigned to the level of knowledge-based behavior. In case of a familiar road environment however, the navigation task may be viewed at as already accomplished. The process of navigation can be characterized as a temporally discrete activity usually initiated by distinct road markings.

The actual dynamic process of driving takes place on the two levels of the task hierarchy described by the terms **guidance** and **stabilization** (the term "stabilization" is used here instead of the usual "control" because each of the three levels can be envisaged as control tasks). Moving on a static environment and foreign objects in motion cause a continuous change of the constellation of sensory input information for the driver, in particular in the perspective representation of the three-dimensional world in the eyes of the driver (Donges 1977, 1978b). This field of view with its continuous changes contains the guidance variables as well as the deviation of the vehicle's motion state from the driver's intention.

With the navigational schedule in mind, the core task of the driver on the **guidance** level is to derive reasonable and safe desired functions like target track and target speed from the traffic scenery ahead (in control systems theory often called "forcing functions") and at the same time to intervene previewingly by an anticipatory open-loop control in order to create favorable preconditions for minimal deviations between desired and actual output functions.

On the task level of **stabilization**, the driver has to align vehicle motion with the selected guidance variables. That means he has to ensure by a closed-loop control process that occurring deviations are stabilized and compensated for at an acceptable level via appropriate corrective motoric actions.

Human driver behavior within these two task levels has been simulated in the form of continuous quantitative models on the basis of control systems theory as illustrated exemplary in the next section.

The extent to which the subtasks of guidance and stabilization are performed in the various behavioral categories specified in Rasmussen (1983) heavily depends on the experience of the individual driver and on the frequency with which the driver has already encountered the particular traffic situation. The novice driver will at first mainly practice his driving activity in the knowledge-based behavioral level. With increasing routine, he will thereafter gradually build up a repertoire of behavioral rules and in the long run develop the ability to employ skills unconsciously. As soon as the appropriate experience has been gained, participation in undisturbed road traffic becomes a day-to-day routine which can be carried out almost entirely on the level of skill-based behavior. An estimation of the duration of this learning process can be derived from the accident involvement of novice drivers: therefore it takes about 7 years or 100,000 km mileage

(Anonymous 1976; Willmes-Lenz 2003) until a driver reaches the completion of his individual capacity for skill-based behavior.

Only the unexpected emergence of critical conditions disturbs the driver's subcortical equilibrium and forces him onto the more demanding levels of either rule-based behavior if his associative memory contains an appropriate sample of response patterns to the situation at hand or even knowledge-based behavior if the triggering event is unfamiliar and of a complex nature. In road traffic, the level of knowledge-based behavior must always be classified as critical or likely to cause an accident whenever the speed and distance to the danger zone no longer allows enough time to weigh up alternatives. Therefore Förster (1987) claims: "In road traffic the need for knowledge-based action has to be minimized!"

As the foregoing considerations show, the guidance level of the driving task takes on enormous importance with regard to the safety of the driving process, because it is on this level that a decision is taken as to whether the driver can draw the necessary conclusions from the sensory input information in time and whether the guidance variables selected by the driver lie in the objectively safe or unsafe zone.

For this level of the task hierarchy, the human driver is gifted with the outstanding ability of preview (anticipatory perception) of the traffic environment that enables him to apply open-loop control inputs (feed forward) in order to proactively come close to the desired motion state and to compensate for system-inherent time delays – as has been demonstrated empirically in Donges (1977, 1978b).

On the task level of stabilization, the driver and vehicle constitute the well-known closed-loop control system which accounts for the dynamic stability of the ongoing process. An experienced driver fulfills this task mainly on the skill-based behavioral level.

The preceding considerations about the intensity of interaction between the behavior and task levels of the two models are indicated by the thickness of the gray-backed connection arrows in Fig. 1.

The significance of the combination of the two model approaches in Fig. 1 expands substantially beyond engineering aspects of the vehicle-road system. For example, they stimulate new methods for driver education programs currently under development in traffic psychology (Bahr and Sturzbecher 2013; TÜV DEKRA arge tp 21 2011).

4 Example of a Control Systems Theory-Based Model for the Guidance and Stabilization Levels of the Driving Task

For the simulation of driver behavior in road traffic, a series of model approaches on the basis of control systems theory have been developed such as Kondo (1953), Fiala (1966), Weir and McRuer (1970), and Mitschke and Niemann (1974). The special capabilities of this method allow identification of causal correlations between input and output functions, without knowing exactly the internal structure of the human information perception, its mental processing, and its transposition into motoric actions. A simplification like this obviously underlies the restriction that it captures only those phenomena that are observable by input and output functions. In state-space representation of control systems theory, three criteria are functional prerequisites: **observability**, **controllability**, and **stability. Observability** refers to the knowledgeable part of the entire system by a limited number of measureable state variables. The resulting model is inevitably incomplete, but nevertheless it creates important quantitative insight into the transfer characteristics of human control behavior in terms of amplitude and time. It clearly shows human adaptability as well as its limitations.

The earliest approach of a driver model originates from Japan (Kondo 1953) (quoted in Jürgensohn (1997)) and describes the driver's steering response under side-wind disturbances. It already presents a principle for the simulation of the human anticipatory perception capability for the driving environment ahead in the simple form of a preview distance. This means that the driver extrapolates the lateral deviation from the desired path to a point at the preview distance and thus compensates for an estimated future control error. In German language this approach has been called "Deichselmodell," i.e., "carriage pole model."

In contrast, the driver model invented in Donges (1977) (short versions in Donges (1978a, b)) separates the two task levels of guidance and stabilization in two sub-models: it presents the guidance level in a form of an "anticipatory open-loop control" and the stabilization level as a "compensatory closed-loop control" in Fig. 2. Because both sub-models will not completely reproduce the driver's action, a "remnant" quantity is a third integral part of the model. The remnant comprises shortcomings of the two sub-models as well as driver-induced signals which are not task related.

This driver model originally describes the lateral dynamics share of the driving task only, but its basic structure is equally suitable for the simulation of longitudinal dynamics. The empirical database for this model comes from a series of driving simulator experiments. The simulated test course was a winding closed-circuit two-lane road located in a horizontal plane without other traffic. In order to define unique forcing functions, all test drivers were firmly instructed to drive along the centerline of the circuit and to follow a preset speed profile as accurately as possible.

Basics for modeling the derivation of desired trajectories and speed profiles according to the target imagination of a driver were invented much later, e.g., Prokop (2001), by minimizing cost criteria built of weighted control errors and constraining overriding of safe motion space.

Here the single input variable for the sub-model "anticipatory open-loop control" is the curvature (i.e., reciprocal of radius) of the prescribed desired trajectory (road centerline). The previewing feed-forward control action starts at a certain time interval, called "motoric anticipation time T_A" prior to changes of desired path curvature and is then transferred by a gain factor and a smoothing lag time constant.

To avoid misunderstanding it should be mentioned that the total anticipation time comprises additional time slices for arousal, perception, mental processing,

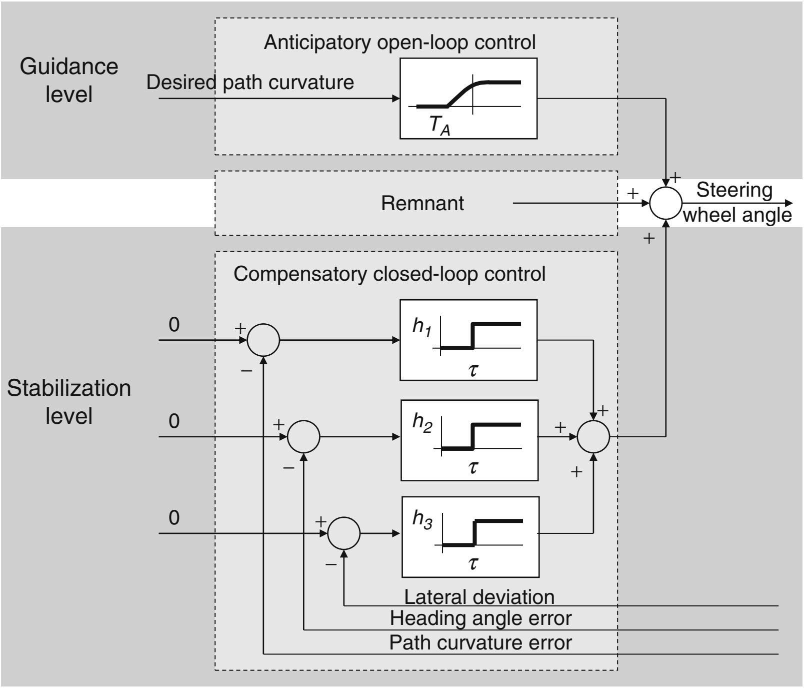


Fig. 2 Block diagram of a two-level model for driver steering behavior (Donges 1978a, b)

and muscular innervation, before a motoric reaction on a foreboding event can be measured on the corresponding control element.

The "compensatory closed-loop control" sub-model receives three actual error state variables simultaneously: path curvature error (difference between curvatures of desired and actual path), heading angle error (angle between tangents of desired and actual path), and lateral deviation. Each of them is fed back via structurally identical transfer functions: a time delay τ (closed-loop reaction time) – the same for the three error states – and individual gain factors.

As has been mentioned before, the driver perceives the input variables for both sub-models from static and moving patterns in the perspective view of the traffic environment ahead (Donges 1977, 1978b).

The model parameters determined from the database of the driving simulator experiments show the following properties (Donges 1978a, b):

For the **anticipatory control sub-model**:

- The gain factor comes very close to the reciprocal of the vehicle's gain factor. This is a must because desired and real curvatures in stationary state have to be close to each other because otherwise the vehicle would leave the road.
- The identified motoric anticipation times of preview steering are in the order of magnitude of 1 s largely regardless of the test conditions. In terms of the earliest

driver model approach mentioned above (Kondo 1953), this means a preview distance or “carriage pole length” growing proportional with vehicle speed.

- The time lag constant decreases with increasing vehicle speed, i.e., the slope of the anticipatory steering action is faster the higher the vehicle speed.

For the **compensatory control sub-model**:

- The feedback loop of heading angle error contributes by far the highest share to compensatory steering angle, i.e., heading angle error can be seen as the main control variable, path curvature error as a D-portion, and lateral deviation as an I-portion of a classical PID controller (PID is an abbreviation for **p**roportional, **i**ntegral, and **d**ifferential).
- The gain factor of path curvature error increases significantly with vehicle speed, i.e., the lead portion of compensatory steering angle grows correspondingly with vehicle speed.
- The driver’s time delay decreases significantly with growing vehicle speeds.

Higher vehicle speeds require less time-shift in the response of the driver in order to satisfy the stability criterion. This necessity can also be explained by the so-called crossover frequency model (McRuer and Krendel 1962): the stability margin of vehicle lateral dynamics diminishes with increasing vehicle speed. This effect must be compensated for by increasing the differential portion and by shortening delay times in driver response in order to maintain sufficient stability margins of the overall driver-vehicle control system.

5 Time Criteria

The correlations between the time behavior of the driver and the driving speed as just described are representative examples of the human adaptation capabilities according to the particular circumstances. The mean values of the motoric action anticipation times of 1 s and the closed-loop time delays of 0.5 s identified from the driving simulator measurements are initial clues for the following considerations about human timing behavior.

Figure 3 conveys an overview of the time horizons characterizing the three levels of the driving task. The typical time horizon of the navigation level extends from the possible duration of a complete ride in the order of several hours to the range of minutes if sudden announcements request for route changes, e.g., by means of traffic signs. Accordingly, today’s navigation systems start early with a first indication that is repeated and substantiated while approaching the decisive point.

In case of favorable visibility conditions, the driver changes from navigation to guidance task as soon as his eyes can perceive the corresponding road and traffic scene ahead. He derives the desired target variables and initiates anticipatory control actions. Speed corrections by releasing accelerator pedal or pressing brake pedal usually demand longer anticipation times than steering.

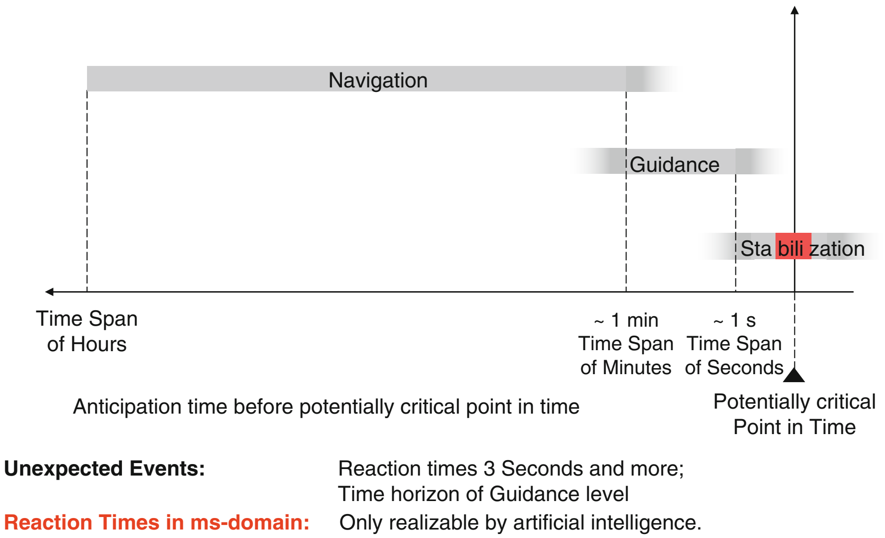


Fig. 3 Time horizons typical for the task levels of navigation, guidance, and stabilization (Braess and Donges 2006)

In view of motoric action anticipation times in the order of 1 s, the process of traffic-environment perception must start distinctly earlier, especially in case of unexpected events. An approximate minimum value for the time needed between perceptive arousal and muscular innervation can be estimated from the results of simple reaction time measurements under controlled laboratory conditions: single-channel display, single-channel control switch, hand on switch, no distraction from outside, and test person fully concentrated. The reported reaction times show mean values of roughly 2 s. That is, information and warning systems in the complex environment of road traffic should enable total anticipation times longer than about 3 s.

The results reported in Wakasugi (2005) – measured under the enhanced arousal of controlled test conditions – at least support this requirement: in case of decisions for lane change maneuvers, warning signals must be given no later than 2 s before.

If total anticipation times of more than about 3 s cannot be achieved (e.g., because of the limited range of environment sensors), only a spontaneously stimulated reaction by an intuitive action recommendation device may be helpful, for example, in the form of a haptic stimulus as by active force feed-in via accelerator pedal or steering wheel.

Typical motoric reactions for the compensation of closed-loop control errors on the task level of stabilization are carried out with time delays of some 100 mins. These fast closed-loop delays represent the lowest time limit of human response because the skill-based behavioral level is least time-consuming. Therefore, any requirement for reaction times in the ms domain can only be fulfilled by technical control systems as, for example, by technical means such as driving dynamics control systems via braking and steering (red zone in Fig. 3).

As mentioned previously, response times to unforeseeable events range from a lowest limit of about 3 s up to significantly higher values largely depending on the complexity of the situation. The importance of short reaction times for accident avoidance has been pointed out by Enke (1979): a reduction of the drivers' reaction times between half a second and one second would help to avoid around half of all collision accidents. Abbreviations of driver reaction times in this order of magnitude seem achievable only by augmenting anticipatory action, i.e., by empowering the guidance level of the driving task by artificial intelligence.

6 A New Approach for the Quantification of Skill-, Rule-, and Knowledge-Based Behavior in Road Traffic

The three-level hierarchy of human target-oriented behavior of Rasmussen, as described in Sect. 2, is a qualitative model originally. In Braess and Donges (2006), a new, admittedly bold approach to a quantification of the terms skill-, rule-, and knowledge-based behavior in road traffic has been proposed. This proposal was intensified by initiating empirical acquisition of driving behavior data collectives (Wegscheider and Prokop 2005), which up to then were rather rare in German engineering literature (Burckhardt 1977; Hackenberg and Heißing 1982).

The g-g-diagram (longitudinal versus lateral accelerations) in Fig. 4 from Wegscheider and Prokop (2005) serves as an example to explain this approach.

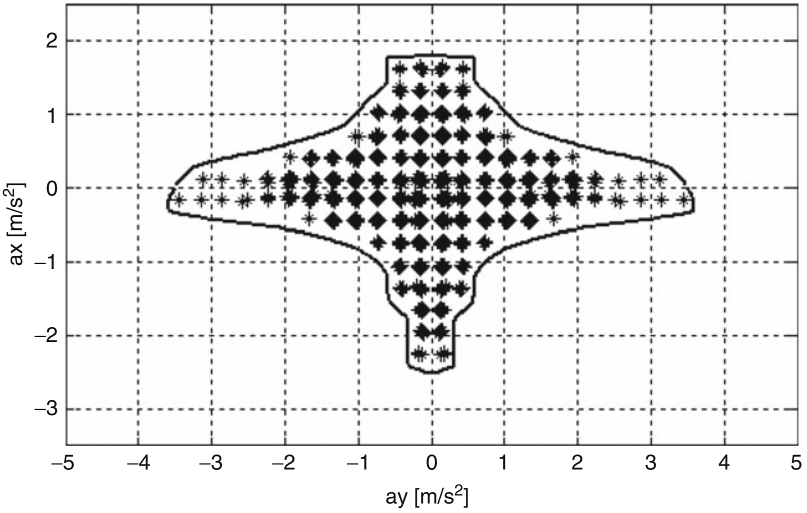


Fig. 4 g-g-diagram of normally experienced drivers (longitudinal versus lateral acceleration) (Wegscheider and Prokop 2005)

This diagram shows the range of longitudinal versus lateral accelerations that twelve drivers with average experience – labeled by the term "normal" – applied during test runs on winding country roads and highways in everyday traffic. The circumferential line represents an 85-percentile line, i.e., all drivers remain below this outline in 85 % of the test driving time.

The outline itself shows the basic contour of a cross with its corners somehow rounded off. This means that the group of subject drivers is limited in applying combined steering and braking or steering and accelerating maneuvers simultaneously but prefers singular actions influencing either lateral or longitudinal dynamics. This indicates the human preference of single-channel control and human limitations in multichannel motoric action. These observations are corroborated by earlier measurements of drivers' behavior collectives in Burckhardt (1977) and Hackenberg and Heißing (1982).

Now imagine that in a similar way the statistical frequency distribution and their percentile outlines of the driving behavioral collective are registered and stored for an individual driver, and this is done not only in two dimensions for longitudinal and lateral accelerations but also for other relevant state variables (such as speed, speed difference and reciprocal of distance to a leading vehicle ahead, steering wheel angle and rate, etc.) in a multidimensional state-variable histogram memory characterizing the behavior of this individual driver. In this way a more or less comprehensive image of the personalized experience horizon will be gained, and thus a personalized objective quantification of the traffic competence of the corresponding individual driver can be determined. One could speak of a personal driving competence "fingerprint." This opens up a new realm for the development of driver assistance systems that seems to have been partially neglected in the past.

Based on the results in Fig. 4, the following pragmatic approach for a quantification of the three behavioral levels is proposed: the skill-based area includes the 80-percentile outline of the longitudinal and lateral acceleration collective of an individual driver, the rule-based range reaches up to the 95th percentile, and all driving states beyond this 95-percentile outline can be looked at as rare events and belong to the knowledge-based behavioral level *(the numeric percentile values of 80 and 95 are arbitrary proposals that have to be certified by profound empirical research).*

The individual experience horizon will differ depending on the drivers' personalities. It will enclose a rather small outline for a careful and cautious driver and expand up to technical limits for an ambitious, sporty driver, Fig. 5. The driving style of individual drivers may also vary according to their mental disposition in a span from defensive (within their 80-percentile outline) over offensive (within their 95-percentile outline) up to aggressive (transgressing their 95-percentile outline) behavior as related to their personal traffic competence.

The empirical results at hand show consistently that drivers' behavior collectives in traffic on public roads usually remain well below tire adhesion limits (Kamm's limit circle) as long as the pavement is dry. However, if weather conditions change, wet, snowy, or icy road surfaces will substantially reduce adhesion limits. Thus it may happen that even the narrow behavioral collective of a cautious driver will

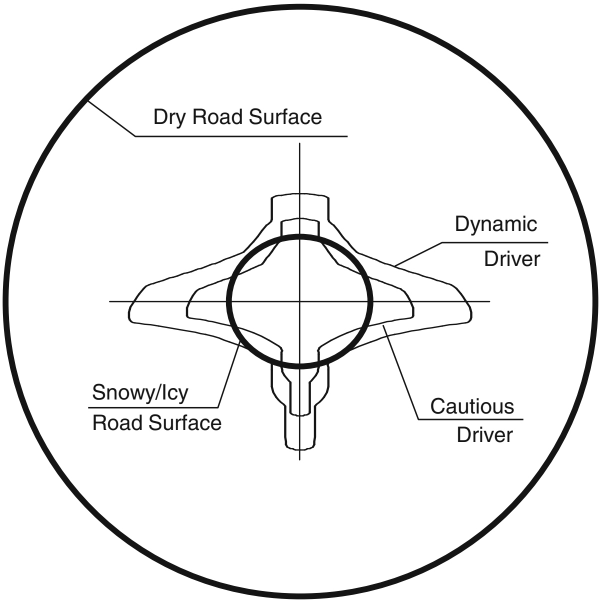


Fig. 5 Driving behavior collectives compared to adhesion limit circles (differently skilled drivers, road surfaces with high and low friction potentials) (Braess and Donges 2006)

exceed the limit of adhesion, Fig. 5. Hence the risk of accidents will grow considerably under these circumstances.

On the basis of these findings, the necessity of a bifacial view shall be highlighted again: in addition to the technical improvements of motor vehicles and other engineering features of the road traffic system, there is a second important influence factor on road traffic accident prevention, i.e., the individual driver's traffic competence. Unfortunately, it seems that up to now this topic has found less attention than the engineering efforts. Many statistical investigations on road traffic accident causation suggest a higher share of human rather than engineering factors – a controversially discussed statement. Very often the reason might be the lack of knowledge on both sides causing a lack of adaption between human and technical requirements.

7 Conclusions for Driver Assistance Systems

The driver behavior models and their corresponding quantifications previously described originate from the challenging field of driver-vehicle motion dynamics (vehicle "handling"). Although this is only a restricted view on the much more complex road traffic system, the findings can be looked at as examples of the recognition of specific capabilities and limitations of both the human driver and the surrounding technical components. Driver assistance systems demand the art of integrating the know-how of experimental psychology with the continuously increasing power of technology and the necessity for a successful adaptation between both faculties.

In conclusion, a few aspects of the findings should be emphasized:

On the **navigation** level:

While traveling, the navigation task aims at initiating and raising the consciousness of an approaching change of circumstances or a potential danger within a time span of a minute or more ahead. The database of navigation system maps could be useful for the driver in order to proactively prepare for maintaining safety margins toward either technical boundaries or the limitations of his individual driving competence.

On the **guidance** level:

As much as ever, there is strong evidence that the task level of guidance could be the most effective contribution to support human drivers' performance. Since around 1990 high R&D effort has been invested into technical sensors for the recognition of the motor vehicles' environment and road infrastructure. This may assist the driver to complete his knowledge about the surrounding situation and to transfer information that the driver might have missed because of inattentiveness, or mental overload, or obstacles hiding relevant cues, or even facts, which are inaccessible for human senses. An important aspect shall be emphasized again: any prolongation of driver anticipation time will lift up the probability to solve an upcoming traffic conflict by deriving safe guidance variables and performing proactive control inputs.

A crucial point is how to present complex information like warnings or action recommendations in a self-explaining and – in any case – non-disturbing display by visual, acoustical, or haptic/kinesthetic means. Promising methods for innately understandable information presentation and habitual or reflex-like response automatisms are still to be discovered.

On the **stabilization** level:

The most efficient form of satisfying the stability criterion in the compensatory driver-vehicle control loop is characterized by reflex-like stimulus – response mechanisms that can only be reached on the skill-based behavioral level after a long-lasting learning process. This level of human performance can only be applied as long as sufficient safety margins with respect to limitations of both human and technical capabilities are kept up.

Two observations from above may serve as indicators of human limitations:

- Only on the skill-based behavioral level are drivers able to react with delay times as low as about half a second. Requirements of shorter response times exaggerate human reaction capabilities.
- The example of the cross-shaped driver behavior outlined in Fig. 4 presumes human drivers' preference for single-channel actions. Simultaneous multichannel motoric output and multichannel sensory input risk overtaxing the human driver.

On the other side, here we find the special capabilities of artificial intelligence: cycle times in the range of milliseconds or less, multichannel measurements by means of sensor fusion and state estimation algorithms, and multichannel controls by a set of actuators. Even relevant information that is not accessible for the human sensorium may be acquired by technical means.

The efficiency of these technologies has been shown impressively by chassis control systems and their simultaneous observation and control of individual wheel speeds, steering angles, and brake or acceleration forces on each wheel (Donges 1995, 1998). Such control systems have contributed already and will help further on to suppress the effects of nonlinear characteristics or cross-couplings between vertical, longitudinal, and lateral vehicle dynamics that are unpredictable even for skilled drivers (Donges 1992), i.e., they might be forced from the level of skill-based behavior onto the more time-consuming levels of rule- or even knowledge-based behavior.

The current and future development of driver assistance systems could be the true connecting link for the further improvement of ease and safety in the dynamic interaction between human and technological capabilities and limitations in road traffic.

References

Anonymous (1976) Unfalldisposition und Fahrpraxis (Accident proneness and driver's experience). Automob Z 78:129

Bahr M, Sturzbecher D (2013) Bewertungsgrundlagen zur Beurteilung der Fahrbefähigung bei der praktischen Fahrerlaubnisprüfung (Evaluation basics for judgement of driving competence in driver‘s license examination). In: Winner H, Bruder R (eds) Maßstäbe des sicheren Fahrens. Darmstädter Kolloquium „Mensch + Fahrzeug. Technische Universität Darmstadt, Ergonomia Verlag, Stuttgart

Braess HH, Donges E (2006) Technologien zur aktiven Sicherheit von Personenkraftwagen – „Konsumierbare“ oder echte Verbesserungen (Technologies for active (primary) safety – „consumable“ or true improvements)?, 2. Tagung „Aktive Sicherheit durch Fahrerassistenz“ 4.–5. April 2006 TU München, Garching bei München

Burckhardt M (1977) Fahrer, Fahrzeug, Verkehrsfluss und Verkehrssicherheit – Folgerungen aus den Bewegungsgesetzen für Fahrzeug, Straße und Fahrer (Driver, vehicle, traffic flow and safety – conclusions from vehicle-motion-dynamics laws for vehicle, road and driver. In: Interfakultative Zusammenarbeit bei der Aufklärung von Verkehrsunfällen, vol XXX. AFO, Köln

Donges E (1977) Experimentelle Untersuchung und regelungstechnische Modellierung des Lenkverhaltens von Kraftfahrern bei simulierter Straßenfahrt (Experimental investigation and control-system-theory-based modeling of driver's steering behavior in simulated road driving). Dissertation TH Darmstadt

Donges E (1978a) Ein regelungstechnisches Zwei-Ebenen-Modell des menschlichen Lenkverhaltens im Kraftfahrzeug (A control-system-theory-based two-level model of human steering behavior in motor vehicles). Zeitschrift für Verkehrssicherheit 24:98–112

Donges E (1978b) A two-level model of driver steering behavior. Hum Factors 20(6):691–707

Donges E (1982) Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen (Aspects of active (primary) safety in motor-vehicle guidance). Automob Ind 27:183–190

Donges E (1992) Das Prinzip Vorhersehbarkeit als Auslegungskonzept für Maßnahmen zur Aktiven Sicherheit im Straßenverkehr (The predictability concept – an integrated approach for improving active (primary) safety in road traffic). VDI Berichte 948 (Anlage), 1. Fachtagung „Das Mensch-Maschine-System im Verkehr“ 19.-20. März 1992, Berlin

Donges E (1995) Supporting drivers by chassis control systems. In: Pauwelussen JP, Pacejka HB (eds) Smart vehicles. Swets & Zeitlinger B.V, Lisse, pp 276–296

Donges (1998) A conceptual framework for active safety in road traffic. In: Proceedings of 4th international symposium on advanced vehicle control AVEC’98, SAE of Japan, Nagoya

Enke K (1979) Possibilities for improving safety within the driver vehicle-environment control loop. ESV-Conference, Washington, pp 789–802

Fiala E (1966) Lenken von Kraftfahrzeugen als kybernetische Aufgabe (Steering of motor vehicles as a cybernetic task). Automob Z 68:156–162

Förster HJ (1987) Menschliches Verhalten, eine vergessene Ingenieur-Wissenschaft (Human behavior, a forgotten engineering science)? Abschiedsvorlesung Universität Karlsruhe, Januar 1987

Hackenberg U, Heißing B (1982) Die fahrdynamischen Leistungen des Fahrer-Fahrzeug-Systems im Straßenverkehr (Driving dynamics performance measures of driver-vehicle-system in road traffic). Automob Z 84:341–345

Johannsen G (2006) Fahrzeugführung und Assistenzsysteme (Vehicle guidance and assistance systems). In: Zimolong B, Konradt U (eds) Ingenieurpsychologie – Enzyklopädie der Psychologie. Hogrefe, Göttingen

Jürgensohn T (1997) Hybride Fahrermodelle (Hybrid driver models). Pro Universitate Verlag, Sinzheim

Kondo M (1953) Directional stability (when steering is added). J Soc Automot EngJpn (JSAE) Tokyo, Jidoshagiutsu, 7(5,6): 104–106, 109, 123, 136–140 (in Japanese, cited by (Jürgensohn 1997))

McRuer DT, Krendel ES (1962) The man-machine system concept. Proc IRE 50:1117–1123

Mitschke M, Niemann K (1974) Regelkreis Fahrer-Fahrzeug bei Störung durch schiefziehende Bremsen (Driver-vehicle control-loop under disturbance by left-right unsymmetric brake forces). Automob Z 76:67–72

Prokop G (2001) Modeling human vehicle driving by model predictive online optimization. Veh Syst Dyn 11(1):1–35

Rasmussen J (1983) Skills, rules and knowledge; signals, signs and symbols and other distinctions in human performance models. IEEE Trans Syst Man Cybern SMC 13(3):257–266

TÜV DEKRA arge tp 21 (2011) Innovationsbericht zur Optimierung der Theoretischen Fahrerlaubnisprüfung – Berichtszeitraum 2009/2010 (Innovation report for the optimization of the theoretical driving test). Dresden

Wakasugi T (2005) A study on warning timing for lane change decision and systems based on driver’s lane change maneuver. ESV-Conference, paper 05–0290

Wegscheider M, Prokop G (2005) Modellbasierte Komfortbewertung von Fahrer-Assistenzsystemen (Model-based evaluation of ease of use of driver assistance systems). VDI-Berichte Nr. 1900: 17–36

Weir DH, McRuer DT (1970) Dynamics of driver steering control. Automatica 6:87–98

Willmes-Lenz G (2003) Internationale Erfahrungen mit neuen Ansätzen zur Absenkung des Unfallrisikos junger Fahrer und Fahranfänger (International experience with new approaches for lowering the accident risk of young and novice drivers). Berichte der Bundesanstalt für Straßenwesen, Bergisch Gladbach, Heft M 144

Framework Conditions for the Development of Driver Assistance Systems

3

Tom Michael Gasser, Andre Seeck, and Bryant Walker Smith

Contents

T.M. Gasser (✉) • A. Seeck
Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany
e-mail: gasser@bast.de; seeck@bast.de

B.W. Smith
Stanford Law School, Center for Internet and Society, Stanford, CA, USA
e-mail: bryantws@civil.co.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_3

Abstract

The term driver assistance systems in the chapter title shall be understood to include vehicle automation. This chapter starts with a homogeneous and consistent classification and nomenclature of all kinds of driver assistance systems known and under discussion today (including vehicle automation). It thereby builds upon familiar classification schemes by the German Federal Highway Research Institute (BASt) and the standardization body SAE international. Detailed evaluation of the German legal situation for driver assistance systems and vehicle automation is provided in the following Sect. 2.

In Sect. 3, an overview is given on the legal system in the US to reveal aspects relevant for vehicle automation. This is intended as initial information for those not acquainted to the US legal system which has been the first to regulate automation in several federal states.

Finally, in Sect. 4, the current rating scheme of the European New Car Assessment Programme (EuroNCAP) is presented in comparison to legal instruments. The model of a consumer protection based approach proves to be a flexible instrument with great advantages in promoting new technologies. Technical vehicle regulations on the other hand rule minimum requirements. Both approaches are needed to achieve maximum vehicle safety.

The term driver assistance systems in the chapter title as well as in the present Handbook shall be understood to include vehicle automation. In order to promote a homogeneous understanding, this chapter proposes a classification of systems from the viewpoint of their effect on vehicle guidance. This classifies and describes the nomenclature of automation levels developed by the BASt project group “Legal consequences of increasing vehicle automation” (Gasser et al. 2012). The relevant legal framework, especially the regulatory and liability laws according to German legislation, is then presented, and the significance of the different categories is explained. The subsequent overview describes the current state of legislation in certain federal states of the USA (as of early 2014), which for the most part allows the use of automated vehicles at least for research, development and testing purposes. Next, the overall framework of consumer protection in Europe is explained. The rating scheme created within the scope of Euro NCAP also increasingly takes into account driver assistance systems for assessment of vehicle safety, and constantly develops the requirements.

1 Classification and Nomenclature of the Systems

The literal meaning of the term “driver assistance system” already suggests interaction of a system with the human driver. To be more general and specific, it is basically necessary to look into various forms of task division between the human and the automatic system, see (Kraiss 1998). Furthermore, a distinction is drawn between “autonomous assistance” and “telematics” (Albrecht 2005), or there is a further distinction made between conventional driver assistance systems and those with environment sensing by machine (Maurer 2009). In individual cases, such further distinctions may be of importance for the technical understanding of the system or relevant for the legal assessment (Albrecht 2005); they do, however, point to elements beyond the vehicle control itself. It is proposed to concentrate exclusively on the division of tasks between the human and the machine for a comprehensive understanding of driver assistance systems and vehicle automation.

In terms of vehicle guidance, the different modes of system operation can be described and distinguished in a way relevant for legal classification:

Operation mode A: Informing functions:

Informing functions only affect the vehicle guidance "indirectly," namely via the driver. They regularly perform the task of environment perception in vehicle guidance as is simultaneously performed by the driver. The information is made available to the driver via the so-called human-machine interface.

Operation mode B: Continuously automated functions:

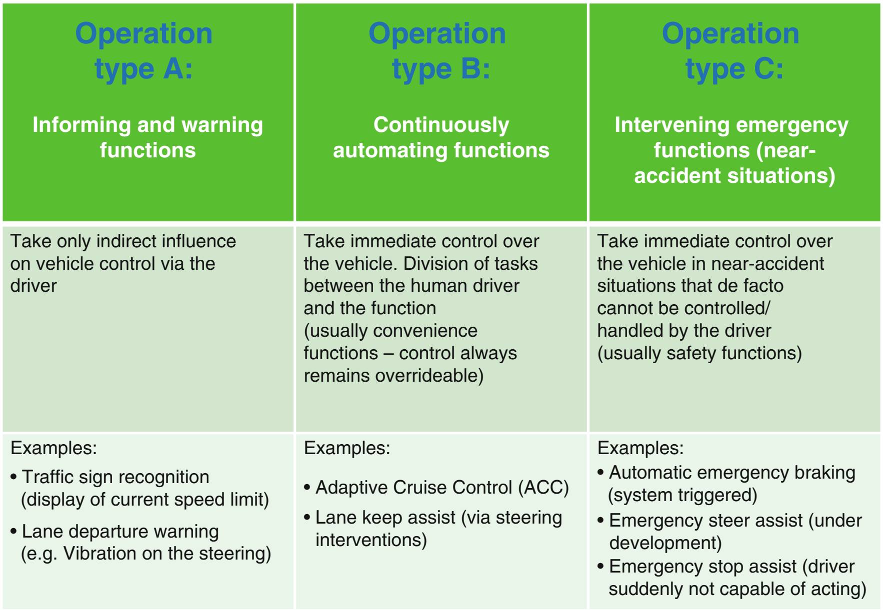
Continuously automated functions are characterized as directly intervening. They directly influence the vehicle control over long periods or parts of the journey. Assisting functions are described as performing a "redundant-parallel" division of tasks between man and machine (Maurer 2009). Both, the status of the system and the interventions in vehicle control are also regularly provided as information via a human-machine interface in order to improve interaction between man and machine; the vehicle control and any change is, however, also immediately perceptible for the driver, partly via a feedback from the controls (e.g., on the steering wheel), but at the same time also by the driving dynamics perceptible by the driver.

Operation mode C: Intervening emergency functions:

From the standpoint of the classification proposed here, the fundamental idea of parallel performance of duties by man and machine in the case of intervening systems should not be left aside. With intervening emergency systems however, this simultaneous performance of duties is indeed not, or no longer, completely achieved. This is crucial for classification. In sudden emergencies, to be described as near-collision situations, the human can only react with a delay (so-called reaction time). During these periods, the intervening systems, here more specifically referred to as emergency functions, are superior to drivers, and their action is temporarily not subject to human control. This special quality of intervening emergency functions justifies their classification as a distinct class in the present categorization. A characteristic example is emergency brake assistance.

The sudden inability of the driver to act (e.g., due to a pathological state) equally leads to the absence of human control, and justifies the inclusion of these systems in operation mode C. Including functions that take effect in such situations (e.g., the so-called emergency stop assistance) is justified in spite of the fact that situations might not yet be as close to a collision. However, they can be described as potentially accident-prone. It is characteristic that in these cases a (significant) reaction can no longer be expected from the driver.

These three higher level categories are summarized in the table (Fig. 1).



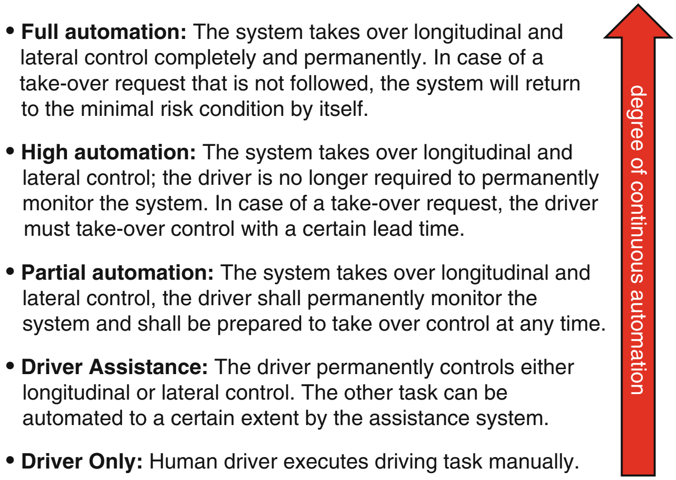
Operation type A: Informing and warning functions	Operation type B: Continuously automating functions	Operation type C: Intervening emergency functions (near-accident situations)
Take only indirect influence on vehicle control via the driver	Take immediate control over the vehicle. Division of tasks between the human driver and the function (usually convenience functions – control always remains overrideable)	Take immediate control over the vehicle in near-accident situations that de facto cannot be controlled/ handled by the driver (usually safety functions)
Examples: • Traffic sign recognition (display of current speed limit) • Lane departure warning (e.g. Vibration on the steering)	Examples: • Adaptive Cruise Control (ACC) • Lane keep assist (via steering interventions)	Examples: • Automatic emergency braking (system triggered) • Emergency steer assist (under development) • Emergency stop assist (driver suddenly not capable of acting)

Fig. 1 Summary of three top level operation modes

The (second) operation mode B presented here, the continuously automating functions, has so far been subjected to in-depth examination and classification. In the context and for the purpose of carrying out the BASt Project Group “Legal consequences of increasing vehicle automation” (Gasser et al. 2012), four different automation levels (plus level zero in the absence of continuous automation) have been described. This classification only concerns the abovementioned operation mode B of continuously automated systems. The illustration, Fig. 2, presents the nomenclature jointly developed by the Project Group in a simplified version.

The continuous increase in the degree of automation has been described within the scope of the BASt Project Group, taking into account the respective changes in the driver’s task. The classification has thus been done under the aspect of task distribution between the driver and the system in respect of vehicle control. Therefore, the division of tasks between the human driver and the machine is again the underlying basic principle for classification (as already suggested above for the top level categorization into the three operation modes). For continuously automating systems (operation mode B), this is now broken down further into the following levels or automation degrees:

Level 0: It begins with the absence of any automation at the lowest level of the “driver only”; here, the driver alone is in charge of the longitudinal and lateral vehicle control.



- **Full automation:** The system takes over longitudinal and lateral control completely and permanently. In case of a take-over request that is not followed, the system will return to the minimal risk condition by itself.
- **High automation:** The system takes over longitudinal and lateral control; the driver is no longer required to permanently monitor the system. In case of a take-over request, the driver must take-over control with a certain lead time.
- **Partial automation:** The system takes over longitudinal and lateral control, the driver shall permanently monitor the system and shall be prepared to take over control at any time.
- **Driver Assistance:** The driver permanently controls either longitudinal or lateral control. The other task can be automated to a certain extent by the assistance system.
- **Driver Only:** Human driver executes driving task manually.

Fig. 2 Simplified nomenclature of continuous vehicle automation by the BASt Project Group

Level 1: Driver assistance as the lowest degree of automation covers systems that continuously automate either longitudinal or lateral vehicle control. The driver is in charge of the remaining control task, whilst the automated task is executed by the function within the technical limits of the respective system.

Level 2: Partial automation is the lowest of the three automation levels explicitly named "automation": Here, the system is in charge of both longitudinal and lateral vehicle control for certain periods of time or situations. The driver's task is to continue monitoring the surrounding traffic in order to be ready for an immediate takeover of vehicle control at all times (e.g., in the form of corrective interventions, or complete takeover of control). In the highest stage of partial automation, the driver is therefore no longer required to control the vehicle actively, but only to monitor, mentally control, and to immediately correct or take over active control in case the system reaches its limits or in the event of a malfunction.

Level 3: In the case of the next higher level of automation, which the BASt-project group referred to as high automation (corresponding SAE-wording: "conditional"), the driver no longer needs to monitor the surrounding traffic and the vehicle continuously and should be prepared to resume vehicle control in case of a takeover prompted by the system following a relatively short lead time (that has yet to be determined). Level 3 systems detect all system limits; but may be unable to reach the safe or risk-minimal state from every initial condition imaginable. Driver takeover of control is indispensable in these cases. The drivers' possibility to regain control over the vehicle immediately and at any time remains untouched.

Level 4: The last degree of automation described by the BASt Project Group is full automation (corresponding to SAE-wording "high automation"). With this level of automation, the driver also no longer needs to monitor the execution of the

> driving task. The characteristic feature here, in contrast to level 3 is, that the system will automatically return to the risk-minimal state from any initial situation, even in case the driver does not respond to a takeover request. Even in the case of level 4, the driver is still able to take over vehicle control immediately and at any time.

The nomenclature with respect to an even higher degree of automation consciously remained undone at the time of the classification. The underlying objective was to describe only those levels of automation that were considered realistically foreseeable at the time of specification. In the ongoing discussion, a further degree (level 5) is being considered. The technical independence is then even higher. This level of the machine's autonomy might amount to "driverless" vehicles.

It should be further noted that currently a common understanding can largely be expected from the international development in the field of classification. This includes in particular the "Preliminary Statement of Policy Concerning Automated Vehicles" (NHTSA 2013) published by the US National Highway Traffic Safety Administration "(NHTSA), and the Standard J3016 "Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems" (SAE 2014) published in January 2014 by the SAE International, more specifically the "On-Road Automated Vehicle Standards Committee," a standardization body. The gradation of automation levels in the said steps or levels is used therein; however, it should be noted that in comparison to (Gasser et al. 2012) different terms are used for the levels 3 and 4, and a fifth level (or degree) is described. For orientation on the international level, it is therefore recommended to take into account the numbering of levels as described above, since the underlying description is congruent in contrast to the use of terms (this can avoid misunderstandings).

2 Legal Framework and Evaluation

Regarding the legal framework, the BASt Project Group "Legal consequences of an increase in vehicle automation" (Gasser et al. 2012) has so far dealt with functions of operation mode B in legal terms under German law, and has pointed out some important legal consequences (without claiming to be exhaustive). Also systems of intervening emergency stop assistance have been discussed there (so far, however, partly in opposition to the understanding proposed here that these systems should indeed be assigned to operation mode C). Hereinafter, an attempt at a comprehensive description of the current state of knowledge across all categories is made. The legal assessment is thereby carried out by focusing on aspects of regulatory (behavioral) and liability law. Presentation of other aspects is largely avoided for the sake of brevity and clarity (e.g., aspects of criminal law, constitutional law, etc., are not discussed).

The regulatory (behavioral) law discussed here covers regulations that rule human behavior (of road users). Regulatory (behavioral) regulations in the field of road traffic determine the behavioral conduct criteria also for drivers. Liability law is understood here in a broader sense than the duty to compensate for damages in various situations. Hereinafter, mostly situations of liability for damages in road traffic are considered as well as liability in case of product defects. This focus on regulatory (behavioral) and liability law seems appropriate, since driver assistance systems in the broad sense defined here, with the division of tasks they bring about, influence the driver in the loop directly or indirectly. Provisions relating to vehicle control systems are of particular interest, because in some cases fundamental changes and contradictions – depending on the operation mode – can be demonstrated from a legal perspective.

2.1 Informing Systems (Operation Mode A)

For legal classification, let us first sum up the characteristic function of informing systems: As stated already, these systems act – because they inform the driver – only indirectly on the vehicle control, by informing and warning the driver. The choice of appropriate action is left to the driver. In addition, redundancy of the information provided must be considered. Herewith it should be specified to what extent the system provides the driver with information he could have perceived himself without the system (which will be the general case), or whether the information increases the driver's perception horizon to an area that lies beyond own immediate perception.

It shall not be discussed here where the information or warnings are generated, i.e., whether they are vehicle-autonomous (e.g., based on vehicle's own sensors), or whether the information reaches the vehicle through communication with other facilities (in other vehicles or traffic facilities equipped for this purpose). In legal terms, such networking of the vehicle (vehicle-to-vehicle communication) may potentially broaden the number of liable parties. Even then, however, the mechanism of action underlying informing systems will still apply.

2.1.1 Evaluation of Regulatory Law Aspects

Regulatory legal aspects are of comparatively minor importance in the case of informing systems: Since the information provided always requires the driver's intervention for a causal effect on the course of events, this corresponds fully to the guiding principles of Road Traffic Codes: The central role of vehicle control lies with the driver, who is supported in this task by the information provided by the system. The driver is always responsible for the decision whether and in what way to respond to an information or warning.

From the perspective of regulatory (resp. behavioral) law, the design of the so-called human-machine interface of a system and its use by the driver are important. Thus, there is a relation to provisions of the German Road Traffic Code (StVO): Thus, section 23 para. 1 StVO describes the duty of the driver not to affect sight and hearing by the use of "devices." However, the resulting requirements are essentially limited to cases with clearly discernible adverse impact. Obstructing the view by devices brought into the vehicle – as may occur for example in the case of mobile navigation devices – will fall within the regulatory scope (within the wide discretion combined with case-specific legal assessment). In principle, the German road traffic regulations currently leave the use of informing systems to the driver as their own responsibility (at least as long as they are not designed to display or interfere with traffic surveillance measures, section 23 para. 1b StVO, or as long as the devices do not combine telecommunication functions since hand-held communication devices are banned, section 23 para. 1a StVO).

With regard to the design of the human-machine interface, there is only a regulation in the form of a legally non-binding EU recommendation – the European statement of principles on human-machine interface (ESoP 2007). The recommendations therein remain abstract and general and leave the manufacturers of informing systems the necessary freedom of design (in the sense of allowing for the economically necessary freedom in order to distinguish from other manufacturers). The recommendations formulate in an abstract manner which design goals are to be considered under the aspect of road safety. As stated in the introduction of the European Statement of Principles, these are minimum requirements that need additional consideration of standards – cited for the most part – and national legislation. Regarding the few aspects contained that allow for a binding regulation, and the related challenges see (Gasser 2008).

Regarding administrative fines enacted in case of traffic offenses – e.g., for speeding – a specific feature in case of informing systems, should be pointed out. Although it may be rather theoretical, the legal consequence ultimately results from the redundancy of information supplied: If the driver is alerted to a breach of traffic rules (such as the applicable speed limit) by an informing system (and in addition to visible road signs), the circumstantial evidence indicates that the offense is being committed intentionally (at least if the violation subsists after the system's warning). Since the German penalty catalogue regulation (BKatV) only assumes negligence, the standard rates according to section 1 BKatV in conjunction with the annex, part I and according to section 3 para. 4a BKatV would basically need to be doubled in cases of intentional perpetration. This also complies with the requirement of section 17 para. 2 of the German Administrative Offences Act (OWiG). For this, it would be sufficient that the existence of such a system in the vehicle is known, the function is designed so that the information in such a situation cannot be left unnoticed by an average driver, and there is no indication that the feature was faulty in the individual case.

2.1.2 Evaluation of Liability Aspects

In respect of liability, the indirect nature of informing and warning systems is as crucial as it is for the regulatory legal aspects already discussed. With regard to legal liability, both from the point of view of the manufacturer's (fault based) liability under section 823 para. 1 German Civil Code (BGB), and under the German Product Liability Act (strict liability), it must be first determined whether a product is to be classified as defective in the first place. The (objectively to be considered) presentation of the product, which determines legitimate product-related expectations, should be taken into consideration (see section 3 para. 1 lit.

a Product Liability Act). This includes, inter alia, the user manual that describes the system (as well as advertising statements, etc.). If – as normally to be expected – in the case of informing and warning systems it is described that the driver must react appropriately in the event of a system failure or a malfunction, it seems questionable whether lacking and erroneous warnings or information not or wrongly provided by a system can justify a claim relevant to product liability at all. In these cases user manuals determine to a large extent which expectations are considered justified. In the case of redundancy by a parallel information perception of the driver and the assistance system, it is not clear why it should not be possible for the driver to compensate for errors of the assistance system.

This is certainly plausible in the case of redundant-parallel assistance systems, i.e., when the driver is provided only with information which he can perceive himself. Systems that expand the driver's information in areas located beyond own perception abilities, liability-issues might be assessed differently. In this case, the redundant-parallel task execution cannot be assumed in the same way; rather, the information or warning as a basis for a control action acquires another quality. The legitimate expectations of a driver described above may not be limited in a corresponding manner. The circumstances of the case are therefore crucial.

For the so-called cooperative systems (vehicle-to-vehicle and vehicle-to-infrastructure communication), the described effect of indirect information is the same. Here as well, the indirect provision of information and warnings has a strong effect (Kanz et al. 2013). The question of broadening the circle of potentially liable parties is yet to be considered on a case-by-case basis; this cannot be addressed in the present context.

2.2 Continuously Automating Systems (Operation Mode B)

The continuously automating systems of operation mode B were evaluated in legal terms by a project group installed by the Federal Highway Research Institute (BASt) on the basis of the automation levels described (see Sect. 1) (Gasser et al. 2012). Regarding the results presented below, as a preliminary point, a restriction must be understood: Legal assessment of functions which, albeit temporarily, exercise independent control of the driver, can only be done sensibly if they are overall consistent with the legal system. The results presented below demonstrate that from level 3, i.e., beginning from high automation, this is not the case in some situations. The following individual assessment still has its value, as it uncovers contradictions and explains which aspects need to be resolved in legal terms.

The joint evaluation of the BASt-Project Group has focused on the legal aspects of the regulatory (behavioral) law and liability law, without claiming to be exhaustive. Rather, when drafting the report, it became apparent that other aspects – such as the expressly mentioned aspects of driver licensing laws – must likewise be assessed for higher degrees of automation. The limitation in legal assessment seemed appropriate according to the state of the art foreseeable when tendering.

2.2.1 Evaluation of Regulatory Legal Aspects

From a legal standpoint, two basic aspects must be distinguished: on the one hand, the obligation of the driver to control the vehicle at all times, on the other, the (very specific) question of the duty of a driver to perform two-handed steering.

Duty of the Driver to Control the Vehicle

The duty of the driver to control the vehicle is expressly stated only in section 3 of the German road traffic regulations (StVO). Thereby, the driver is required to drive only as fast that the vehicle can "constantly be controlled" (section 3 para. 1 sentence 1 StVO). Drivers must therefore always "control" their vehicle, and for that reason the speed should be selected so that it is suitable, depending on the subjective (driving skills, driving experience, etc.) and objective factors (road condition, weather, route, etc.) and meets all not entirely unlikely "situations on the road [. . .]" (Bouska and Leue 2009). The concept underlying the Road Traffic Code is influenced, among others, by the basic rule for traffic behavior, the caution and consideration requirement of section 1 para. 1 StVO, which is to be taken into account in the interpretation of all specific requirements and prohibitions directed to the driver (König 2013). It also captures basically any not specifically regulated neglect of driving tasks by the driver. Such behavioral requirements are found in the following, more specific provisions of the Road Traffic Act (on safe distance, section 4 StVO; overtaking, section 5 StVO; etc.). The rules have in common that they apply to road users, in so far as relevant here, in particular to drivers of motor vehicles on public roads. Their purpose is to avert dangers for public order and safety (see section 6 para. 1 no. 3 StVG – German Road Traffic Act).

The concept of control over the vehicle is also found in the Vienna Convention on Road Traffic (Federal Law Gazette 1977), which regulates authorization of drivers and vehicles for cross-border traffic (the cross-border element is clearly shown in national laws, e.g., in the recognition of international driving licenses according to art. 41 and appendix 7 of the Vienna Convention in section 29 para. 2 German driving license regulation (FeV)). Article 13 para. 1 of the Vienna Convention requires that the driver "shall in all circumstances have his vehicle under control so as to be able to exercise due and proper care and to be at all times in a position to perform all maneuvers required of him." The wording chosen here identifies – to this extent comparable to section 3 StVO – the speed and distance behavior of the driver as an obligation addressed to him. To this extent, art. 8 of the Vienna Convention must be further considered, which states in paragraph 1 that each vehicle must have a "driver," and stipulates in paragraph 5 that "every driver shall at all times be able to control his vehicle." The "driver" is defined in Art. 1 lit. v) of the Vienna Convention as "any person who drives a motor vehicle or other vehicle (including a cycle)[. . .]." Both, Art. 8 and Art. 13 of the Vienna Convention are included in "Chapter II Traffic Regulations," and thus are aimed at those who drive vehicles.

Already at a very early stage in the development of driver assistance systems it has been pointed out that there would be a contradiction to the control over the vehicle, if non-overridable systems would be approved, since this would make it impossible for a driver to act in line with his duties (Albrecht 2005) and (Gasser 2009). The continuously automating systems considered here (operation mode B) are and can be interpreted as overridable at all times, so that there is no contradiction to the vehicle control by the driver under this aspect.

It must, however, be questioned in how far these provisions of the Road Traffic Code and the Vienna Convention can take effect on driver assistance systems and vehicle automation at all. Starting from the scope of the systems, we will find that the present systems of continuous vehicle automation are designed to be "intervening" and directly influence the vehicle control. In that regard, tasks are performed by technical systems which are assigned solely to the driver according to the underlying assumptions of both the German road traffic regulations as well as the Vienna Convention (without this fact being expressly stated). The systems assume – with varying technical range and depending on the degree of automation – the task of vehicle control. According to legislative concepts and/or the concepts of the contracting parties, this task obviously lies as the sole responsibility of the driver. According to the concept of classification proposed here, it must be examined to what extent the systems which assume control-related tasks of the driver still correspond to the statutory or contractual concept. This approach seems appropriate, since the technical requirements do not refer to control-relevant processes (so far with very few, narrowly defined and isolated exceptions in the area of electronic braking and steering systems).

Driver Assistance Systems

According to the definitions presented above, the driver assistance systems of operation mode B assume either longitudinal or lateral control of the vehicle, while the driver performs the other respective driving task. It is important to understand that here even the automated driving task is executed by the driver assistance system to a limited extent ("within technical limits"). By way of illustration, the limited maximum acceleration or deceleration performance of an adaptive cruise control (ACC) (see ▶ Chap. 45, "Adaptive Cruise Control") can be mentioned here, or the limited steering torque applied by a lane keeping system (see ▶ Chap. 48, "Lateral Guidance Assistance").

The first automation level (driver assistance systems) therefore leaves vehicle control largely to the driver, and can be overridden at all times. If a driver makes use of these systems as intended by the manufacturer, this will usually increase his driving comfort but he must execute the respective non-automated part of the driving task further. Additionally, his obligation remains to continually monitor and correct the system, in the execution of the automated driving task (longitudinal or lateral control). Driver assistance systems feature system limits which allow safe vehicle control only in cooperation with the driver. In addition, there is the fact that the driver, according to the current design of driver assistance systems, serves as a fall-back in case of system failure. What changes basically is that the execution-related part of the driving task changes with regard to active execution. What remains unchanged is the duty of the driver to gather all driving-relevant information in parallel with the system. The driver is thereby able to regain control over the vehicle at any time, since he is also technically on a higher level position than the system: Driver's control operations always "override" the system. This is consistent with the driver's obligations from the German Road Traffic Code.

Partial Automation

The concept of partial automation according to the definitions presented above refer to systems capable of managing both the longitudinal and lateral control of the vehicle, while the driver continuously monitors system control. When the automatic driving no longer corresponds to the driver's wish, the driver corrects or takes over.

The change in the driver's task is relevant to the extent that the driver, in the highest development stage of this degree of automation, no longer needs to actively execute the vehicle control. The system is yet – so far equivalent to driver assistance systems – designed so that the driver has the task to complement the control of the system at system limits and as a fall-back in case of failure. For this, the driver must – as in the case of driver assistance, level 1 – in parallel take up all information relevant for the control of the vehicle. Only in this way is it possible to make use of the system as intended, and to take over control or correct the automated task execution, if this becomes necessary. As in the case of the driver assistance, this is technically possible at any time. The fact that, in the case of partial automation, the control over the vehicle is carried out by human action only in case of correction or takeover, and in other cases by refraining from action, does not lead to a fundamentally different legal assessment: Ultimately, active engagement in task execution and refraining from action are both just "different manifestations of wilful decision" (Wessels and Beulke 2012). Since the driver does not lose control over the vehicle due to ongoing monitoring, the intended use of partially automated systems remains compatible with driver's obligations under the German Traffic Code.

High Automation/SAE: Conditional Automation

In case of high automation ("conditional automation" according to SAE), the driver could, according to the definitions adopted, use the system as intended without continuously monitoring the driving behavior. The driver, however, is required to resume vehicle control after a takeover is prompted by the system within a specified lead time. This is necessary because the vehicle that recognizes all the system limits and prompts the driver to take over, it is not yet capable of automatically returning to the risk-minimal state from *any* initial state.

The technical effort to implement high automation is tremendous. A benefit for the driver (such as increased comfort) results only if it is possible for the driver to cease monitoring of the traffic environment. The German Road Traffic Code does not allow a driver to turn away from the driving task while driving. If the driver ceases monitoring, the control over the vehicle by the driver is dispensed at least temporarily: It is not to be expected that the driver will still be able to take over vehicle control immediately; initial studies demonstrate this (Gold et al. 2013). As a result, vehicle control can no longer be assumed, as soon as a driver turns away from the task of driving (and the turn-away time is not purely marginal, or expressly prescribed under certain circumstances, such as in the case of looking back to watch out for oncoming traffic when taking a left turn according to section 9 para. 1 sentence 4 German Road Traffic Code).

From a legal perspective, there is no longer a wilful decision combined with vehicle control when turning from the task of driving. In contrast to partial automation, the driver no longer merely refrains from action. The driver's action has no relation to vehicle control at all and can thus no longer be categorized as human control behavior. Rather, highly automated driving is a case of exclusively automated vehicle guidance: as soon as the driver takes his attention from observing roadway and traffic, he can no longer adjust the control by the system (which possibly remains partially perceptible through other senses) to the specific traffic situation developing on the roadway. Human will is no longer related to vehicle control. This speaks very clearly for the complete absence of human action in the specific driving situation (at least in case the driver makes use of the highly automated function as intended by the manufacturer by ceasing to monitor system control). This use of the system no longer allows to be considered as "redundant-parallel" division of tasks: Due to the fact that the driver no longer monitors the surroundings as the fundamental basis of vehicle control no task remains that is in any way related. Since the driver is slipped as the relevant controller, vehicle control is neither parallel nor redundant with the driver as fall back.

With a focus on regulatory law, evaluation therefore results in a fundamental incompatibility in legal terms. This relates more specifically to the driver's obligations under the German Road Traffic Code.

Full Automation/SAE: High Automation

Additionally, in the case of full automation ("high automation" according to SAE), in comparison to level 3 adds as further criteria that the driver is no longer ultimately required to perform a takeover upon request, because the system is capable of returning to the risk-minimal state by itself. As with the level 3, the driver can turn to other activities that have not yet been determined in detail. However, the driver remains in the effective position to gain control over the vehicle at all times.

From a legal perspective – as in the case of level 3 – vehicle control by the driver is slipped, as soon as the driver turns from vehicle control when using the system as intended. As with level 3, no wilful decision of the driver is present, but only automated vehicle control, which also can no longer be classified as a "redundant-parallel" division of tasks between the human and the machine. Again, this results in a fundamental incompatibility with the duties of the driver according to Road Traffic Regulations.

Driver's Obligation to Two-Handed Steering

The German Road Traffic Code (StVO) includes only the express obligation not to drive single-track vehicles hands-free (applicable to cyclists and drivers of motorcycles); see section 23 para. 3 sentence 2 German Road Traffic Code (StVO). A similar arrangement has not been made for drivers of double-track vehicles, so that the loophole prevents analogous application of this prohibition (section 23 Road Traffic Code regulates other obligations, including those of the drivers of multi-lane vehicles, while the ban of hands-free driving is explicitly limited to the said drivers of single-track vehicles).

The legal opinion that hands-free driving of multi-track vehicles is a breach of duty has been derived from section 1 paragraph 1 of the German Road Traffic Code (StVO): This perception is based on the assumption that hands-free driving is a violation of the requirement to drive cautiously and considerately. Such a breach of duty is, however, not prohibited (nor will this be considered a breach of the obligations in the sense of section 24 of the German Road Traffic Act (StVG)) – unless this coincides with a concrete endangerment according to section 1 paragraph 2 of the German Road Traffic Code (StVO) (König 2013).

It is consequential here to assume a breach of duty, when measured by vehicles that are used on the road nowadays: As a general rule, those are vehicles with steering systems of the levels 0 and 1 (“driver only” and “assisted”). Based alone on the definitions stated above, it is clear that even in the case of a driver assistance system the automation is only “limited,” so that the driver must take corrective action when reaching system limits and in case of system failure. Thus, the assumption of a breach of duty in case of hands-free driving stands to reason, as far as systems are used that require driver’s intervention without a delay to deal with system limitations and in case of system failure. If such systems are used for driving “hands-off” this is likely as much contrary to duty, as is hands-free driving without an automated steering system (of operation mode B).

The question remains, however, to what extent a breach of duty of driving “hands-off” is to be assumed in case a steering system can automatically handle all disorders which have short term effect, or in case the effects remain marginal in the concrete situation.

Therefore, it would be difficult to identify a breach of duty if all situations are equally well controlled as in the case of driving “hands-on.” The same would apply, in case a specific driving situation is mastered by the driver driving “hands-off” at system limits and in the event of a malfunction as well as would be possible in case of driving “hands-on.” For example at a very low speed (as is the case nowadays already when using parking steering assistance systems), or with a significant lane width and in the absence of other road users, which might be endangered in case of a system limit or system failure.

2.2.2 Evaluation of Legal Liability

In respect of legal liability with relation to continuously automated systems, two aspects must be considered above all: On the one hand, the liability (especially) acc. to the German Road Traffic Act (StVG), and the manufacturer’s liability for defective products.

Liability Under the German Road Traffic Act (StVG)

The civil liability of the vehicle holder in accordance with section 7 para. 1 German Road Traffic Act (StVG) assigns the risk of operational hazard to the holder.

The operational risk includes both driving errors of the driver and technical defects. It can be assumed that a technical defect of automatic vehicle control – regardless of automation degree – will be counted offhand to the operational risks, because section 7 para. 1 StVG defines "vehicle operation" broadly (Albrecht 2005). Therefore the requirements for evidence to be provided by the injured party are the lowest within the scope of the owner liability, because only the operation of the vehicle and damage causation must be proven.

In addition to the vehicle owner, also the driver is liable in cases of section 7 para. 1 StVG according to section 18 German Road Traffic Act (StVG). (Liability of the driver can additionally be based on section 823 para. 1 BGB or section 823 para. 2 German Civil Code (BGB), the latter in conjunction with the German Road Traffic Code (StVO) as protective law, both shall not be discussed further here). According to section 18 para. 1 sentence 2 StVG, there is a legal presumption against the driver to have been at fault while driving. This can be applied consistently up to and including level 2 automation (partial automation, see above). At higher degrees of automation, level 3 or higher (beginning from high automation, resp. "conditional automation" according to SAE), the presumption of fault against the driver no longer seems appropriate in all cases: In a case of third parties misconduct causing an accident, the driver can be expected to exonerate himself from the presumption of fault by producing evidence of the said misconduct. This will apply in conventional vehicle control level 0 ("driver only") up to and including partially automated vehicle control of level 2 ("partially automated"). These levels of automation require constant traffic monitoring from the driver, and (in levels 1 and 2) additional monitoring of the control system and, where appropriate, immediate error correction or system override. Therefore the driver may be expected to observe a misconduct of a third party leading to an accident at any time. It is possible for the driver to provide information in this regard.

The situation is radically different at the higher levels of automation, beginning from level 3 (high automation, resp. "conditional automation" according to SAE). This presupposes, such use would be legally permissible, i.e., the driver would not be required to monitor traffic in all driving situations, see above Sect. 2.2.1. The driver's tasks when using level 3 functions are limited to the readiness to resume vehicle control after a reasonable lead time. The driver is thus freed from any active vehicle control and cannot – as long as he is not actively overriding the system or fails to resume vehicle control – act negligently from the outset of the automated phase. In addition, the driver, who detracts his attention which is the intended use of level 3 systems, cannot observe the misconduct of a third party, and therefore must produce evidence that an automatic control system was activated at the time of the accident.

This situation is inconsistent due to the legal presumption of driver's fault and no longer seems justified. However, it is not unacceptable, since the driver may possibly succeed to produce exculpatory evidence – possibly with the aid of technical means – under civil law. Moreover, the situation is not fundamentally different from the conventional vehicle guidance ("driver only," level 0), if fault of a third party is not provable under civil law. Ultimately an economically unsustainable situation is not present since the driver is co-insured in the motor vehicle liability insurance (section 2 para. 2 no. 2 of the German Regulation on the insurance coverage in the motor vehicle liability insurance (KfzPflVV)).

The motor vehicle liability insurance nowadays comprises those systems, which are covered by vehicle type-approval. In that regard, future automatic systems will be included in insurance because these systems will (by then) be covered by the vehicle type-approval (which they are not today). Another prerequisite is that the legal framework conditions for the intended use of these systems of level 3 and above are created.

Manufacturer's Liability for Defective Products

Apart from claims against the manufacturer on the basis of contractual liability, which shall not be discussed further here, the manufacturer's liability exists with respect to any third party for damages based causally on a product defect. As relevant basis for claim, either strict liability (regardless of culpability) under the German Product Liability Act, or the manufacturer's liability under section 823 para. 1 German Civil Code (BGB) (or section 823 para. 2 BGB in conjunction with a protection law) come into question. Manufacturer's liability is applicable for the culpable breach of duty of care by placing a defective product on the market, assuming the damage is caused by the product's defectiveness. Both bases for claim are very close today, also due to the reversal in the burden of proof for the requirement of fault in case of manufacturer's liability. Therefore there is no need to go into the specifics here.

Defectiveness is the central concept of product liability; it must be proven by the injured party. Automated systems from level 1 (assisted), up to level 3 (highly automated, resp. "conditional automation" according to SAE), assume interaction with the driver, so as to ensure a safe ride. This is the necessary consequence of the division of tasks between the driver and the machine in the case of these levels of continuous automation. On the part of the driver, knowledge and awareness of the performance of each system (especially regarding system limits) is essential. Only this knowledge allows the driver to recognize the need to monitor a system, to initiate corrective actions when needed or to resume vehicle control and to take operating actions as required. In this respect, the driver's instruction is crucial in order to influence the users' expectations appropriately and to avoid the "instruction error" as a subset of the product liability errors.

Another relevant type of error in product-liability in this context is the design error. In this context, the standard of what is to be considered an error is particularly relevant. The German Federal Court of Justice has ruled with regard to passive safety systems, that a manufacturer "[. . .] must take measures, already within the design and planning phase of the product, that are objectively necessary to prevent dangers, and remain reasonable according to objective standards [. . .]" (German Federal Court of Justice 2009). In order to avoid dangers, a solution must therefore be applied that is considered a series operational solution among experts. This broad requirement finds its limit in the reasonableness of such measures, depending on the risk level of the product and the economic impact of a safety measure (Lenz 2009).

If dangers cannot be avoided even with the best available science and technology, "[. . .] the manufacturer is generally obliged to warn the product users of those dangers that may occur under normal conditions of use or as foreseeable misuse, and which do not pertain to the general risk knowledge of the user group [. . .]" (German Federal Court of Justice 2009). Apart from the resulting uncertainty as to which measures must be taken in each individual case in order to prevent danger, there is also the fact that the present statements relate directly only to passive safety systems in motor vehicles. The statements can, therefore, only to a limited extent be applied to automated controls that require direct interaction with the driver. However, the significance and scope of driver instructions while operating still has very high importance in case of automation levels 1 and 2 (assisted and partially automated): It can be assumed that no operational series solution for risk avoidance (by means of technology) is available for these continuously automated systems. This raises the importance of user instructions. The same is true for level 3 systems (high automation, "conditional automation" according to SAE), in so far as the transfer of vehicle control back to the driver is affected.

Continuously automated systems up to and including level 3 (high automation, "conditional automation" according to SAE), are technically limited in their means to return to the risk-minimal state by autonomous action. For level 3, therefore, the instruction is inextricably linked to constructional requirements as far as the determination of defectiveness is concerned. Both potential sources of defectiveness must therefore be tackled together.

In case the intended use of level 3 and 4 systems (high and full automation, "conditional" and "high" automation according to SAE) is made, this means the driver will turn from the task of traffic monitoring (provided the other framework conditions for this use would be created, see above Sect. 2.2.1). Therefore the consequences that arise from the fact that the driver is (temporarily) no longer in the driver-vehicle control loop must be considered. The partition of tasks affects the requirements for system construction: In the light of the risk level originating from such independent machine control on public roads, level 3 and 4 systems need to be designed in a manner that they are automatically able to deal with all situations that may occur whilst driving automatically.

At the same time, this finding inevitably leads to the assumption that any type of damage that occurs during periods of level 3 or 4 automation implies some kind of causally underlying product defect. This assumption applies, unless the damage is caused exclusively by a third party in the traffic situation or the driver has been overriding the system (including the well instructed driver's failure to resume control after a "sufficient lead time"). This assumption of defectiveness will, of course, only apply if aspects of procedural law are taken into account, in particular the burden of providing evidence of circumstances and the burden of proof in a civil case (Gasser et al. 2012). If, therefore, level 3 systems do not work correctly, with the result of an accident – and this is provable procedurally – this would suffice to sue the manufacturer for damage since the driver will have been entitled to turn from the task of driving.

From level 3 automation ("conditional automation" according to SAE), another important question arises from the extent to which such systems may still exhibit system limits. Again, the definition made allows the conclusion that system limits are possible only to the extent of the described resumption of vehicle control by the driver "after sufficient lead time." System limits, however, which cause immediate deactivation or require immediate error correction or immediate resumption of control by the driver, are either not feasible, or only to a very limited extent in rare cases.

2.3 Intervening Emergency Functions (Operation Mode C)

Intervening emergency stop systems (operation mode C) are usually also referred to as "driver assistance systems." This is consistent in the strict sense of the term, since they also support the driver. In fact, the benefits are shown in situations that have a special quality, as the driver can react to a trigger in road traffic only with a delay in certain situations (or not at all in case of an inability to act for health reasons). For this reason it does not seem justified to assume a division of tasks in these modes of operation.

2.3.1 Importance of Controllability

The legal debate on these systems has so far been conducted primarily in relation to the requirement of controllability, which has been derived from art. 8 para. 1 and para. 5 and art. 13 para. 1 of the Vienna Convention on Road Traffic (VC). The requirement of constant control, (art. 8 para. 5 VC) and command over the vehicle (art. 13 para. 1 VC) to be applied by a driver (art. 8 para. 1 VC) has been understood to rule out the use of non-overridable systems in road traffic. Respective system design can prevent the driver from performing duties in road traffic (Albrecht 2005; Gasser 2009). A contradicting legal opinion questions the applicability of these rules to driver assistance systems from the outset, since they are stated in the second chapter of the Vienna Convention on Road Traffic, and consequently refer to the driver alone, without developing any significance for driver assistance systems. Furthermore it is argued that there is no similar injunction in the third chapter of the Vienna Convention, which would limit the (technical) question of admission to the cross-border traffic, see (Bewersdorf 2005).

A decision on one of these contentious legal opinions regarding intervening emergency systems, from today's perspective, seems irrelevant: All intervening emergency systems discussed in this section, are designed today to be overridable. As far as has previously been unclear, whether interventions of emergency systems in time-critical situations are possible, an important indication was provided by the so-called "General Safety Regulation" (EC directive 661/2009 of 13th July 2009), which, among other things, makes emergency brake assistance systems compulsory for buses and commercial vehicles of the categories M2, M3, N2 and N3, see art.

10 section 1 EC Regulation 661/2009. Although there is no distinction according to whether such interventions are initiated by the system or by the driver, there is much evidence that even system-initiated interventions do not contradict the cited provisions of the Vienna Convention on Road Traffic.

In general terms, it could be inferred that in other cases of driver's inability to act it must be possible to intervene in order to avoid an accident or to mitigate consequences of an accident. A contradiction to the regulatory legal requirements in these special situations is then excluded, especially if the driver would have the ability to override (even if then only of theoretical importance).

2.3.2 Special Case: Emergency Stop Assistance

Unlike other emergency functions, emergency stop assistance intervenes in situations that are characterized primarily by the driver's inability to act (e.g., for health reasons). The intervention time of these systems is in contrast to other emergency systems (e.g., automatic emergency braking or in the future possibly also emergency steer assistance, etc.) not necessarily short. Moreover the situation cannot necessarily be described as near-collision.

The Project Group report assigned the emergency stop assistance systems erroneously to operation mode B (without explicitly making this distinction), but it treated them as a special case of full automation (level 4), see (Gasser et al. 2012). This fails to take the characteristics of these systems into account: Emergency stop assistance is not continuous automation consciously being applied by the driver, in the sense of (partial) transfer of the control task to the system, but rather, an automated system intervention in case of a recognized inability to act on the part of the driver. These circumstances and the inability of the driver to take the appropriate control action himself underlie the decision to assign these systems to the intervening emergency systems (operation mode C).

Functions of emergency stop assistance are not yet available on the market at the time of writing this chapter. Scenarios that describe this type of operation mode so far assume that emergency stop assistants intervene in case of a physiologically-induced loss of control by the driver. They carry out a fully automatic braking of the vehicle into a safe, so-called minimum risk state in order to prevent accidents. A risk-minimal state, for example on multi-lane motorways, means a moderate deceleration, which – if this is possible with a minimum risk due to traffic conditions and system state – may include a lane change to the breakdown lane or hard shoulder. The vehicle is decelerated to a standstill there. If it is not possible to change lanes, the vehicle comes to a halt on its current traffic lane. The prerequisite for the system intervention is the driver's loss of control due to medical impairment, which can be detected based on various indicators. A termination of the automatic control operation would be possible at any time by the driver by deactivating the emergency stops maneuver, or overriding it, respectively (Bartels 2012).

From the viewpoint of regulatory (behavioral) law, it must be stated that an addressee of the rules of conduct, e.g., the driver according to the German Road Traffic Code (StVO) is no longer capable of acting. The driver cannot be accused of this lack of control, as long as there have been no clear signs that would cast doubt on his driving ability before.

As far as the liability of the vehicle keeper according to section 7 para. 1 StVG is concerned, it must be assumed that the automatic vehicle control implemented by the emergency stop system belongs to the "operation" of the vehicle, provided the system has been licensed for road traffic in the course of the vehicle type approval. Therefore, according to the rule of holder liability, damage brought about triggers liability of the vehicle "holder." In the context of vehicle operation, it should be considered that the liability also covers the continued operational danger of a motor vehicle standing on an express road, such as a motorway. Consequently, where the braking or standing vehicle causes damage because of low speed (or standstill), this would also be attributable to vehicle operation within the meaning of the provision.

The situation is different in the case of the driver's liability: If the driver could not at least foresee that a loss of consciousness or inability to act could occur while driving, no fault may be attributed to the driver. This evidence should, however, under applicable liability law, be provided by the driver in accordance with section 18 para. 1 sentence 2 StVG, and therefore it is subject to the procedural uncertainties in the context of the burden of proof in civil proceedings.

In the context of motor vehicle liability insurance, it is assumed that control operations of an emergency stop assistant would also be insured, provided that these lead to damage to third parties (assuming again the vehicle would, under the vehicle type-approval, be approved for road traffic). However, only "vehicle use" is insured under the motor vehicle liability insurance. A conscious use of an emergency stop assistant is certainly not made by drivers. However, the driving state which is to be brought to the risk-minimal state through the function is apparently triggered directly by the preceding vehicle use. The situation is therefore comparable to the completely uncontrolled skidding of a vehicle due to high speeds, whereby the skidding can be considered just as uncontrollable for the driver. Furthermore, swerving can be assigned to the preceding use made of a vehicle too (e.g., driving at too high speed). Thus it can be assumed that control operations of emergency stop assistants are likewise to be considered included in today's scope of insurance.

Product liability is affected by the fact that emergency stop assistants act beyond drivers' control in cases of physiologically-induced inability to act. Therefore, the control of the vehicle in case of emergency stop assistance might not be measured against the same hazard avoidance scale applied to continuous level 3 or 4 automation (cp. Sect. 2.2.2 on operation mode B). Rather, it can be assumed that such a product, provided that at the time of entry into traffic it represents an operational solution to the state of science and technology, corresponds to the safety necessary according to product liability law. This should apply even if in individual cases, accidents occur that result from system limits. This legal result must be considered appropriate, because, as of today, even emergency stops which appear merely satisfactory may already help to improve road safety by avoiding the otherwise inevitable high risk of vehicles running completely out of control.

2.3.3 Other Emergency Systems

Finally, it should be noted that the comments on emergency stop assistance are also applicable to the legal situation of other systems within operation mode C: The starting point, the driver's lack of ability to act and the character of these systems that are not designed to distribute workload between the driver and the machine, is the same in both cases. An important distinction, however, must be made in the case of other emergency systems (e.g., for emergency braking or in future emergency steering systems): Basically the driver is still available, but cannot react appropriately due to the framework conditions. Either the driver has misinterpreted the situation or the task becomes apparent very suddenly, so that the driver can only respond to the situation with a delay in time. This therefore leads to another important consideration: The vehicle control may need to be transferred back to the driver after an intervention, possibly after the reaction time already. This represents a further challenge to human-machine interaction. In addition, the aspect of override in cases of improper interventions by the system requires special attention, since an override by the driver can be very risky, and the system manufacturer will usually intend to maintain the ultimate decision-making authority of the driver.

3 Regulation in the USA

Although vehicle automation is discussed all over the world, an express legal framework has been adapted for this purpose in only some US states (and even then only partially). This section briefly describes these particular statutes and rules.

Regulation of automated driving in the United States must be understood in the context of the overall legal system. This system seeks both to avoid injuries through preventive measures that improve road safety and to compensate for some of the harms that do occur. Responsibility in a legal sense is concerned with obligations, including duties of contract and of due care, and with liabilities, including criminal and civil judgments. However, responsibility in this legal sense is not necessarily coterminous with technical requirements or moral obligations.

Figure 3, which illustrates this understanding in the context of automobiles, identifies four quadrants of regulation (Smith 2013a). Rules may be established publicly (by public authorities) or privately (by a natural or legal person); they may act prospectively (before a crash or other incident) or retrospectively (after a crash or other incident). Public and prospective measures include regulatory performance standards, process standards, and conditions of market entry. Public and retrospective measures include criminal penalties, legally mandated recall actions, and legislative or regulatory hearings and investigations. Private and prospective measures include contractual terms, industry standards, and insurance conditions. Private and retrospective measures include civil claims in the case of product defect or negligence as well as reputation or sales impacts to a manufacturer, its suppliers and other companies. These measures in turn interact with those in the other quadrants. Thus, for example, statutory insurance mandates may strengthen the

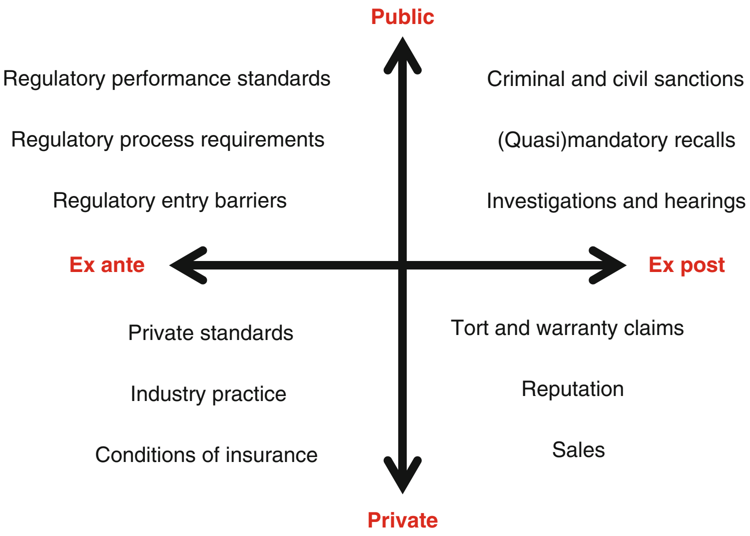


Fig. 3 Taxonomy of regulation

U.S. Constitution	• Constitutional due process protections
Federal statutes and some international agreements	• National Traffic and Motor Vehicle Safety Act (1966) • Geneva Convention on Road Traffic (1949)
Regulations of federal agencies	• Federal Motor Vehicle Safety Standards (FMVSSs)
Activities of federal agencies	• NHTSA investigations
U.S. state constitution	• Diverse regulatory content
State statutes	• Rules of the road and state automated driving statutes
Regulations of state agencies	• Rules for automated driving in Nevada and California
Activities of state agencies	• Licensing of drivers and registration of vehicles
Common law	• General tort law
(Industry norms and private standards)	• (ISO / SAE / ANSI Standards)

Fig. 4 Hierarchy of legal norms in the United States

position of an insurance company towards its policyholders, industry standards may influence the outcome of civil suits, and hearings may weaken consumer demand.

Statutes at the federal level and in the individual US states are only one – albeit important – part of regulation in this broad sense. Figure 4 presents a hierarchy of legal norms in the United States (Smith 2012a). The US constitution, which structures the relationship among the federal government, state governments, and individuals, occupies the highest level. On the second highest level are statutes enacted by Congress, such as the Traffic and Motor Vehicle Safety Act of 1966, and certain treaties confirmed by the Senate, including the Geneva Convention on Road Traffic of 1949. Statutes can establish federal agencies like the National Highway Traffic Safety Administration (NHTSA), which is responsible for motor vehicle safety. These agencies carry out their statutory duties by enacting regulations and making administrative decisions, which constitute the next two legal levels. NHTSA, for example, is responsible for the federal motor vehicle safety standards (FMVSSs) to which vehicle manufacturers self-certify their products (rather than obtaining prior approval of an authority, see Sect. 4.1). NHTSA also initiates investigations of possible defects that could lead to recall actions. These rules and actions can to varying degrees be reviewed and interpreted by federal and state courts.

Subordinate to federal law is state law, which likewise has multiple levels: A typically more comprehensive constitution, statutes enacted by the state legislature, and the rules and decisions of state agencies. Licensing, registration and the conduct of non-commercial drivers and their vehicles are largely regulated by the states. These laws are definitively interpreted by the courts of the respective state but also provisionally interpreted by other courts. In almost every US state, an even lower level of regulation exists: The common law, which is specific to each state, is not codified, and is developed further by the courts. The common law is very important for questions of civil liability (Smith 2013b). For example, the US Supreme Court has struggled to clarify the respective roles for federal safety standards and state common law in determining product defectiveness in product liability cases (Geier v. American Honda Motor Co 2000), (Williamson v. Mazda Motor of America 2011). In addition, community norms and private standards may affect this common law. Finally, an important principle must be understood that runs through the entire legal system (in the United States as well as in Germany): Everything is allowed unless forbidden (Smith 2012b).

Regulation of automated driving takes place against this background. In February 2014, NHTSA cautiously announced that it would “begin taking steps to enable vehicle-to-vehicle (V2V) communication technology for light vehicles” (US DOT 2014), which could be of importance for automated driving. A “Preliminary Statement of Policy Concerning Automated Vehicles,” a nonbinding document published in May 2013 by NHTSA and attributed to the entire US Department of Transportation, expresses hope for the long-term benefits of automated vehicles, describes degrees of automation in a narrative style, and identifies relevant research projects initiated by NHTSA (2013). However, it also states that “detailed regulation” of “self-driving vehicle technologies” is not feasible in light of their “rapid evolution and wide variations.” It further advises US states not to “authorize the operation of self-driving vehicles for purposes other than testing” (NHTSA 2013).

Figure 5 shows the US states in which (as of April 2014) there are or were legislative bills on automated vehicles (Weiner and Smith 2014). The states of Nevada, Florida, California and Michigan as well as the District of Columbia have

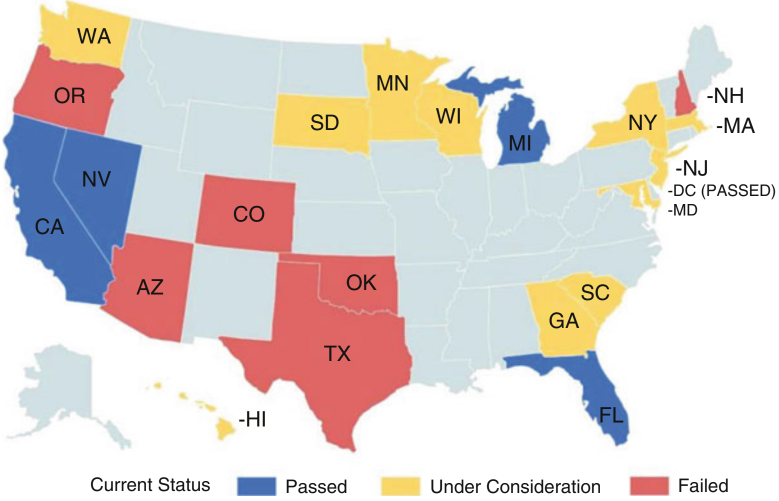


Fig. 5 Status of state legislation on automated driving

enacted such bills, and their regulatory approaches agree only in part. Each regime envisions a human driver of an automated vehicle and clarifies that only vehicles that can operate without the active “control and monitoring” of this person are considered automated or “autonomous.” In contrast to the vehicle automation levels developed by the Germany Federal Highway Research Institute (BASt), SAE International and NHTSA, these regimes differentiate only between automated and nonautomated vehicles (with the line falling roughly between SAE levels 2 and 3). The laws also specify requirements for automated vehicles and their drivers, and some codify a common law rule on the original manufacturer’s limited liability in cases where a vehicle or its parts are modified. Nevada and California also directed their road traffic authorities to establish specific regulations, whereas Florida and Michigan required their agencies to produce only reports on vehicle automation. Michigan (after publication of NHTSA’s policy statement) limited approval only to the operation of automated vehicles for experimental purposes; California’s Department of Motor Vehicles, on the other hand, is enacting rules that, pursuant to the state’s automated driving statute, also anticipate the general consumer operation of such vehicles.

Although these legislative acts are of considerable symbolic importance and practical relevance, they should not be overstated. In the US states that do not have such statutes, operation of certain automated vehicles is not necessarily prohibited and, depending on the context, probably permitted (Smith 2012b). The relevant authorities in these states likely already have authority to enact similar rules without a statutory basis specific to vehicle automation, and some are in contact with key firms. Furthermore, important questions, such as how the human driver must behave during automated operation, remain unanswered even in these recent statutes. Legal uncertainties persist everywhere, and basic principles of reasonability and prudence, which underlie both common law and state vehicle codes, will gradually evolve along with the relevant technologies. For these reasons, legislatures, regulators, and courts at both the federal and state levels will necessarily remain involved in the ongoing clarification of legal responsibility in the broadest sense.

4 Requirements for Driver Assistance Systems in the Context of "Ratings" and Legal Regulations

The requirements for the safety of vehicles, which are reflected in the specifications for the development of new vehicles, can be divided into the following three groups:

- Requirements due to the type approval regulations,
- Requirements of consumer organizations (e.g., Euro NCAP) and
- Manufacturers' internal requirements.

4.1 Type Approval Regulations

The approval of vehicle types and components is nowadays almost exclusively done on an international level through EU directives or regulations, drafted by the European Commission in Brussels (EC directives 2014), or through UN regulations (e.g., UN R 13-H), created by the United Nations Economic Commission for Europe (UNECE) in Geneva (UN-Regulations 2014).

What role the type-approval provisions play in the introduction of new driver assistance systems, depends on whether the functions of a driver assistance system fall in an area controlled by the type-approval or not. For example, in the field of lighting technology, there is a large number of requirements which must be followed in the type-approval, so that innovative lighting systems usually can only be approved (without exemption) if the respective legal approval conditions have been adjusted accordingly. Other driver assistance systems (e.g., in the field of informing or warning systems, such as displaying the speed limit in the vehicle or a blind spot monitoring function) can be readily introduced, because their functions do not fall, or only to a limited extent, in the area covered by type-approval regulations.

For electronically controlled assistance systems which intervene in vehicle components and functions that are safety-related a new approach has been taken, among other things, in the regulations on vehicle braking and steering systems. In annex 8 of the 13H UNECE regulation (on braking systems) and in annex 6 of the 79 UNECE regulation (on steering systems) generic requirements were defined, which must be complied with in the case of complex electronic vehicle control systems. These generic safety-related requirements go beyond pure performance requirements. This makes it possible for example to use the vehicle braking system for driver assistance functions like ESC, ACC or brake assistance and emergency braking assistance.

The process of stipulating new safety-enhancing vehicle systems and features of vehicle equipment with the objective of type-approval is often lengthy due to the necessary national and international coordination process. This is challenging especially with respect to the rapidly advancing technical development of new driver assistance systems. Improvement of type-approval regulations in terms of technical requirements can enhance the safety level of nearly all new vehicles.

4.2 Euro NCAP Requirements

With the requirements of the type-approval the legislative authorities only define minimum standards that must be complied with in order to obtain access to the market with a new vehicle model. Accordingly, all new vehicles introduced to the market must meet the legal requirements. However, these tests for type approval initially say nothing about the differences in the safety level of various type-approved vehicle models that have gained access to the market by meeting the legal requirements. This is the starting point where consumer associations come into play: Through their own (crash) testing, the different safety levels of already type approved vehicle models are determined, and published as nuanced consumer information. Because of the consumer organizations' objective, it becomes apparent that it would make little sense if the type-approval tests with their requirements were merely repeated in a consumer protection test. The result of such an approach would be "test passed" which does not allow for differentiation. A useful differentiation of products in terms of safety is often possible by tightening the test requirements and assessment criteria in comparison to the type-approval test. A consumer test must allow for a gradual differentiation of products, while the test according to legislation merely allows a binary differentiation between "passed" or "failed."

In the European New Car Assessment Programme (Euro NCAP) (2014), the gradual differentiation of test results is achieved by setting upper and lower performance limits. Within these limits assessment is based on the test result which is calculated by means of a linear interpolation ("sliding scale"). See also ► Chap. 11, "Test Methods for Consumer Protection and Legislation for ADAS."

The strategic objective of consumer organizations is to supplement the manufacturer's internal requirements by requirements, deemed useful from the consumer perspective. This is the purpose of ratings, which may increase the value of a vehicle in the eye of the consumer. A bad rating, or even the absence of a rating, may for example result in the effect that a vehicle cannot be procured by fleet customers due to internal standards. Thus, these ratings make the abstract value of "road safety" tangible, which justifies investment by the vehicle manufacturer. As this process has a strong economic impact, a transparent review process is mandatory.

Therefore, the requirements for vehicle safety, which are put forward by regulators and consumer organizations, have been and are a driving force for many innovations in automotive engineering. For example, many technical innovations in the field of passive pedestrian protection in the past were due to corresponding new requirements in European legislation. However, this example also reveals that requirements of the legislators and consumer organizations were significantly related more to the field of passive vehicle safety in the past. Innovations in the field of active safety and driver assistance systems, such as the case with ESC, which has demonstrably brought about significant safety benefits in real accident situations, have emerged thanks to the creativity and efficiency of the automotive industry and their suppliers – though in this example consumer information has led to the rapid spread of the ESC systems in almost all vehicle classes.

In the past decade, emergency braking systems that affect certain types of accidents by automatic braking, or animate the driver to brake by issuing appropriate warnings, have become increasingly important. Euro NCAP has responded to this dynamic, and, in 2014, released a test method for the first emergency braking systems, which are designed to influence collisions in longitudinal traffic. It has further been announced that from 2016 onwards emergency braking systems for pedestrian protection will also be evaluated, see, e.g., ▶ Chap. 11, "Test Methods for Consumer Protection and Legislation for ADAS."

A transparent assessment process with a resulting market value of the safety effect allows the comparison of different measures that serve the same goal. For example, protection of vehicle occupants through restraint systems may in principle be confronted with protection through emergency braking systems. The choice of means in the vehicle safety may therefore be left to the manufacturer.

Still, many significant innovations to increase road safety remain to be developed in the areas of active, integrated safety, and for driver assistance systems, which are frequently legally regulated only to a very restricted extent, and which have so far been tested and evaluated in a rudimentary form in consumer tests. This finding always brings new challenges to both the legislative authorities and the consumer protection organizations.

4.3 Manufacturers' Internal Requirements

Internal manufacturer vehicle safety requirements always contain the legal requirements of the relevant region in which the vehicle is to be sold, and frequently also selected requirements, which are known from the field of consumer tests. In addition, however, many car manufacturers also have their own in-house safety standards that go beyond the legal requirements or consumer tests, and in some cases also relate to further or other aspects of vehicle safety. These additional internal requirements of manufacturers are based, among other things, on manufacturers' own assessment regarding product liability, the presumed customer requirements, and thus the market strategy or findings from manufacturers' own accident research.

4.4 Beyond NCAP: Consideration of New Safety Features in Consumer Protection

Many important innovations in the field of passive safety have therefore been introduced to the market, because the corresponding test methods and assessment criteria, which serve as the basis for legislation or for assessment in consumer protection tests, have promoted or even expedited the development. In contrast, many safety systems in the field of active and integrated safety and in the field of Driver Assistance Systems are being developed by the creativity of engineers in the automotive industry alone – also with the expectation of economic benefit. Many experts estimate that the greatest potential for further improvement of road safety lies in these areas, and that they will continue to develop very dynamically.

Against this background – and because Euro NCAP wants to remain a major force in the assessment of safety-related vehicle systems in the future – Euro NCAP has initiated the development of a generic approach to the creation of new test procedures and assessment criteria for systems in the field of active and integrated safety, and in the field of driver assistance (Seeck 2007). In Euro NCAP, this activity is referred to as "Beyond NCAP." A potential Beyond NCAP assessment concept could be complementary to the known crash test review process and used additionally. The objective of developing a Beyond NCAP rating method is to define a flexible, transparent and predictable process that is able to identify and label innovations by means of a safety-assessment as shortly as possible after introduction to the market. This assessment gives the enhanced safety of a new function a market value, as classical assessment does.

Previously, Euro NCAP has implemented both the specification of the assessment process and performed the assessment (see Fig. 6, left side). The vehicle

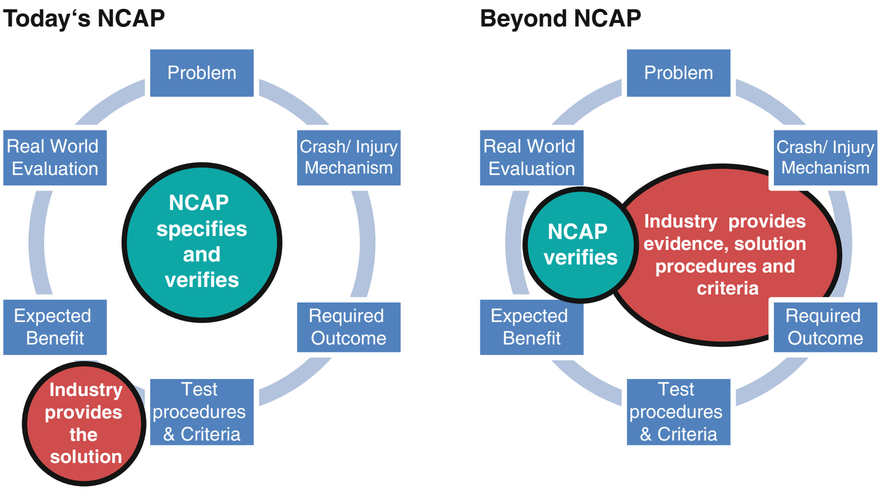


Fig. 6 Comparison of the current valuation method with the beyond NCAP method

manufacturer has "only" offered a technical solution, which – if it was positively evaluated at Euro NCAP and the assessment method was correct – has proved beneficial in real world accidents.

According to the Beyond NCAP concept, the vehicle manufacturer should not only develop a new safety system and introduce it to market. The manufacturer should rather provide scientifically reliable data, with which he reveals the expected benefits for real accidents, and propose a test method, with which the new safety system can be tested and evaluated. Euro NCAP, in this case, only assumes the role of verifying all information provided (see Fig. 6, on the right), to perform an assessment on this basis.

Through a robust Beyond NCAP rating method, which complements the existing Euro NCAP rating method, new safety systems can be evaluated faster and better, and can facilitate marketing by an independent quality seal. Fundamental prerequisite for a functioning Beyond NCAP method is, trust and partnership between Euro NCAP and the industry.

Introduction of New Test Procedures The aim of Euro NCAP in the Beyond NCAP process apart from producing incentives for the introduction of new safety features is essentially the continuous improvement of the classical assessment method. This is achieved by collecting and evaluating submitted information and by reviewing the proposals on assessment methods. As soon as an appropriate maturity of a safety system is reached, this information is then used for the development of new specific test procedures and the identification of their importance within the rating. The first successful implementation of the Beyond NCAP process is the introduction of tests for emergency braking carried out in 2014, after the first emergency braking system was successfully taken into account by the Beyond NCAP process in 2010. Taking emergency braking into account for the vehicle safety rating – compared to Beyond NCAP award – promises a significantly higher market value, because this new assessment immediately flows into the score of the so-called "star" rating. This further increases the requirements for a very good assessment. A vehicle manufacturer that fails to invest in new safety technologies in the future will fall behind in terms of star-rating compared to competitors.

The introduction of new safety features usually spread down from the premium segment to the vehicles of the mass market. This can take some years to happen. Therefore the mandatory introduction of rating criteria can systematically disadvantage mass market manufacturers. To mitigate this disadvantage, Euro NCAP has introduced the concept of "fitment rates": For functions newly introduced in the assessment, no standard use is required within the first years, but an equipment rate, which is somewhere between 50 % and 70 %. In 2016, the principle of "fitment rate" will be replaced by "dual rating." With dual rating, only the safety systems of a vehicle model are considered for the basic rating that are marketed as standard in Europe (EU-28) to 100 %. In addition, however, the manufacturer has the option to get a second rating for a vehicle model including the safety systems that are available only as an option, and thus at extra price. Thus, Euro NCAP offers the vehicle industry an opportunity to market expensive safety equipment at additional cost for price-sensitive models and/or markets, and to use the second, so-called Euro NCAP "safety pack" rating for this purpose. However, the use of "safety pack" rating is subject to the following requirements which must be observed by the vehicle manufacturer:

1. The potential safety systems, which are taken into account in a "safety pack" rating, are set by Euro NCAP.
2. The vehicle manufacturer agrees to always use the two assessments (basic rating and "safety pack" rating) in advertising.
3. For the safety package-rating, a minimum equipment rate is required by Euro NCAP to be achieved by the vehicle manufacturer through appropriate marketing measures over the production time of the vehicle model.

With the instruments Beyond NCAP, "fitment rates" and, from 2016, "dual rating," Euro NCAP has found a solution with which the assessment can be continuously adapted to technical progress. This is meant to achieve a larger and faster market penetration of innovative safety systems (by stimulated demand and thus higher production numbers and accompanied by lower manufacturing costs).

5 Conclusion

In summary, it can be ascertained that the systematic classification of driver assistance systems according to their mode of operation proposed here offers a comprehensive structure that takes into account technical, legal and behavioral aspects. According to this understanding, the legal view on continuously automating systems (operation mode B), considered in depth in this chapter, reveals inconsistencies. These take effect as soon as a degree of automation is achieved that aims at dismissing the driver, even if only temporarily, from the (mental) function of traffic monitoring in the driver-vehicle control loop. For the corresponding degrees of automation (beginning from level 3), contradictions with established law arise. As far as we are aware, only in some states of the United States laws have been adopted so far that allow the operation of such vehicles, with specific limitations and differences.

Vehicle systems that have a positive impact on vehicle safety can be made mandatory by type-approval regulations in the area of the UNECE. These systems are then fitted to almost all new vehicles. However, the instrument of type-approval is comparably inflexible, and the adoption of new or amended legislation is a lengthy process, so that it is challenging to react appropriately on the rapid technological progress in the field of driver assistance systems. In these cases, rating scheme from the consumer protection point of view – such as EuroNCAP – have the advantage of making the safety level of vehicles transparent for consumers. The active and integrated vehicle safety is subject to the constantly evolving area, called "Beyond NCAP." This field already significantly affects the design as well as the equipment of vehicles with driver assistance systems by providing stimulus. “Beyond NCAP” thus plays an important role in the consideration of framework conditions for the development of driver assistance systems.

Under the aspect of increasing vehicle automation, a need for research on the continuously automating operation mode B can presently be identified in four key areas: In the first place this involves the driver’s performance in the cooperation with driver assistance systems; technologically, the vehicle and infrastructure-related requirements for safe operation, and in general the societal acceptance of this development as such.

5.1 Need for Research on Human-Machine Interaction

Need for research on human-machine interaction is of paramount importance. This finding is not surprising in view of the initially described task division between the human and the system, which will be part of the systems until very high levels of automation are reached. This research has, at the same time, significant effect on the area of system development: For example, the question of hazard prevention measures necessary under the aspect of product liability law can be answered on the basis of research in the field of human-machine-interaction. This includes the field of possibly foreseeable misuse of systems. This also results in a greater predictability of legal decisions in the face of product liability risks to be met once these systems are introduced to the market. In addition, research on human-machine interaction can provide a basis for the estimation of the safety effect systems take – a point of view which plays a special role in the context of the societal acceptance of this development.

5.2 Need for Research on Technical Safeguarding Strategies

A need for research on technical safeguarding strategies is also of paramount importance in order to make automation technologies available with the highest possible safety and reliability. This raises the question of how systems which do not require the driver as a fall-back system can be made safe and reliable. This concerns the continuously automated systems from automation level 3 and higher. More specifically, this involves the question of how future test procedures can be designed to demonstrate technical reliability and availability.

5.3 Need for Research to Identify the Necessary Measures in the Road Traffic Infrastructure

It must also be identified, whether and which measures are needed in road traffic infrastructure to enable certain automated functions. Thereby, it must be considered that the automated systems proposed today are very strongly vehicle-based, and – if at all, and in comparison with the current transport system – provide few additional requirements in regard to infrastructure. However, the question of which measures in detail may be safety-relevant, require further investigation.

5.4 Need for Research on Societal Acceptance of Automated Systems in Road Traffic

It is already apparent today that automated systems, on the one hand, can considerably increase traffic safety. It is, however, equally apparent that these systems introduce a previously unknown automation risk to road traffic. This automation risk in vehicle control may be clearly less significant than the benefits for road safety; however, the question arises whether this new risk will be socially accepted. The analysis of the current legal situation under German law shows that the current legal system is, in some specific cases, not designed to address this novel form of automatic vehicle control properly (see above Sect. 2.2.1). Legal changes currently being implemented in some US states – although sometimes very limited – create a legal framework for such systems; however, they rely, potentially wrongfully, on unattainable technical perfection, thus on the absence of the automation risk. The question of acceptance is therefore significant, and requires parallel consideration.

References

Albrecht F (2005) Driver assistance systems for influencing speed. German Car Law [Deutsches Autorecht] (DAR) (4):186–198

Bartels A (2012) System description of automated driving functions. Basics, technical design and requirements [Grundlagen, technische Ausgestaltung und Anforderungen]. BASt research report: FE 88.0006/2009. Published as part 2 of the BASt project group report (BASt, issue F 83). Legal consequences of an increase in vehicle automation. Wirtschaftsverlag NW, Bergisch Gladbach (project 1), pp 27–44

Bewersdorf C (2005) Approval and liability for driver assistance systems in road traffic. Duncker & Humblot, Berlin

BGH (German Federal Court of Justice) judgment of 16 June 2009, ref. no.: VI ZR 107/08. On liability of vehicle manufacturers for erroneous activation of airbags. Available in the Federal Court of Justice ruling database using file number or date of judgment. http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288. Accessed 1 Feb 2014

Bouska W, Leue A (2009) German road traffic code text with explanatory notes and general administrative regulations on the road traffic regulations and traffic regulations of the GERMAN FEDERAL POLLUTION CONTROL ACT, 23rd edn. Jehle Publications, Heidelberg

EC Directives and Regulations. http://ec.europa.eu/enterprise/sectors/automotive/documents/directives/motor-vehicles/index_en.htm. Accessed 1 Feb 2014

ESoP (2007) Recommendation of the European Commission on safe and efficient on-board information and communication systems (2006). Consolidation of the European statement of principles on human-machine interface. Official Journal of the European Union, 2007 L 32/200 (2007/78/EC)

Euro NCAP. http://www.euroncap.com/home.aspx. Accessed 1 Feb 2014

Gasser TM (2008) Driver's resilience – legal situation and solution proposal in view of increasing electronic possibilities. Road Traffic Law [Straßenverkehrsrecht] (SVR) (issue 6):201–206

Gasser TM (2009) Legal aspects of the introduction of driver assistance and information systems [Rechtliche Aspekte bei der Einführung von Fahrerassistenz- und Fahrerinformationssystemen]. Traffic Accid Automot Eng [Verkehrsunfall und Fahrzeugtechnik] (VKU), (issue 07–08):224–231

Gasser TM, Arzt C, Ayoubi M, Bartels A, Bürkle L, Eier J, Flemisch F, Häcker D, Hesse T, Huber W, Lotz C, Maurer M, Ruth-Schumacher S, Schwarz J, Vogt W (2012) Legal consequences of an increase in vehicle automation. Consolidated final report of the BASt project group: "legal consequences of an increase in vehicle automation", part 1(issue F 83). Wirtschaftsverlag NW, Bergisch Gladbach

Geier v. American Honda Motor Co (2000) 529 U.S. 861, 120 S. Ct. 1913

German Law Act Approving the Vienna Convention on Road Traffic (1977) Federal Law Gazette II, 1977, p 810. The convention text is therein in English, French and German. The convention in original version and all authentic languages (English, Chinese, French, Russian and Spanish) in the United Nations treaty collection. "B Road Traffic" no. 19. https://treaties.un.org/pages/CTCTreaties.aspx?id=11&subid=B&lang=en. Accessed 1 Feb 2014

Gold C, Damböck D, Lorenz L, Bengler K (2013) "Take over!" How long does it take to get the driver back into the loop? Proc Hum Factors and Ergon Soc Annu Meet 57(1): 1938–1942

Kanz C, Marth C, von Coelln C (2013) Liability in the case of co-operative traffic and driver assistance systems [Haftung bei kooperativen Vekehrs- und Fahrerassistenzsystemen]. Research report for FE project 89.0251/2010. http://www.bast.de/cln_032/nn_42642/DE/Publikationen/Download-Berichte/downloads/haftung-assistenzsysteme.html. Accessed 29 Jan 2014

König P (2013) In: Hentschel P, König P, Dauer P (eds) Road traffic law commentary [Straßenverkehrsrecht Kommentar], 42nd edn. CH Beck, Munich

Kraiss K-F (1998) User-friendly automation – fundamentals and implementation concepts. Automat Eng [Automatisierungstechnik] 46(10):457–467

Lenz T (2009) On liability of vehicle manufacturers for erroneous activation of airbags [Zur Haftung des Fahrzeugherstellers für die fehlerhafte Auslösung von Airbags]. Prod Liabil Int [Produkthaftung International, (PHi)]: 196–198

Maurer M (2009) Design and testing of driver assistance systems. In: Winner H, Hakuli S, Wolf G (eds) Handbook of driver assistance systems [Handbuch Fahrerassistenzsysteme], chapter 5, 1st edn. Vieweg + Teubner, Wiesbaden, pp 43–52

National Highway Traffic Safety Administration (NHTSA) (2013). Preliminary statement of policy concerning automated vehicles. Published online on 30 May 2013. http://www.nhtsa.gov/About+NHTSA/Press+Releases/U.S.+Department+of+Transportation+Releases+Policy+on+Automated+Vehicle+Development. Accessed 21 Jan 2014

Press Release: U.S. Department of transportation announces decision to move forward with vehicle-to-vehicle communication technology for light vehicles, February 2014. http://www.nhtsa.gov/About+NHTSA/Press+Releases/2014/USDOT+to+Move+Forward+with+Vehicle-to-Vehicle+Communication+Technology+for+Light+Vehicles. Accessed 7 Apr 2014

Recommendation of the European Commission on Safe and Efficient on-Board Information and Communication Systems (2006) Consolidation of the European statement of principles on human-machine interface. Off J Eur Union 2007 L 32/200 (2007/78/EC)

SAE International "On-Road Automated Vehicles Standards Committee" (2014) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. Document no. J3016_201401 under development since 16 July 2012. http://www.sae.org/works/documentHome.do?comtID=TEVAVS&docID=J3016_201401&inputPage=wIpSdOcDeTaIlS. Accessed 21 Jan 2014 and: Bryant W S (2014) SAE Levels of Driving Automation. http://cyberlaw.stanford.edu/loda. Accessed 21 Jan 2014

Seeck A (2007) The future of vehicle safety assessment for type approval and Euro NCAP. In: 6th international VDI conference on vehicle safety 18–19 October 2007 – innovative vehicle occupant and third party protection, Berlin

Smith BW (2012a) Autolaw 3.0, Transportation research board workshop on road vehicle automation. http://onlinepubs.trb.org/onlinepubs/conferences/2012/Automation/presentations/WalkerSmith.pdf. Accessed 7 Apr 2014

Smith BW (2012b) Automated vehicles are probably legal in the United States. 1 Tex A & M L Rev (2014). http://cyberlaw.stanford.edu/publications/automated-vehicles-are-probably-legal-united-states. Accessed 7 Apr 2014

Smith BW (2013a) Regulatory approaches. Transportation research board workshop on road vehicle automation, July 2013. http://bit.ly/1dUIvvn. Accessed 7 Apr 2014

Smith BW (2013b) Proximity-driven liability. Geo Law Rev (2014) 102. http://ssrn.com/abstract=2336234. Accessed 7 Apr 2014

UN/ECE Regulations. http://www.unece.org/trans/main/welcwp29.htm. Accessed 1 Feb 2014

Weiner G, Smith B W (2014) Automated driving: legislative and regulatory action. https://cyberlaw.stanford.edu/wiki/index.php/Automated_Driving:_Legislative_and_Regulatory_Action. Accessed 07 April 2014

Wessels J, Beulke W (2012) German criminal law general part (margin no. 93), 42nd edn. C.F. Müller, Heidelberg

Williamson v. Mazda Motor of America, Inc. (2011) 131 S. Ct. 1131

Driver Assistance and Road Safety

4

Matthias Kühn and Lars Hannawald

Contents

Abstract

In order to make any statements about the effect of advanced driver assistance systems (ADASs) on road safety, it is important to understand the accidents that happen. In 1970, 21,332 people were killed on German roads (West Germany). In 2013, 3,339 people were killed on the roads. The number of motor vehicles in Germany increased during this time: from 16.8 to 53.8 million. In order to get a better understanding of accident statistics, a shift of focus is required from the general view obtained from a country's accident statistics to detailed analysis of the accidents that happen. This field extends from the representative surveys of

M. Kühn (✉)
Unfallforschung der Versicherer, Gesamtverband der Deutschen Versicherungswirtschaft e.V., Berlin, Germany
e-mail: m.kuehn@gdv.de

L. Hannawald
Verkehrsunfallforschung an der TU Dresden GmbH, Technische Universität Dresden, Dresden, Germany
e-mail: hannawald@vufo.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_4

the German Federal Statistical Office (Destatis), based on road accident reports, to the in-depth analyses of different stakeholders on road safety. Accident analysis was carried out for cars, trucks, buses, and powered two-wheelers based on the German In-Depth Accident Study (GIDAS) and the German insurers' accident database (UDB). The safety potential of advanced driver assistance systems (ADASs) can be ascertained in a variety of ways. For the results here, an alternative, "what if" method, was used in order to quantify the effect of different generic ADASs for cars, trucks, and buses.

1 Accident Statistics

Before we make any statements about the effect of advanced driver assistance systems (ADASs) on road safety and their potential for the future, it is essential to know about and understand the accidents that happen. The accident patterns identified should then be addressed by the specific functions of the ADAS. In order to do this, a shift of focus is required from the general, representative view obtained from a country's accident statistics to detailed analysis of the accidents that happen. That, in turn, requires surveys of varying degrees of quality to collect accident data, resulting in accident statistics that provide specific information. This field extends from the representative surveys of the German Federal Statistical Office (Destatis), based on road accident reports, to the in-depth analyses of different accident researchers working on road safety. In Germany, that means using, above all, the German In-Depth Accident Study (GIDAS) and the German insurers' accident database (UDB).

GIDAS stands out in terms of its level of detail and the usefulness of its data. The data collected at accident locations for a representative selection of road accidents in a specific region is unrivaled. This joint research project of the German Federal Highway Research Institute (BASt) and the Forschungsgemeinschaft Automobiltechnik (FAT), a research alliance of all German car makers, is thus extremely useful for accident research. The UDB, on the other hand, is based on the claims data of insurers. The data is based on a representative selection of motor third-party liability claims involving damage costs of at least 15,000 euros and at least one case of injury. However, the UDV (German Insurers Accident Research) does not investigate these cases at the accident locations. Consequently, certain kinds of statement cannot be made about the vehicle, etc., or at least only with reservations. This dataset thus describes more severe claims and cannot be compared for all issues with the official road traffic accident statistics or GIDAS.

1.1 Accident Statistics in Germany

A glance at the accident statistics in Germany over recent decades tells us that the number of fatalities has declined almost continuously (see Fig. 1).

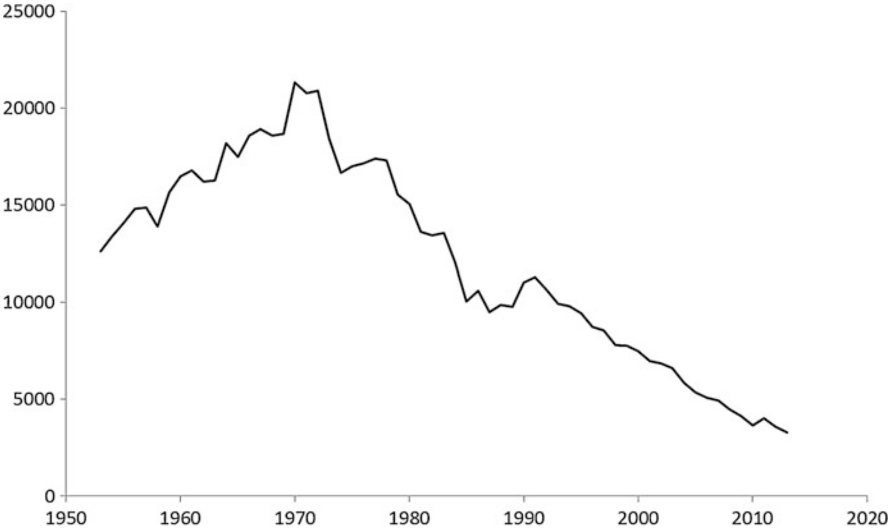


Fig. 1 Number of fatalities on the roads in Germany over time (Destatis 2013)

Whereas 21,332 people were killed on German roads (West and East Germany together) in the year 1970, by 2013 this had fallen to 3,339. The number of accidents involving injury fell from 377,610 in 1970 to 291,105 in 2013 (Destatis 2013). When we also consider the increase in the distance traveled over the same period, the improvements in these figures are even more impressive. The number of motor vehicles in Germany increased from 16.8 million in 1970 to 53.8 million in 2012. And the distance traveled by motor vehicles almost trebled from just under 251 bn km in the year 1970 to 719.3 bn km in 2012. Most of these are passenger cars, of which around 43 million were registered in 2012, traveling a total distance of 610.1 bn km and thus accounting for around 85 % of the total distance traveled by all motor vehicles.

The distribution of fatalities for the year 2012 gives an initial indication of where to focus efforts to further improve road safety in Germany (see Fig. 2).

It is clear from this that 60 % of fatalities occurred on roads outside built-up areas. Over a quarter of these cases occurred as a result of collisions with trees. Just over 1,000 people were killed on the roads in built-up areas. These were primarily unprotected road users such as pedestrians and cyclists. The remaining 387 fatalities occurred on freeways (autobahns). If the fatalities are analyzed by type of road user, we see that around half of the 3,600 fatalities in 2012 were car occupants (1,791), roughly 20 % were on motorcycles (679), and around 25 % were nonmotorized, unprotected road users (520 pedestrians and 406 cyclists).

The distribution of fatalities by accident type in 2012 is shown in Fig. 3. Driving accidents (where the driver loses control) are the most common type of accident resulting in fatalities, followed by accidents in longitudinal traffic and turning-into/crossing accidents.

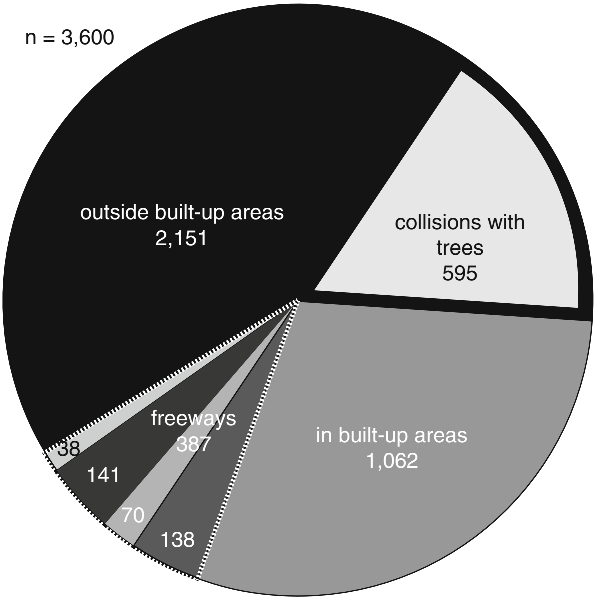


Fig. 2 Fatalities on German roads by location in 2012 (Destatis 2012)

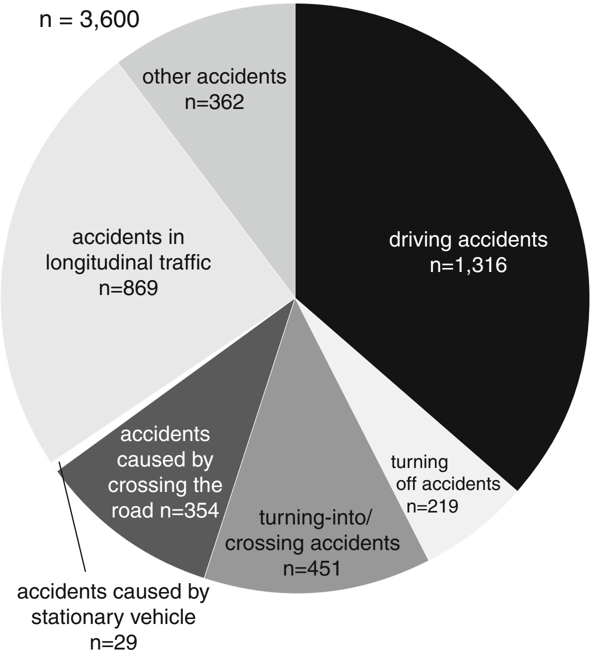


Fig. 3 Fatalities on German roads by accident type in 2012 (Destatis 2012)

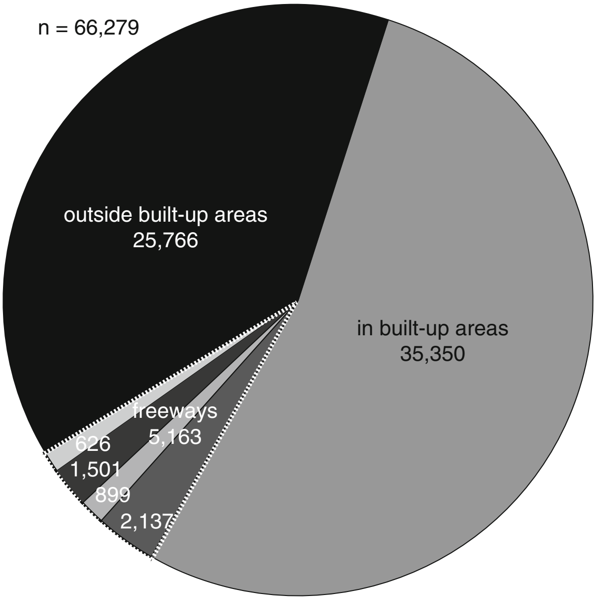


Fig. 4 Seriously injured accident victims in Germany by accident location in 2012 (Destatis 2012)

In the case of driving accidents, in particular, which are categorized as such in police accident reports when the driver loses control of the vehicle as a result of driving at a speed that is inappropriate for the course, cross section, inclination, or condition of the road, it is possible to see the importance of speed as a factor contributing to accidents.

If we look at the situation in relation to serious injuries on the roads in 2012, the picture changes (see Fig. 4). Over half of all seriously injured road users suffered their injuries in accidents in built-up areas, around two-fifths of injuries occurred in accidents outside built-up areas, and less than 10 % happened on freeways.

In addition to the location, which is an accident-specific parameter, various participant-specific parameters can be analyzed. The age of the person responsible for the accident is an important parameter. The age distribution of the people primarily responsible for all accidents involving injury in 2012 is shown below (see Fig. 5).

It is clear from this that the 18–25 age-group is responsible for significantly more accidents than other age-groups. The low proportion of older people primarily responsible for these accidents can also be partially explained by the fact that older people drive less. Age risk groups for causing road accidents thus cannot be specified accurately unless adjustments are made to take into account the yearly distances traveled by these age-groups. For example, the number of people responsible for an accident in each age-group as a ratio of the number of people in that age-group not responsible for an accident indicates the relative risk of someone in that age-group being responsible for an accident. Assuming that the distance

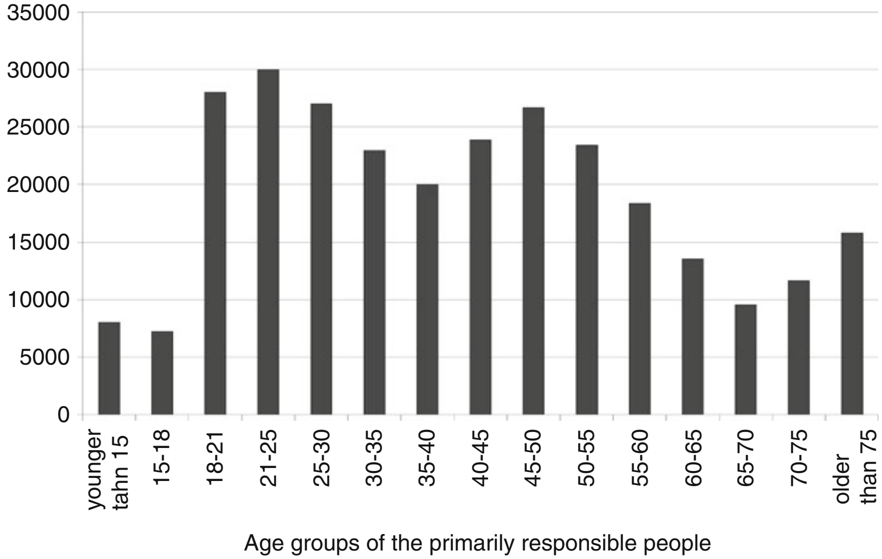


Fig. 5 People primarily responsible for accidents involving injury by age-group in 2012 (Destatis 2012)

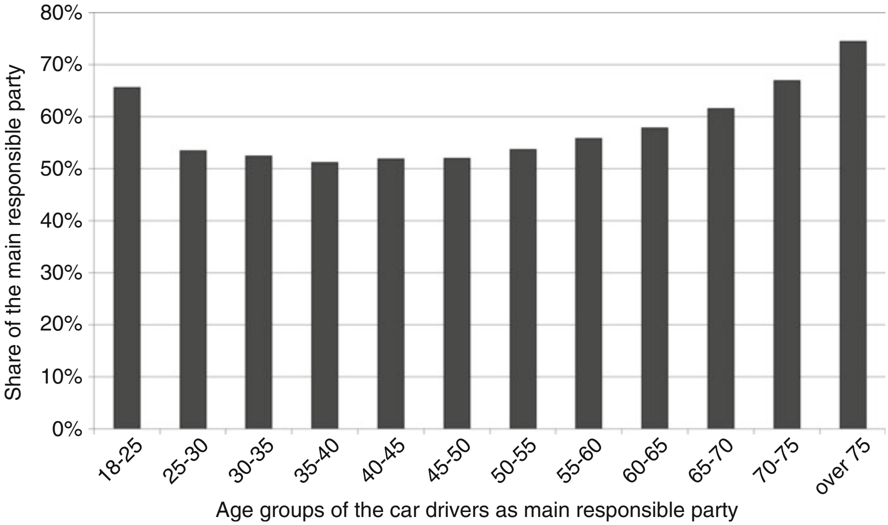


Fig. 6 Percentage of car drivers by age-group have been primarily responsible for an accident involving injury (Hannawald 2014)

traveled by those responsible for an accident does not differ from that of those not responsible for an accident in a specific age-group but that it does differ between the age-groups, the effect of distance traveled can be eliminated. The age risk groups are then shown in Fig. 6.

1.2 Forecast

The major trends affecting individual motorized transport – driverless cars, electric vehicles, and demographic change – will have a significant impact on the development of road safety in the future. Forecasts of accident figures for the years 2015 and 2020 assume a reduction in the number of accidents involving injury to 279,000 and then 234,000 (Meyer et al. 2012). On the basis of this forecast, 3,212 fatalities are expected in 2015 and 2,497 in 2020. A significant drop is thus expected in the number of car occupants and pedestrians involved in accidents in 2015 and 2020. The same applies to the number of fatalities and casualties with serious injuries. The number of pedestrians and car occupants suffering serious injuries is falling significantly as a proportion of all road users suffering serious injuries. On the other hand, the proportion of cyclists and, in particular, motorcyclists suffering serious injuries is rising. This is not due to an increase in the absolute numbers; instead, it is the result of differences in the rates at which casualties among different types of road users are declining.

In order to be able to understand accidents even better in the future and target measures more effectively, it is necessary to look at severely injured road users as well as fatalities. One consequence of improved vehicle technology and better rescue services is that although the number of fatalities is falling, the number accident victims surviving with severe injuries is rising. Thus, the quality of road safety work cannot be measured by the reduction in the number of fatalities alone. The results of studies suggest that around 10 % of the road users officially recorded as seriously injured had life-threatening injuries. For the year 2012, that amounts to 6,000–7,000 polytraumatized patients (Malczyk 2010). The current definition of seriously injured used in the official statistics is based solely on the criterion of whether a patient receives inpatient hospital treatment for at least 24 h. This large group is thus very heterogeneous and does not permit any statements to be made about those with life-threatening injuries. Consequently, efforts are under way in Germany to introduce a subgroup for accident victims with very severe injuries based on the Maximum Abbreviated Injury Scale (MAIS). This would permit targeted accident analysis, the development of subgroup-specific preventative measures, and more precise estimation of the economic costs of serious road accidents, for example. In a similar vein, the Europe-wide harmonization of the definition of seriously injured road accident victims, also on the basis of MAIS (Auerbach 2014), is expected in 2015.

1.3 Accident Statistics by Vehicle Type

To gain a better understanding of the accidents of the various vehicle types, it helps to shift the level of focus and make use of in-depth accident data. Table 1 shows an overview of the kind of accidents for different vehicles on the basis of a GIDAS analysis (Hannawald 2014). Differentiating between vehicle types in this way reveals that they have their own specific accident patterns. The proportion of

Table 1 Distribution of the kind of accidents by vehicle type based on an analysis of GIDAS (Hannawald 2014)

Accidents involving	Cars	Trucks	Buses, trams	Motorcycles	Bicycles
Collision with another vehicle that is pulling away, stopping, or standing in stationary traffic	4.1 %	5.0 %	3.7 %	3.7 %	5.0 %
Collision with a vehicle moving ahead or waiting	16.1 %	27.4 %	5.7 %	11.7 %	2.9 %
Collision with another vehicle traveling in the same direction next to the first vehicle	4.7 %	10.6 %	7.6 %	7.4 %	5.3 %
Collision with an oncoming vehicle	7.5 %	9.4 %	5.1 %	6.8 %	5.5 %
Collision with another vehicle that is turning into or crossing a road	39.3 %	27.2 %	34.2 %	38.4 %	62.2 %
Collision between a vehicle and pedestrian	11.0 %	6.9 %	23.1 %	2.6 %	4.3 %
Collision with an obstacle on the road	0.2 %	0.3 %	0.2 %	1.2 %	1.0 %
Coming off the road to the right	7.9 %	6.1 %	0.8 %	8.2 %	1.6 %
Coming off the road to the left	6.4 %	3.9 %	0.3 %	4.0 %	1.0 %
Accidents of a different type	2.9 %	3.1 %	19.1 %	15.9 %	11.2 %

collisions with vehicles in front, either moving or stationary, is significantly higher for trucks (27.4 %) than for other vehicle types. For buses and trams, collisions with pedestrians are much more common than they are for other vehicle types. Collisions with other vehicles that are turning into or crossing a road are far more common than all other accident types for cyclists, and bicycles are much more likely than other vehicle types to be involved in this kind of accident.

The insurers' accident database (UDB) was used to carry out in-depth analyses of the accidents involving different vehicle types. The UDB is based on motor third-party insurance damage claims and provides clearly more detailed information than the German federal statistics. It is comparable with GIDAS, although it is less useful in some respects because no analysis is carried out at the scene of the accident.

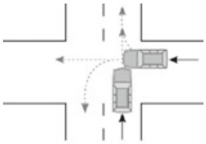
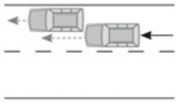
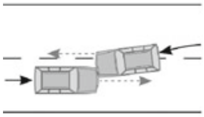
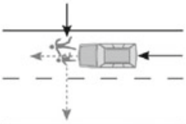
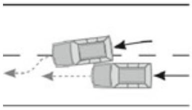
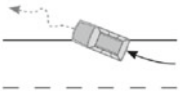
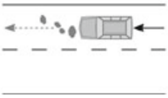
1.3.1 Cars

Using 1,641 damage claims in the UDB as a basis, accidents involving cars were subdivided by parameter "kind of accident" (see Fig. 7).

Over 50 % of the accidents involving cars are collisions with a vehicle that is turning into or crossing a road or collisions with a vehicle that is pulling away, stopping, or standing in stationary traffic.

Figure 8 shows the most common mistakes made by car drivers that lead to accidents. Turning, U-turn, reversing, and starting is the most common category, just ahead of ignoring the right-of-way. Not too far behind these come unadapted speed and insufficient safety distance from the vehicle in front. The category "other

Fig. 7 The most common accident scenarios involving cars in the insurers' accident database (UDB) (Hummel et al. 2011)

The most frequent accident scenarios $N_{\text{data pool}}$ = 136,954 [100%]		Percentage share
(1) Collision with another vehicle that turns into or crosses a road		34.5
(2) Collision with another vehicle: - which starts, stops or is stationary - moving ahead or waiting		22.2
(3) Collision with another oncoming vehicle		15.5
(4) Collision between vehicle and pedestrian		12.1
(5) Collision with another vehicle moving laterally in the same direction		6.9
(6) Leaving the carriageway to the right or left		6.3
(7) Collision with an obstacle in the carriageway		0.1

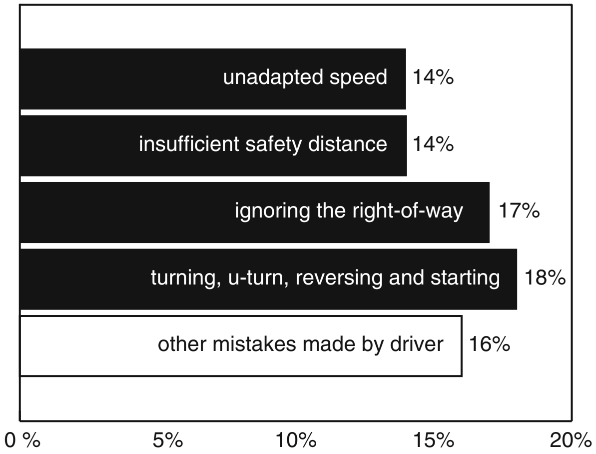


Fig. 8 The most common causes of car accidents involving injury (Destatis 2012)

mistakes made by driver" features prominently for all vehicle types. It is clear from both of these that the picture would be different if these causes were known and that there are limits to the usefulness of the national road accident statistics if we want to find out the causes of accidents. The information on unadapted speed, for example, is also worthy of further critical analysis here.

1.3.2 Trucks

Figure 9 shows the most common accident scenarios involving trucks.

The analyses show that around 50 % of the accidents are collisions with a vehicle traveling in the same direction next to or in front of the truck. Accidents involving vehicles that are turning into or crossing a road are the second most common accident type. Four hundred forty-three claims cases involving trucks were analyzed in detail here (Hummel et al. 2011). The most common accident causes were failure to drive at a safe distance from the vehicle in front and mistakes when turning, making a U-turn, reversing, and starting (see Fig. 10).

1.3.3 Buses

The analysis of accidents involving buses revealed that around 30 % of them were accidents in longitudinal traffic, 18 % were turning-off accidents, and 17 % were driving accidents, in which the driver lost control of the vehicle. When turning-into/crossing accidents are taken into consideration as well, which accounted for around 15 % of all accidents, it is clear that around 33 % of all accidents took place at junctions and intersections (UDV 2011).

The most common causes of accidents are shown in Fig. 11. As already mentioned, knowledge of the causes of the accidents in the largest group, the "other mistakes by driver" category, could change the picture and have a significant impact on the measures required.

Fig. 9 The most common accident scenarios for trucks in the insurers' accident database (UDB) (Hummel et al. 2011)

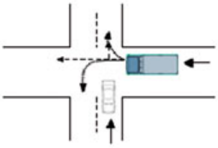
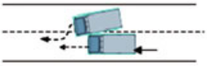
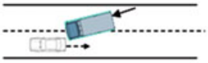
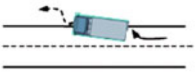
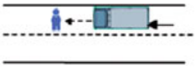
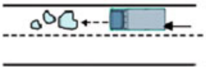
The most frequent accident scenarios $N_{\text{data pool}} = 18{,}467$ [100%]		Percentage share
(1) Collision with another vehicle that is: -moving ahead or is waiting -starts, stops, or is stationary		31.6
(2) Collision with another vehicle that turns into or crosses a road		22.3
(3) Collision with another vehicle moving laterally in the same direction		18.5
(4) Collision with another oncoming vehicle		14.3
(5) Leaving the carriageway to the right or left		5.1
(6) Collision between a vehicle and pedestrian		4.4
(7) Collision with an obstacle in the carriageway		0.4

Fig. 10 The most common causes of truck accidents involving injury (Destatis 2012)

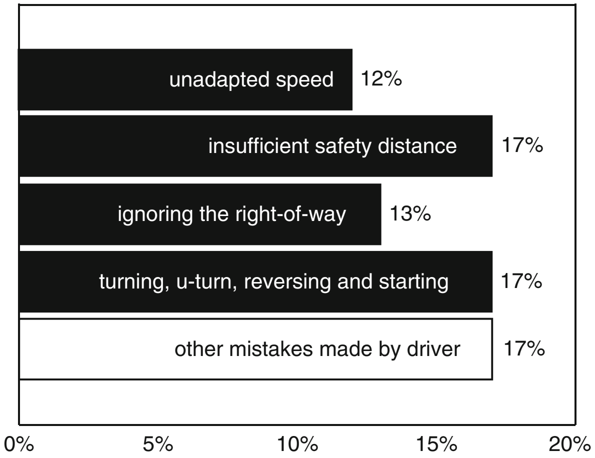
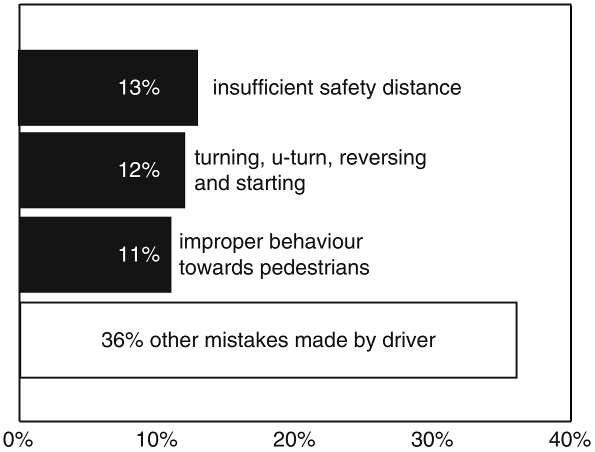


Fig. 11 The most common causes of bus accidents involving injury (Destatis 2012)



1.3.4 Powered Two-Wheelers (PTWs)

As road users, motorcyclists are extremely vulnerable. Motorcycle accidents, which often result in serious or fatal injuries for the rider, are caused by acceleration, speed, the narrowness of the motorcycle's silhouette, and errors of judgment on the part of both motorcyclists and other road users. It therefore makes sense to examine the accident data by looking more closely at who causes the accidents and who is involved (see Fig. 12). If the single-vehicle accidents and the accidents involving two road users caused by the rider of the powered two-wheeler are grouped together, it is clear that 51 % of all accidents involving no more than two road users were caused by the rider of the powered two-wheeler.

To identify the typical accident scenarios, we analyzed single-vehicle accidents and accidents involving two road users, subdivided on the basis of who primarily

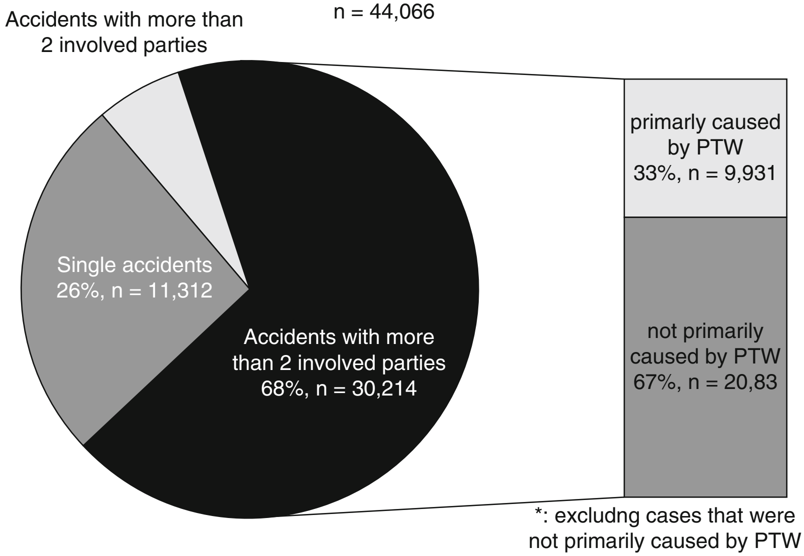


Fig. 12 Involvement in accidents of powered two-wheelers in Germany in 2012 (Destatis 2012)

caused them, as recorded in the insurers' accident database (UDB) (Hummel et al. 2011). The underlying accident material consists of 880 accidents involving powered two-wheelers.

The analyses of the single-vehicle accidents show that 56 % of them involved crashes when traveling straight ahead. Leaving the carriageway to the right (26 %) and left (12 %) were the second and third most common scenarios. These two scenarios were characterized by inappropriate speeds in bends and unfavorable weather conditions (see Fig. 13).

In accidents involving two road users and primarily caused by the powered two-wheeler rider, the most frequent scenario was a collision with an oncoming vehicle (41 %), followed by a collision with a vehicle traveling in the same direction (24 %) and a collision with a vehicle coming from the right (16 %). Further scenarios were a collision with a vehicle that was stationary, parking, or stopping for traffic (8 %) and a collision with a vehicle coming from the left (also 8 %) (see Fig. 14).

The analysis of the accidents involving two road users that were not primarily caused by the powered two-wheeler rider revealed that the most common accident scenarios were a collision with a powered two-wheeler coming from the left (32 %) and a collision with an oncoming powered two-wheeler (29 %). These were followed by a collision with a powered two-wheeler traveling in the same direction (20 %) and a collision with a powered two-wheeler coming from the right (17 %) (see Fig. 15).

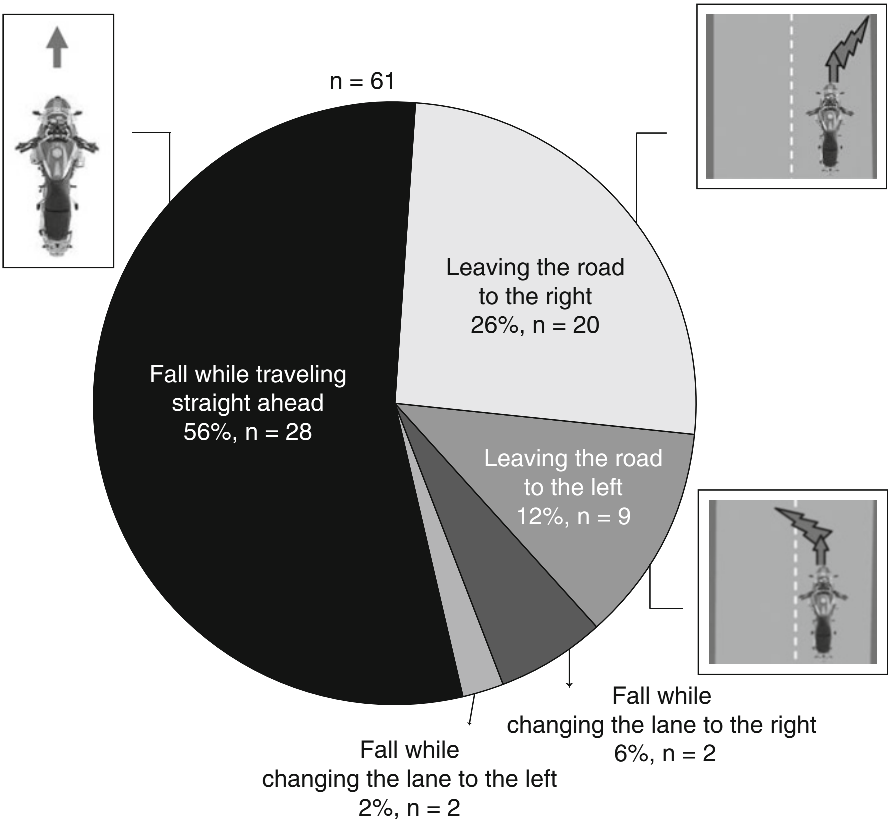


Fig. 13 Accident scenarios for single-vehicle accidents involving powered two-wheelers (Hummel et al. 2011)

2 Safety Potential of Advanced Driver Assistance Systems

Advanced driver assistance systems are electronic systems in the vehicle that are designed to help the driver to drive. The aim is often to make driving easier and improve safety or economy. This section focuses exclusively on the safety aspect.

There are direct links between the safety of vehicles and accident situations. The figure below shows the phases involved in an accident (see Fig. 16). It was produced by the European Automobile Manufacturers Association (ACEA). The idea is that every accident goes through the different phases, beginning with a “normal driving” phase in which the accident is not yet foreseeable for the driver but in which certain conditions, such as the length of time for which the driver has been driving, are already having an effect on the driver. This phase ends with the accident-triggering critical situation that precedes every accident. For example, the driver may be too late in noticing that the driver in front has braked or that a child

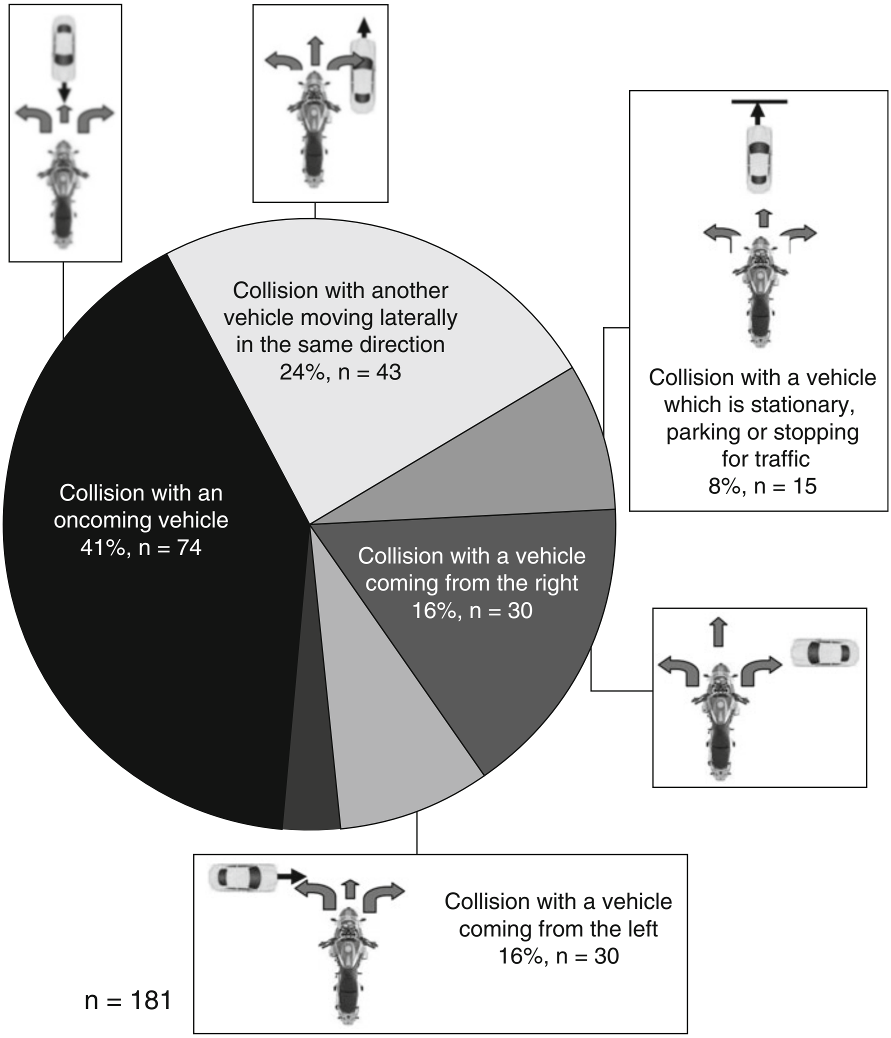


Fig. 14 Accident scenarios of accidents involving two road users and primarily caused by the powered two-wheeler rider (Hummel et al. 2011)

has run onto the road. This situation is followed by the danger phase. These two phases occur relatively frequently in everyday traffic and do not always result in an accident. The critical threshold of an accident is passed when the “point of no return” is reached and an accident becomes unavoidable. This is followed by the pre-collision phase, which may be relatively short, depending on the accident. The impact is followed by the “during collision” phase and ends with all road users involved coming to a standstill in the final situation of the accident. The greatest stresses – and thus the injuries of those involved – usually occur in this phase. The

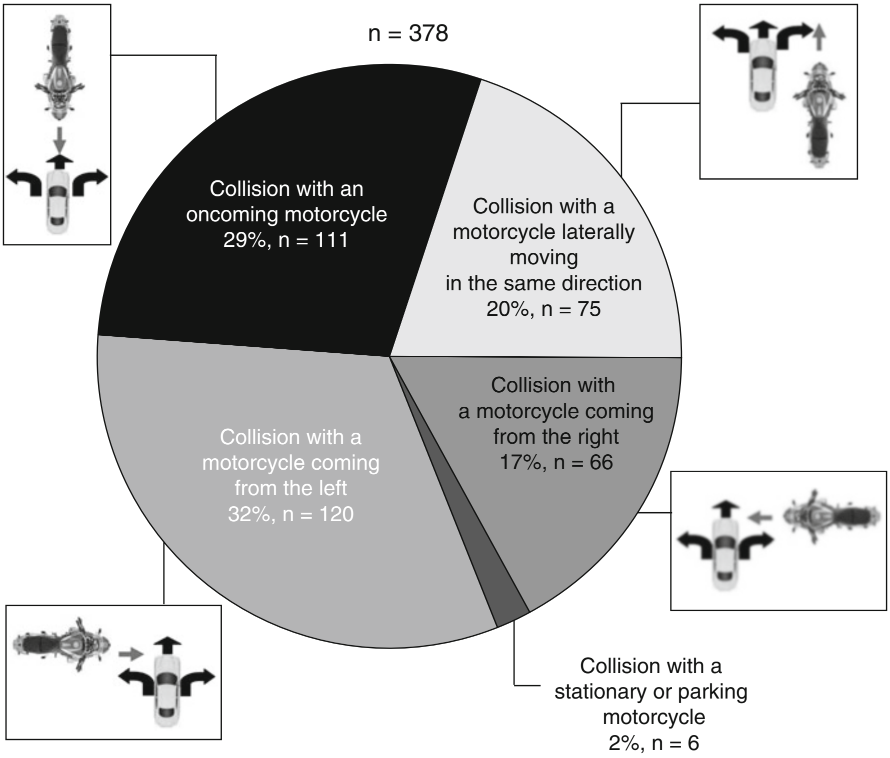


Fig. 15 Accident scenarios of accidents involving two road users and not primarily caused by the powered two-wheeler rider (Hummel et al. 2011)

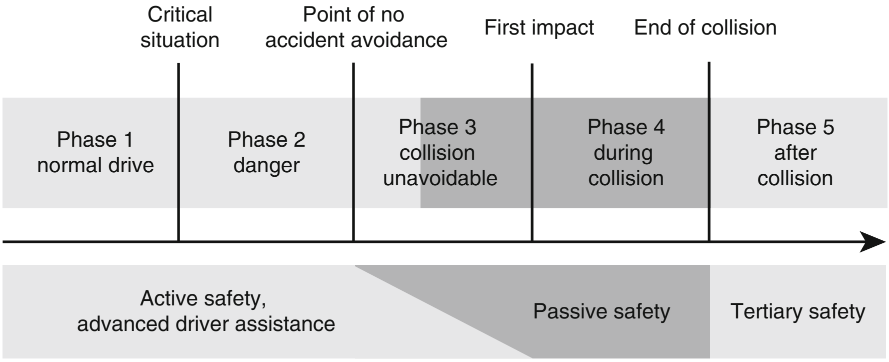


Fig. 16 Accident phases in the ACEA model and relevance of active safety and advanced driver assistance systems and passive and tertiary safety measures

phase following the collision involves any rescue measures taken, such as the making of an emergency call.

Active safety and advanced driver assistance systems are relevant in the pre-collision phases, 1–3, but no longer once the first impact takes place. Depending on how the system takes effect, it may be able to prevent the critical situation from arising (e.g., a navigation system minimizes the extent to which the driver is distracted from driving, and adaptive cruise control ensures that a sufficient distance to the car in front is maintained). Alternatively, it may defuse a critical situation (like an ESC system) or reduce the force of the impact once the point of no return is passed (like a brake assist system). As a result of this diversity and the range of different ways in which advanced driver assistance systems take effect, special methods are required in order to ascertain the safety potential.

2.1 Methods of Assessing the Safety Potential of ADASs

The safety potential of advanced driver assistance systems (ADASs) can be ascertained in a variety of ways. For example, a retrospective comparison of two accident groups can be carried out: vehicles with ADASs and vehicles without them. Lie et al. used this approach to prove the effectiveness of electronic stability control (ESC) on the basis of Swedish accident data (Lie et al. 2005).

For the results shown below, an alternative, "what if" method, was used (Hummel et al. 2011). In this approach, the course of an accident as it happened in reality is examined and contrasted with what would have happened with a generic advanced driver assistance system. Generic in this context means a system with a number of selected features rather than a product that is actually available on the market. This makes it possible to determine the effect a particular type of advanced driver assistance system would have on the accident statistics if all cars were fitted with the system. In order to use this method, both the accident circumstances and the features (functionality) of the system to be examined or a generic system must be known. In the multistage procedure adopted in this method, the following distinction was drawn: whether the accident would have been preventable or whether its effects could only have been mitigated. An accident is considered to be theoretically preventable if it would not have happened with an ADAS. However, if the analysis shows that the accident would still have happened but that its consequences may have been less serious, the system is still considered to be capable of having a positive effect.

This can be illustrated with significantly greater precision by means of a simulation. In this case, too, the accidents are examined on a "what if" basis. Now that the accident situation can be portrayed with such detail and precision in the simulation environment, systems with significantly more complex functionality can be assessed with regard to their benefits. In addition, for warning and notification systems, human reactions have to be taken into account in the form of a driver model. The definition of this driver model is a great challenge, since it cannot always be assumed that the reactions of the driver will be suitable. In the study "Equal Effectiveness for Pedestrian Safety," for example, the benefits of a brake assist system were examined in terms of its effect on all pedestrian accidents in GIDAS. In all accident scenarios involving cars, it was examined what positive effects on the pedestrian an emergency braking system would have as a result of the lower speed of impact that would be expected. Following this case-based analysis of over 700 real pedestrian accident scenarios, the reductions in the numbers of seriously injured and killed pedestrians were compared with the known potential of other measures to protect pedestrians (Hannawald and Kauer 2003).

As an alternative, a field operational test (FOT) can also be considered in order to analyze the safety potential. Field operational tests are used primarily to evaluate new technologies such as ADASs (Benmimoun et al. 2011). To this end, the vehicle is equipped with extensive measurement equipment. The driver is then instructed, for example, to drive for a period with the ADAS switched on or off. This type of behavioral observation has become possible as a result of the rapid technical progress made in the collection, storage, and analysis of large quantities of data and the development of a measurement equipment that takes up less and less space. Everything required to explain and describe the driver's driving and the functionality of the ADAS is recorded: from the vehicle's movement (e.g., acceleration, speed, direction, vehicle status, etc.) to eye, head and hand movements, and pedal operation. This data provides information about the interactions between driver, vehicle, road, weather, and traffic not just in normal conditions but also in critical situations and even accident situations. The biggest challenge here is analyzing the very large quantities of data.

2.2 Cars

The most promising advanced driver assistance system for cars is the emergency braking system, followed by the lane departure warning system or lane-keeping assist system and the blind spot warning system (see Table 2). The emergency braking system becomes even more effective if it is also able to address accidents with pedestrians and cyclists. Up to 43.4 % of all car accidents in the database then become preventable (Hummel et al. 2011).

Table 2 Safety potential of ADASs for cars based on all accidents involving cars (Hummel et al. 2011)

ADAS	Theoretical safety potential
Emergency braking system (does not react to stationary vehicles)	17.8 % p
Lane departure warning system	4.4 % p
Blind spot warning system	1.7 % p

p preventable

2.3 Trucks

Using the method described above, it was found that an emergency braking system was the advanced driver assistance system with the greatest safety potential for trucks as well (Hummel et al. 2011). The potential doubled when the system was also able to detect stationary vehicles in front of the truck. The emergency braking system was followed by the blind spot warning system and the turning assistant with cyclist detection when all truck accidents were taken into account (see Table 3). If we look only at accidents between trucks and unprotected road users, the safety potential of a turning assistant with cyclist and pedestrian detection is very high. In order to analyze the cases, a system was assumed that monitors the areas in front of and to the right of the truck and warns the truck driver if there is a pedestrian or cyclist in the critical zone when the vehicle is pulling away or during turning. It was assumed that the driver would make the ideal response to the warning.

It was found that around 43 % of all truck accidents involving cyclists and pedestrians could be prevented if this turning assistant were used and that around 31 % of the cyclists and pedestrians killed in collisions with trucks would not be killed.

There is significant variation in the safety potential of the different advanced driver assistance systems depending on the configuration/type of the truck they are used with (see Table 4).

2.4 Buses

The emergency braking system is the system with the greatest safety potential for buses (see Table 5). It is followed by the blind spot warning system and the turning assistant with pedestrian and cyclist detection. There is also a clear increase in safety potential if the emergency braking system can detect stationary vehicles.

Here, too, the potential of the ADAS varies depending on the type of bus or the purpose for which it is used (see Table 6). Intercity buses benefit significantly more

Table 3 Safety potential of ADASs for trucks based on all accidents involving trucks (Hummel et al. 2011)

ADAS	Theoretical safety potential
Emergency braking system	6.1 % p
Emergency braking system (reacts to stationary vehicles)	12.0 % p
Turning assistant for pedestrians	0.9 % p
Turning assistant for cyclists	3.5 % p
Lane departure warning system	1.8 % p
Blind spot warning system	7.9 % a

p preventable, *a* addressable

Table 4 Safety potential of ADASs for trucks depending on truck type/configuration (Hummel et al. 2011)

ADAS	Box truck (no trailer)	Truck with trailer	Semitrailer truck
Emergency braking system (p)	2.2 %	6.1 %	5.1 %
Emergency braking system (reacts to stationary vehicles) (p)	7.9 %	10.7 %	9.5 %
Turning assistant for cyclists (p)	4.2 %	0.6 %	2.9 %
Turning assistant for pedestrians (p)	0.5 %	0.9 %	0.8 %
Blind spot warning system	6.8 %	5.2 %	6.4 %
Lane departure warning system (p)	1.6 %	1.8 %	1.3 %

p preventable, *a* addressable

Table 5 Safety potential of ADASs for buses based on all accidents involving buses (Hummel et al. 2011)

ADAS	Theoretical safety potential
Emergency braking system (a)	8.9 %
Emergency braking system (reacts to stationary vehicles) (a)	15.1 %
Turning assistant for cyclists and pedestrians (p)	2.3 %
Lane departure warning system (p)	0.5 %
Blind spot warning system (a)	3.8 %

p preventable, *a* addressable

Table 6 Safety potential of ADASs for buses depending on bus type/purpose (Hummel et al. 2011)

	Theoretical safety potential	
ADAS	City bus	Intercity bus
Emergency braking system (a)	11.9 %	4.5 %
Emergency braking system (reacts to stationary vehicles) (a)	16.6 %	17.3 %
Turning assistant (p)	3.4 %	–
Land departure warning system (p)	0.3 %	1.5 %
Blind spot warning system (a)	0.2 %	14.6 %

p preventable, *a* addressable

from a blind spot warning system, for example, while a turning assistant is more beneficial for a city bus.

2.5 Outlook

Across all the vehicle types analyzed, the emergency braking system emerges as the most promising ADAS. It will be found in different forms and with different functionality in all vehicle categories in the future – with justification, as the figures show. The addition of cyclist and pedestrian detection to the functionality of emergency braking systems will increase their safety potential. The turning assistant is particularly effective in trucks for the protection of cyclists and pedestrians. The increasing numbers of cyclists on the roads can only increase its importance. A next step in the development of these systems could be automatic evasion in emergency situations. This could increase the effectiveness of pure emergency braking systems in certain accident-critical situations, since, in terms of pure driving dynamics, evasion can take place at a later point than braking. This functionality has not yet been evaluated in terms of its influence on improving road safety, so that is essential. However, it would place even greater demands on the quality of the methods of analysis and the accident data.

The development of ADASs will benefit from the major trend toward highly automated driving, and the systems found in vehicles in the future will blur the boundaries between systems that make driving easier and pure safety systems as well as between different ADAS functions. Road safety will benefit when it is no longer necessary for drivers to develop an understanding of the various ADAS functions in order to be able to interpret warnings etc. correctly. Instead, it would make sense to have a fluid protection zone around the vehicle that supports the natural responses of the driver in critical situations. This goes hand in hand with the further development of the human-machine interface in order to adjust warnings or interventions so that they cannot be misinterpreted by the driver. The accident analyses show that today’s systems still have clear shortcomings in this respect.

References

Auerbach K (2014) Schwer- und schwerstverletzte Straßenverkehrsunfallopfer (Seriously and severely injured road accident victims). BASt-Newsletter Nr. 2, Bergisch Gladbach

Benmimoun M, Kessler C, Zlocki A, Etemad A (2011) euroFOT: field operational test and impact assessment of advanced driver assistance systems – first results. In: 20th Aachen Colloquium “Automobile and Engine Technology”, Aachen

Destatis (2012) Fachserie 8, Reihe 7. Federal Statistical Office, Wiesbaden

Destatis (2013) Fachserie 8, Reihe 7. Federal Statistical Office, Wiesbaden

Hannawald L (2014) Verkehrsunfallforschung an der TU Dresden GmbH, Analyse GIDAS Stand 31.12.2013 (Road accident research at TU Dresden GmbH, analysis of GIDAS as at 12/31/2013), unpublished, Dresden

Hannawald L, Kauer F (2003) Equal effectiveness study on pedestrian protection. TU Dresden, Dresden

Hummel T, Kühn M, Bende J, Lang A (2011) Advanced driver assistance systems – an investigation of their potential safety benefits based on an analysis of insurance claims in Germany. Research report FS 03. Gesamtverband der Deutschen Versicherungswirtschaft e.V, Berlin

Lie A, Tingvall C, Krafft M, Kullgren A (2005) The effectiveness of ESC (Electronic Stability Control) in reducing real life crashes and injuries. In: 19th international technical conference on the enhanced safety of vehicles conference (ESV), Washington, DC

Malczyk A (2010) Schwerstverletzungen bei Verkehrsunfällen (Severe injuries in road accidents), vol 722, Fortschritt-Berichte VDI-Reihe 12. VDI-Verlag, Düsseldorf

Meyer R, Ahrens G A, Aurich A P, Bartz C, Schiller C, Winkler C, Wittwer R (2012) Entwicklung der Verkehrssicherheit und ihrer Rahmenbedingungen bis 2015/2022 (The development of road safety and the conditions that make up its framework up to the year 2015/2022). Reports of the Bundesanstalt für Straßenwesen, M 224, Wirtschaftsverlag NW, Bergisch Gladbach

UDV (2011) Omnibus-Unfälle (Bus accidents). Study of the UDV (Unfallforschung der Versicherer), unpublished, Berlin

Behavioral Aspects of Driver Assistance Systems

5

Gert Weller and Bernhard Schlag

Contents

Abstract

Driving a car is a task that requires predominantly cognitive resources. Which resources are required exactly and to which amount they are required depends on the characteristics of the situation. The situation comprises the driver, the vehicle, and the environment. Advanced driver assistance systems (ADAS) change the situation and thus the task of driving a car in several aspects. These changes have an impact on the driver and mainly concern workload, situation awareness, and mental models. Changes therein can result in behavioral changes subsumed under the term behavioral adaptation. Behavioral adaptation can be negative or positive or even both depending on the perspective. However, when developing and designing ADAS, the nature of the aforementioned changes resulting in behavioral adaptation must be understood and must be taken into account. Behavioral changes must further be an integral part of evaluating the effect of such systems. Because of the constant technological progress, ADAS increasingly constitute a step toward full automation. Therefore, experiences in other domains in which automation has reached higher levels than in passenger

---

G. Weller (✉) • B. Schlag
Technische Universität Dresden, Dresden, Germany
e-mail: gert.weller@tu-dresden.de; schlag@verkehrspsychologie-dresden.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_5

cars must also be taken into account. This chapter provides the basis to understanding such ADAS-related effects.

In order to assess advanced driver assistance systems (ADAS) from a behavioral science perspective, the term advanced driver assistance system must be defined in a way that is relevant to the behavioral sciences. Modern driver assistance systems are systems which take over key components of human cognition. Engeln and Wittig (2005, quoted in Engeln and Vratil 2008) state that these key components are perception and the evaluation of what is perceived. Thus, purely carrying out an activity does not fall within the definition of driver assistance.

Taking over or automating these core components of human information processing by a technical system inevitably changes the driver's role. The subsequent paragraphs describe positive and negative aspects of this change. The following factors are seen as relevant and are defined in more detail below:

– Visual and cognitive workload
– Situation awareness
– Mental models

Changes in these factors are the basis for measurable changes in driver and in driving behavior. These changes are denoted as behavioral adaptation.

There is an interesting issue associated with automation which is a result of switching between different levels of automation. This is referred to as the "take-over-control problem."

This chapter provides an overview of the aforementioned points and describes them within the context of supporting drivers via advanced driver assistance systems (ADAS).

1 Visual and Cognitive Load

Load (strain) is the behavioral scientific construct associated with advanced driver assistance systems that possibly reveals itself earliest to a layperson. Put more simply, workload mirrors the degree of effort required to carry out a task. The level of difficulty of this task mirrors the stress of this task; the total stress is the sum of all stressors demanded by all tasks and environmental conditions. Although it might seem the case at first glance, stress and workload are not the same. This is because the same stressor can lead to different levels of load depending on the resources that the person can call upon. The amount of resources and access to them varies within an individual as well as between individuals. Thus, the same task done at different times of day can lead to different levels of load. This also applies when there is a change in the number of times a task is carried out.

A formal definition of mental stress and strain (workload) can be found in ISO Standards (ISO 10075 1991):

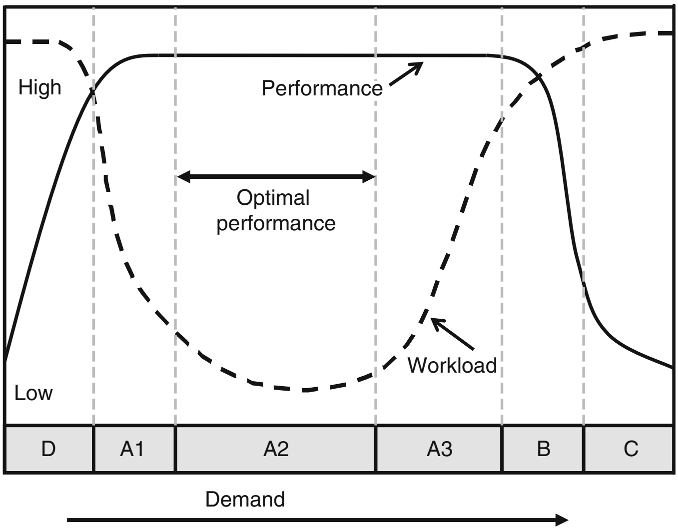


Fig. 1 Relationship between stress, workload, and performance (de Waard 1996)

- Mental stress: The total of all assessable influences impinging upon a human being from external sources and affecting it mentally
- Mental strain: The immediate effect of mental stress within the individual (not the long-term effect) depending on his/her individual habitual and actual preconditions, including individual coping styles

When talking about advanced driver assistance systems, it is important to note that the relationship between stress, load, and performance is not linear but rather u-shaped (stress – load) or inverted u-shaped (stress – performance) (see Fig. 1).

This means that either being under-challenged (underload) or over-challenged (overload) can lead to a drop in performance and that a person only achieves optimal performance with a medium level of stress. However, performance alone is not a valid indicator for unfavorable stress factors. Unfavorable stress factors exist when additional effort must be invested in order to sustain a high level of performance. This is illustrated in areas A1 and A3 in Fig. 1. According to de Waard (1996), "state-related effort" must be called upon in area A1 and "task-related effort" in A3. "State-related effort" expresses the fact that people must activate themselves in order to counteract the threat of being under-challenged. In contrast, "task-related effort" means that effort must be made in order to cope with the additional requirements of the task. In both cases, increased effort made over an extended period of time results in a reduction in performance as shown in areas B and D. In all cases, a limit in available resources is responsible for reduced performance.

As mentioned, the use of advanced driver assistance systems can change the demand placed upon the driving task. If, and to what degree, this happens is dependent upon the characteristics of the system. Due to the fact that, by definition, assistant systems are supposed to relieve driver workload, this is most likely to result in underload (Young and Stanton 2002). Underload is not a problem where there is complete automation because drivers can devote themselves to other tasks. However, in the case of partial automation, being under-challenged can be a safety problem. This applies particularly when the driver must take over control unexpectedly. Underload is a typical consequence of automation which often results in a high proportion of monitoring activities (Wickens et al. 2013).

Overload as a result of an advanced driver assistance system can be dangerous because of two reasons. Firstly, the model that a user makes from the systems' characteristics can be flawed or wrong (see Sect. 3). Secondly, new kinds of requirements result from assistant systems. The most noteworthy of these is the design of the human-machine interface (HMI). Apart from activating the system and its settings, the interface must at least convey the current status of the system. During normal driving this occurs visually because audio signals are more associated with warnings (ISO 15005 2002; ISO 15006 2011; ISO 15623 2013). Displaying additional visual information in vehicles is potentially problematic because the visual channel is the one that is the most in demand (Sivak 1996). In addition to increased visual load due to the optical display of information, there are also the known problems of distraction (passive, bottom-up driven) and aversion (active, top-down driven). Thus, it is important to take into consideration the relevant standard (ISO 15005 2002) when designing how this information is displayed.

When assessing the effect on workload from using an advanced driver assistance system, load should be broken down into different types (Taylor et al. 2013). This subdivision goes back to the multiple resources model of Wickens (2002, quoted in Wickens and McCarley 2008). According to this model, there is not a single resource but rather many different resources that are, to a great extent, independent of one another. Wickens differentiated between resources depending on how information is coded (codes), how they are displayed/emitted (modalities), how they are processed (stages), and what answers are given (responses) (Wickens and McCarley 2008).

When designing advanced driver assistance systems, the goal should be to optimize workload and not necessarily reduce it. Approaches to optimizing workload can be found in Piechulla et al. (2003) and in Hajek et al. (2013) or in the field of aviation in Liu et al. (2012).

Regardless of the type of workload, it can be measured using three different metrics (de Waard 1996):

- Performance metrics
- Physiological metrics
- Subjective metrics

An overview of each approach can be found in Gawron (2008). There is also an overview of the different methods in ISO 17287 (2003). Metrics for measuring visual attention are given their own category there.

2 Situation Awareness

Situation awareness in terms of "needing to know what's going on" is undoubtedly a must for safe driving. The term situation awareness, at first, seems to be self-explanatory; however, established definitions reveal a far more complex picture that goes beyond the everyday meaning. For example, Mica Endsley defines situation awareness as follows (1988, quoted from Endsley 1995):

> Situation Awareness is the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future.

According to this definition, there are three hierarchical levels of situation awareness: perceiving the situation (Level 1), comprehending the situation (Level 2), and projecting the future situation (Level 3). These three levels in the context of action regulation as developed in the situation awareness model of Endsley (1995) are shown in Fig. 2.

The hierarchical structure of situation awareness means that a higher level of situation awareness cannot be achieved without there being situation awareness at the lowest level. This hierarchical definition of situation awareness assumes that all levels of situation awareness must be present in humans.

Advanced driver assistance systems can help drivers at all three levels of situation awareness; however, they can also result in reduced situation awareness. This is the case when erroneous, imprecise, or too little/too much information is provided or when the information is provided at the wrong time or the wrong place.

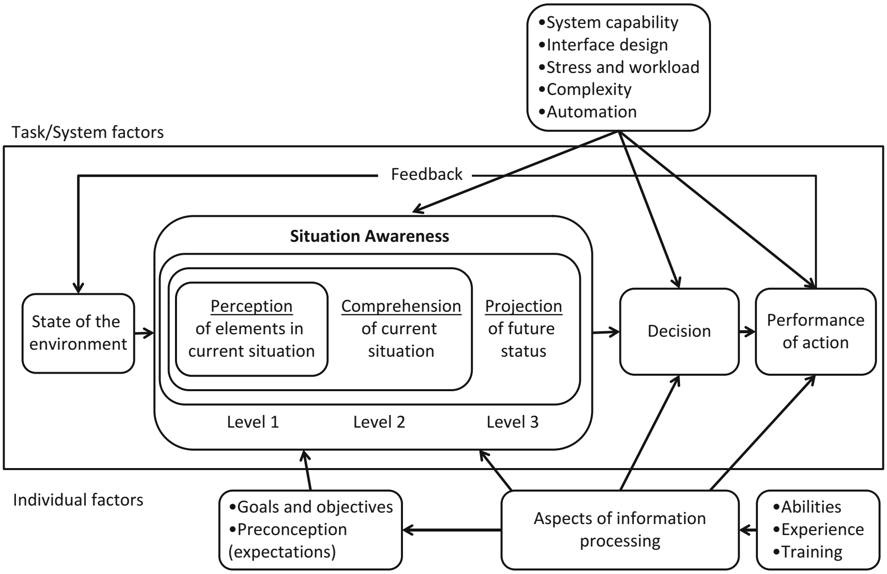


Fig. 2 Framework model of situation awareness (Endsley 1995; Kluwe 2006)

In these situations, advanced driver assistance systems have a direct effect on attention and consequently an indirect effect on situation awareness. According to the model of Wickens (2007), attention is both a filter and a resource. The filter is selective attention. This could be thought of as the beam of light produced by a car's headlights which cannot illuminate all necessary objects at once. If attention is diverted to irrelevant information, then other important information fails to be perceived (for other concepts of attention, see Strayer and Drews 2007 and Müller and Krummenacher 2008). Thus, selective awareness is particularly important for Level 1 of situation awareness. On the other hand, the part of attention which is understood to be a limited resource influences all three levels of situation awareness equally (for an in-depth look at attention as a resource, see the contribution by Abendroth and Bruder in ▶ Chap. 1, "Capabilities of Humans for Vehicle Guidance"). If resources are used up or are committed to other tasks, then there are correspondingly fewer resources available to develop good situation awareness. Although attention as a resource is also linked with load, situation awareness and workload are independent of one another (Endsley and Kiris 1995; Wickens 2008).

In addition to specific effects resulting from advanced driver assistance systems, general effects of automation on situation awareness have also been discussed. Endsley and Kiris (1995) state that the following can lead to a lack of situation awareness:

- False understanding of role allocation (moving from active operator to observer).
- Change from active information processor to passive information receiver.
- Missing feedback or change in quality of feedback related to the system.

When automating systems in general and advanced driver assistance systems in particular, special focus should be given to the implications for situation awareness. It is important that the drivers receive system feedback and, as long as there is no full automation, that they remain in the control loop.

Measuring situation awareness can be done in one of the following three ways (Durso et al. 2007):

- Subjective measurement
- Situation-specific questions
- Performance measurement

Generally speaking, measurement is done via situation-specific interviews with test persons. If this is done as part of a simulation, the simulation is stopped ("Freezing Technique") and the questions are asked. One particular method of doing this is SAGAT ("Situation Awareness Global Assessment Technique") which was developed by Endsley (2000). A prerequisite of using this method is that questions must be defined in advance for the specific situation.

One method that falls within the first category, and as such is largely independent of the specific situation, is SART ("Situation Awareness Rating Technique")

(Taylor 1990, quoted in Salmon et al. 2009). This scale consists of ten dimensions that fall under three primary categories; each dimension is rated on a 7-level scale. It is important to note that both subjective measurement and situation-specific questions require a certain level of consciousness. However, just because a person is not conscious of all elements of a situation does not necessarily mean that there is reduced situation awareness.

Thus, Gugerty (1997) differentiates between direct methods, such as those mentioned above, and indirect methods. The indirect methods include performance measures from which the level of situation awareness must be extrapolated. Performance measures include reaction times as well as other measures of driver and driving behavior as long as they are relevant to situation awareness. One example for using reaction time is SPAM (Situation Presence Awareness Method) by Durso and Dattel (2004) which asks test subjects questions and the time required to (correctly) answer is used as the measure of situation awareness. Durso et al. (2007) advise using glance measurement with caution in conjunction with situation awareness.

Overall, a driver's level of situation awareness can have a significant impact on the performance of the entire system. It is an important prerequisite for interpreting situations and for choosing the right way to act. Advanced driver assistance systems can have either a positive or a negative influence on situation awareness.

3 Mental Models

In the previous chapter on situation awareness, it was explained that advanced driver assistance systems can raise awareness via prominent warning signals. In this situation, attention is drawn through stimuli, that is, bottom-up. However, attention can also be guided top-down, that is, through expectations. These expectations are formed from the beliefs we have of how systems work. The totality of these beliefs is known as the mental model.

Brewer (2002) defines mental models as follows:

> A mental model is a form of mental representation for mechanical-causal domains that affords explanations for these domains. (. . .) The information in the mental model has an analogical relation with the external world: the structure of the mental representation corresponds to the structure of the world. This analogical relation allows the mental model to make successful predictions about events in the world. (p. 5/6)

One definition that explicitly includes the effects of mental models on behavior comes from Wilson and Rutherford (1989):

> [. . .] a mental model is a representation formed by a user of a system and/or task, based on previous experience as well as current observation, which provides most (if not all) of their subsequent system understanding and consequently dictates the level of task performance. (p. 619)

Mental models are important because they control expectations, and these expectations, in turn, influence driver behavior (Weller 2010). Mental models, together with other forms of internal representation such as schemata and scripts, have various advantages which can have a positive influence on the effectiveness and efficiency of dealing with systems (Weller et al. 2006):

- Generally, they are simpler than the reality.
- They are more likely to be used automatically than consciously, and as a result they happen faster and use fewer mental resources.
- They automatically draw attention to relevant stimuli and thus help to use attention resources efficiently.

However, it is because of these reasons that mental models can also be the cause of mistakes (Hacker 2005): if the mental models are incomplete or incorrect, then the way a person acts as a result can also be erroneous. Due to the fact that mental models are activated via signals from the environment, a misleading signal or a false interpretation of a signal can result in the wrong mental model being activated.

Especially in the area of advanced driver assistance systems, it has been shown that a user's mental model of the way a system works can be wrong (Jenness et al. 2008, quoted in Beggiato and Krems 2013). However, more experience with the system can result in them being correctly calibrated (Beggiato and Krems 2013).

Collecting data on mental models is difficult because of a number of reasons. Mental models are as varied as the systems being examined. Due to the fact they are not one-dimensional and they reflect different functional connections, largely standardized approaches must be adapted. Nevertheless, there are several approaches which can be used to examine mental models. These are described in Cherri et al. (2004). Generally, these are done by conducting interviews or through the use of questionnaires. Additional methods are described by Mohammed and Lori Hamilton (2010).

A user's mental model of an assistance system depends on the degree of trust that the user has in the system. Trust should be calibrated so that it corresponds to the actual system characteristics (Lee and See 2004). If this is not the case and trust is too high or too low, it can have negative results on behavior which Wickens et al. (2004) describe as mistrust (not enough trust) and complacency (too much trust) (see Fig. 3).

Mental models also determine the level of situation awareness due to their role in directing attention. Mental models act as a heuristic in the search for information (Stanton and Young 2005). They can only be adjusted through experience and feedback. Therefore, it is important to inform the driver about the state of the system and limitations of the system.

4 Behavioral Adaptation

Adapting behavior to changing conditions is a natural reaction of people and a prerequisite for their development. This characteristic can, however, become a problem if it is not taken into consideration when planning changes or when

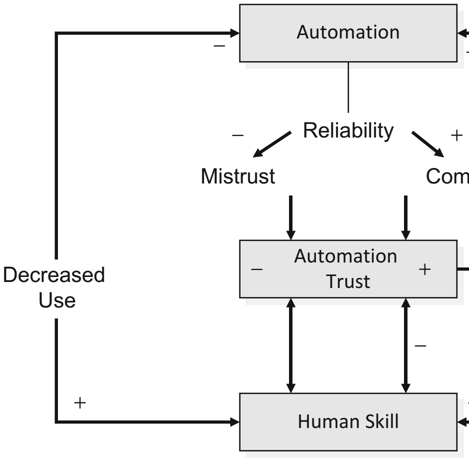


Fig. 3 Relationship between reliability, complacency, trust, and human skill (according to Wickens et al. 2004)

something unexpected occurs. Behavioral adaptation should be defined prior to going into detail about individual aspects of behavioral adaptations. The generally accepted definition comes from the OECD (1990):

> Behavioral adaptations are those behaviors which may occur following the introduction of changes to the road-vehicle-user system and which were not intended by the initiators of the change; Behavioral adaptations occur as road users respond to changes in the road transport system such that their personal needs are achieved as a result, they create a continuum of effects ranging from a positive increase in safety to a decrease in safety. (p. 23)

There are numerous examples of behavioral adaptations in the field of traffic and transportation. Previous studies have shown unexpected behavioral adaptations to ABS (Sagberg et al. 1997) or ACC (Hoedemaeker and Brookhuis 1998; Weinberger 2001), as well as effects for road width (Lewis-Evans and Charlton 2006). A summary of various studies can be found in the OECD report (1990).

Although there is no question that behavioral adaptations can occur, there are two outstanding questions:

- Which factors influence the onset and the extent of behavioral adaptations?
- How can the effects of behavioral adaptations on safety be assessed?

In order to assess the effect of behavioral adaptations on safety, it is necessary to compare the intentional change in behavior (e.g., relieve driver burden) with unintentional changes in behavior (see Elvik and Vaa 2004). Only when the effects of intentional behavioral adaptation are greater (in absolute terms) than the effects of unintentional behavioral changes can it be said that they are successful. Rothengatter (2002) assumes that these unintentional effects are not sufficient to undo the positive effects. However, it should be noted that the impact of unintentional behavioral adaptation tends to be underestimated (Dulisse 1997). In methodological terms, it is not only difficult to tease apart the individual elements but also very difficult to separate the effect of individual interventions from other influences. Approaches for this can be found in Noland (2003).

Precisely because it is difficult to separate changes in behavior into intentional versus unintentional components, it is vital to understand which factors result in behavioral adaptations and how these can be influenced. Advanced driver assistance systems can influence driver workload and situation awareness; these changes may have their origins in false mental models of how things work or the limitations of advanced driver assistance systems. If these changes reach a critical threshold, the result is a change in – or adaptation of – behavior. This type of adaptation has been described in previous chapters.

In addition to behavioral adaptation based on the pattern mentioned above, behavioral adaptation can also result from a change in motivation variables. The most relevant of these variables are perceived risk and the effort to keep this workload at an optimal level. The following three theories are considered to be particularly influential:

- Risk Homeostasis Theory (RHT) by Wilde (1988, 2001)
- Zero-Risk Theory by Näätänen and Summala (1976) and Summala (1988)
- Task-Difficulty Homeostasis by Fuller (2005, 2011)

The following briefly describes the basic assumptions of each model. A more detailed discussion of the models can be found in Weller (2010).

The most influential and most controversial theory is Wilde's Risk Homeostasis Theory. This theory posits that in every society there is a desired level of risk ("target risk"). The goal of social behavior is not to reduce risk but rather to achieve the desired level of risk. Although the theory was originally developed at an aggregated level for an entire society, it was soon used at an individual level. Thus, a driver compares their current perceived risk with their individual target risk. By adapting their own behavior (e.g., increasing or decreasing speed), the driver attempts to minimize discrepancies between the two values. Since target risk is considered stable, a reduction in subjective risk (e.g., through the use of driver assistance systems) means that the driver must compensate for this apparent increase in safety through risky behavior in order to get back closer to their target risk.

Zero-Risk Theory by Näätänen & Summala is another influential theory that assumes subjective risk as one of the relevant behavioral determinants. In contrast to Wilde's theory, the authors postulate that drivers do not seek a certain level of risk but rather behavior is controlled and regulated via safety margins (see Lu et al. 2012) with subjective risk normally at zero. As with Wilde, subjective risk is the product of subjective probability and subjective utility ("subjective expected utility," SEU) of an unpleasant incident. This theory is different to Wilde's in that the driver strives to keep risk low. According to the theory, accidents happen when there are errors in perception and interpretation when determining risk.

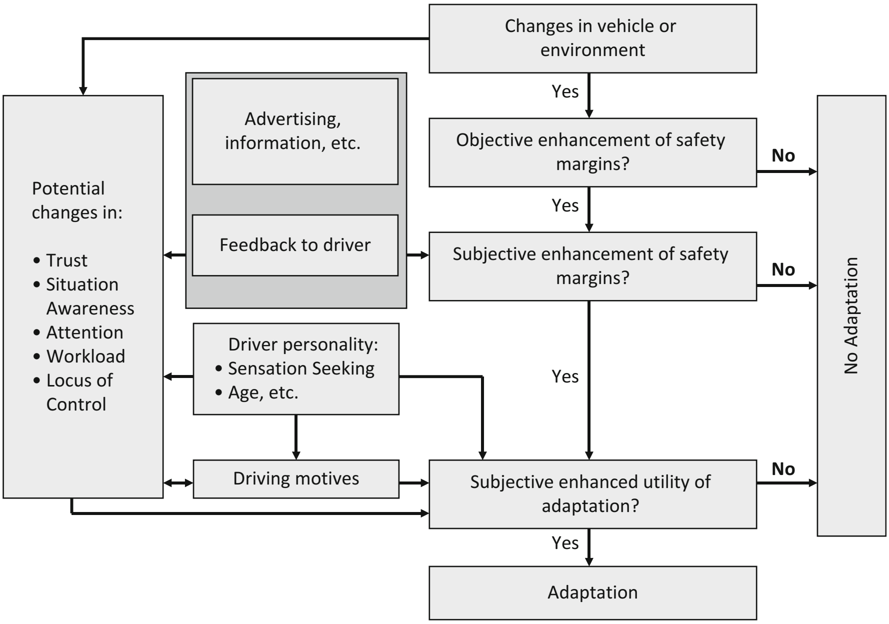


Fig. 4 Process model of behavioral adaptation (Weller and Schlag 2004)

Unlike the aforementioned theories, Fuller does not see risk initially as the determinant of behavior but rather task difficulty. Fuller, at first, assumes that the driver mainly adjusts driving speed, and thus the degree of task difficulty, in order to try to achieve a certain fixed level of task difficulty. This endeavor could be incorporated into area A2 of de Waard's model (1996) (see Fig. 1). Fuller later changed the concept from homeostasis to allostasis. This change takes into account that the desired level of task difficulty can also change. Fuller (2011) gives the example of driving with flashing blue lights where drivers are prepared to seek a level of task difficulty that deviates from that which is normally accepted. In de Waard's model (see Fig. 1), this would correspond approximately to the A3 area, that is, consciously trying for a short period of time to achieve a higher level of task difficulty than is optimal. This does not have negative consequences in the short term, but over the long term, it can have a negative impact. Although Fuller (2011), like the other two authors, comes back to risk as a behavior determinant ("risk allostasis theory," RAT), his initial concept of behavioral adaptation through task difficulty represents an important contribution to explaining behavioral adaptation following the introduction of advanced driver assistance systems.

The approaches to explaining behavioral adaptation were consolidated into a process model by Weller and Schlag (2004) (see Fig. 4). This model explains behavioral adaptation in motivation terms through the subjective benefit that results from a change in behavior. However, the starting point is a change in the vehicle or the driving environment. Only when this change allows an objective increase in the scope for action can an adaptation to behavior occur. The probability of an accident occurring is used as the basis for assessing whether there is an objective increase in the scope for action. This is based on the assumption that drivers strive primarily to avoid a collision rather than to reduce the severity of injuries. This assumption explains why passive safety systems generally lead to less behavioral adaptation.

This objective increase in scope for action must be perceived. It is possible that systems which only very occasionally intervene in an actual emergency do cause less behavioral adaptation because the driver is simply not aware of them. This perception is influenced through external factors such as advertising, system attributes such as the type and level of feedback, and from those driver characteristics which determine whether the driver perceives the scope to act. Finally, the driver must anticipate the benefit of adapting behavior. This is the case, for example, when the driver can drive faster under the same conditions. However, the driver may not experience a benefit if they can drive faster but this is associated with a decrease in driving comfort. The criteria that the driver applies depend on personal characteristics (for more on the impact of sensation-seeking on behavioral adaptation, see Jonah (1997)), motives for driving (e.g., time pressure, see Adams-Guppy and Guppy (1995)), and the effects of driver assistance systems on psychological characteristics as mentioned above (see also Stanton and Young 1998). Subjective benefit is defined as the potential to achieve goals which result from the aforementioned motivation theories of explaining behavioral adaptation.

5 The Take-Over-Control Problem

The take-over-control problem refers to when a human takes over control or must take over control of an automated system. Taking over control is problematic due to a number of reasons. These are closely linked to the topics already mentioned.

- Advanced driver assistance systems and automation are supposed to reduce driver workload. This reduction in workload can lead to the driver being under-challenged and deactivated (cf. Fig. 1). If drivers unexpectedly have to take over control, it can take some time before they are sufficiently reactivated.
- The more the driving task, or part of the driving task, is automated, the less the drivers must inform themselves about the driving situation since knowledge of the current driving is depicted or saved externally. An example of this is an overtaking assistant which warns the driver of vehicles in their dead spot when changing lanes. If drivers rely on the assistant, they also reduce the amount they actively seek information. In the event of the system failing or in situations where the assistant does not operate, there can be problems due to insufficient situation awareness.
- Over the long term, constant use of automation can result in a loss of competencies – that is to say, deskilling (see Fig. 3). If this is the case and the necessary driving skills are no longer present, taking over control can lead to problems.

The points mentioned above are closely linked with what Bainbridge (1983) termed the "Ironies of Automation." These ironies can be summarized as follows:

- The aim of automation is to replace unreliable humans, but it is these same imperfect humans that are responsible for the development, design, and implementation of automation.
- Although humans are seen as unreliable, they are expected to monitor the automation.
- Precisely when the automation fails, that is, in highly complicated situations, is when humans are supposed to take over control.

Finding solutions for these ironies can only be done through human-focused automation as described, for example, by Billings (1997). Bainbridge (1983) herself suggests measures to counter the negative effects of automation, despite pointing out that there are not any simple solutions.

It is vital to inform the user of the system's status and even more so to communicate malfunctions ("automatic systems should fail obviously") (Bainbridge 1983, p. 777). Furthermore, automation must be consistent and predictable. In addition to information, active participation of humans can ensure that the driver remains in the loop which, according to Endsley and Kiris (1995), is a key requirement for automation that minimizes errors. The use of adaptive and adaptable automation is required on occasion (Kaber and Endsley 2004).

As can be seen from above, whether the take-over-control problem occurs or not is highly dependent on the amount of automation and its design.

Especially with monitoring activities, underload and low situation awareness lead to problems when a person needs to take over control. This is particularly problematic when automation is designed with different levels. In this case, the difficulty of simply taking over results in the problem of mode awareness (Sarter and Woods 1995; Sarter 2008). A lack of mode awareness or mode errors occur when a level of automation is assumed that is not there. This is critical when the driver turns to another task during a highly automated phase and then in an emergency has to take over control (Merat et al. 2012).

In order to avoid problems from taking over control and to keep the driver in the loop, Parasuraman et al. (2000) and Parasuraman (2000) suggest differentiated automation. According to this concept, the degree of automation should vary depending on the effects that are expected on load, situation awareness, complacency, and de-skilling. These effects should be taken into consideration for each task that is to be automated and separated by the individual action steps ("information acquisition," "information analysis," "decision and action selection," and "action implementation"). To assist in the decision-making process, MABA-MABA lists ("Men are better at – machines are better at") or Fitts lists (Fitts 1951, quoted in Lee 2006) can be used. This helps to avoid undifferentiated automation and to ensure that only those parts of a task are not automated where there is no technical possibility of doing so (see Ironies of Automation).

Especially with the development toward highly automated driving where the drivers, to a certain extent, can devote themselves to other tasks and the transition to manual mode happens with a time reserve, designing the prompt to take over control will play an important role in acceptance and, thus, the success of the technology. It is vital to take into consideration the facts mentioned above and to involve the user early on in the development process.

References

Adams-Guppy JR, Guppy A (1995) Speeding in relation to perception of risk, utility and driving style by British company car drivers. Ergonomics 38(12):2525–2535

Bainbridge L (1983) Ironies of automation. Automatica 19(6):775–779

Beggiato M, Krems JF (2013) The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. Transport Res Part F 18:47–57

Billings CE (1997) Aviation automation: the search for a human-centered approach. Erlbaum, Mahwah

Brewer FW (2002) Mental models. In: Nadel L (ed) Encyclopedia of cognitive science, vol 3. Wiley, London, pp 1–6

Cherri C, Nodari E, Toffetti A (2004) AIDE deliverable 2.1.1: review of existing tools and methods (EU contract No.: IST-1-507674-IP). Available: http://www.aide-eu.org/pdf/sp2_deliv/aide_d2-1-1.pdf. Mar 2007

de Waard D (1996) The measurement of drivers' mental workload. The traffic research centre VSC, University of Groningen, The Netherlands. Available: http://home.zonnet.nl/waard2/dewaard1996.pdf. Apr 2012

Dulisse B (1997) Methodological issues in testing the hypothesis of risk compensation. Accid Anal Prev 29(3):285–292

Durso FT, Dattel AR (2004) SPAM: the real-time assessment of SA. In: Banbury S, Tremblay S (eds) A cognitive approach to situation awareness: theory and application. Ashgate, Aldershot, pp 137–154

Durso FT, Rawson KA, Girotto S (2007) Comprehension and situation awareness. In: Durso FT, Nickerson RS, Dumais ST, Lewandowsky S, Perfect TJ (eds) Handbook of applied cognition. Wiley, Chichester, pp 164–193

Elvik R, Vaa T (2004) The handbook of road safety measures. Elsevier, Amsterdam

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Hum Factors 37(1):32–64

Endsley MR (2000) Direct measurement of situation awareness: validity and use of SAGAT. In: Endsley MR, Garland DJ (eds) Situation awareness analysis and measurement. Lawrence Erlbaum, Mahwah, pp 147–173

Endsley MR, Kiris EO (1995) The out-of-the-loop performance problem and level of control in automation. Hum Factors 37(2):381–394

Engeln A, Vratil B (2008) Fahrkomfort und Fahrgenuss durch den Einsatz von Fahrerassistenzsystemen (Driving comfort and driving pleasure by using driver assistance systems). In: Schade J, Engeln A (eds) Fortschritte der Verkehrspsychologie (Advances in traffic psychology). VS Research, Wiesbaden, pp 275–288

Fuller R (2005) Towards a general theory of driver behavior. Accid Anal Prevent 37(3):461–472

Fuller R (2011) Driver control theory. From task difficulty homeostasis to risk allostasis. In: Porter BE (ed) Handbook of traffic psychology. Academic, London, pp 13–26

Gawron VJ (2008) Human performance, workload, and situational awareness measures handbook, 2nd edn. CRC Press, Boca Raton

Gugerty LJ (1997) Situation awareness during driving: explicit and implicit knowledge in dynamic spatial memory. J Exp Psychol Appl 3(1):42–66

Hacker W (2005) Allgemeine Arbeitspsychologie. Psychische Regulation von Wissens-, Denk- und körperlicher Arbeit (General work psychology. Psychological regulation of knowledge-, thinking- and physical work), 2nd edn. Hans Huber, Bern

Hajek W, Gaponova I, Fleischer KH, Krems J (2013) Workload-adaptive cruise control – a new generation of advanced driver assistance systems. Trans Res Part F 20:108–120

Hoedemaeker M, Brookhuis KA (1998) Behavioral adaptation to driving with an adaptive cruise control (ACC). Transp Res Part F 1(2):95–106

ISO 10075 (1991) Ergonomic principles related to mental work-load – part 1: general terms and definitions. Geneva: ISO

ISO 15005 (2002) Road vehicles – ergonomic aspects of transport information and control systems – dialogue management principles and compliance procedures. Geneva: ISO

ISO 15006 (2011) Road vehicles – ergonomic aspects of transport information and control systems – specifications for in-vehicle auditory presentation. Geneva: ISO

ISO 15623 (2013) Intelligent transport systems – forward vehicle collision warning systems – performance requirements and test procedures. Geneva: ISO

ISO 17287 (2003) Road vehicles – ergonomic aspects of transport information and control systems – procedure for assessing suitability for use while driving. Geneva: ISO

Jonah BA (1997) Sensation seeking and risky driving: a review and synthesis of the literature. Accid Anal Prev 29(5):651–665

Kaber DB, Endsley MR (2004) The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theor Issues Ergon Sci 5(2):113–153

Kluwe RH (2006) Informationsaufnahme und Informationsverarbeitung (Information uptake and information processing). In Zimolong B, Konradt U (eds) Ingenieurpsychologie (Engineering psychology). Enzyklopädie der Psychologie, Themenbereich D, Serie 3, Bd 2 (Encyklopedia of psychology, Subject area D, series 3, vol 2). Hogrefe, Göttingen, pp 35–70

Lee JD (2006) Human factors and ergonomics in automation design. In: Salvendy G (ed) Handbook of human factors and ergonomics, 3rd edn. Wiley, Hoboken, pp 1570–1596

Lee JD, See KA (2004) Trust in automation: designing for appropriate reliance. Hum Factors 46(1):50–80

Lewis-Evans B, Charlton SG (2006) Explicit and implicit processes in behavioural adaptation to road width. Accid Anal Prev 38(3):610–617

Liu D, Guarino SL, Roth E, Harper K, Vincenzi D (2012) Effect of novel adaptive displays on pilot performance and workload. Int J Aviat Psychol 22(3):242–265

Lu G, Cheng B, Lin Q, Wang Y (2012) Quantitative indicator of homeostatic risk perception in car following. Safety Sci 50(9):1898–1905

Merat N, Jamson AH, Lai FCH, Carsten O (2012) Highly automated driving, secondary task performance, and driver state. Hum Factors 54(5):762–771

Mohammed SF, Lori Hamilton K (2010) Metaphor no more: a 15-year review of the team mental model construct. J Manage 36(4):876–910

Müller HJ, Krummenacher J (2008) Aufmerksamkeit (Attention). In: Müsseler J (ed) Allgemeine Psychologie (General psychology), 2nd edn. Spektrum, Heidelberg, pp 103–154

Näätänen R, Summala H (1976) Road-user behavior and traffic accidents. North-Holland, Amsterdam

Noland RB (2003) Traffic fatalities and injuries: the effect of changes in infrastructure and other trends. Accid Anal Prev 35(4):599–611

OECD (1990) Behavioral adaptations to changes in the road transport system. OECD, Paris

Parasuraman R (2000) Designing automation of human use: empirical studies and quantitative models. Ergonomics 43(7):931–951

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybernet Part A Syst Human 30(3):286–297

Piechulla W, Mayser C, Gehrke H, König W (2003) Reducing drivers' mental workload by means of an adaptive man–machine interface. Trans Res Part F 6(4):233–248

Rothengatter T (2002) Drivers' illusions – no more risk. Trans Res Part F 5(4):249–258

Sagberg F, Fosser S, Saetermo IAF (1997) An investigation of behavioral adaption to airbags and antilock brakes among taxi drivers. Accid Anal Prev 29(3):293–302

Salmon PM, Stanton NA, Walker GH, Jenkins D, Ladva D, Rafferty L et al (2009) Measuring situation awareness in complex systems: comparison of measures study. Int J Ind Ergonom 39(3):490–500

Sarter NB (2008) Investigating mode errors on automated flight decks: illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. Hum Factors 50(3):506–510

Sarter NB, Woods DD (1995) How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum Factors 37:5–19

Sivak M (1996) The information that drivers use: is it indeed 90% visual? Perception 25(9):1081–1089

Stanton NA, Young MS (1998) Vehicle automation and driving performance. Ergonomics 41(7):1014–1028

Stanton NA, Young MS (2005) Driver behaviour with adaptive cruise control. Ergonomics 48(10):1294–1313

Strayer DL, Drews FA (2007) Attention. In: Durso FT, Nickerson RS, Dumais ST, Lewandowsky S, Perfect TJ (eds) Handbook of applied cognition. Wiley, Chichester, pp 29–54

Summala H (1988) Risk control is not risk adjustment: the zero-risk theory of driver behaviour and its implications. Ergonomics 31(4):491–506

Taylor GS, Reinerman-Jones LE, Szalma JL, Mouloua M, Hancock PA (2013) What to automate: addressing the multidimensionality of cognitive resources through system design. J Cogn Eng Decision Making 7(4):311–329

Weinberger M (2001) Der Einfluss von Adaptive Cruise Control Systemen auf das Fahrverhalten (The influence of Adaptive Cruise Control on driving behavior). Dissertation TU München, Berichte aus der Ergonomie (Ergonomic reports). Shaker-Verlag, Aachen

Weller G (2010) The psychology of driving on rural roads. development and testing of a model. VS-Verlag, Wiesbaden

Weller G, Schlag B (2004) Verhaltensadaptation nach Einführung von Fahrerassistenzsystemen (Behavioral adaptation following the introduction of driver assistance systems). In: Schlag B (ed) Verkehrspsychologie. Mobilität – Verkehrssicherheit – Fahrerassistenz (Traffic psychology. Mobility – traffic safety – driver assistance). Pabst, Lengerich, pp 351–370

Weller G, Schlag B, Gatti G, Jorna R, van de Leur M (2006) RiPCORD – iSEREST. Internal report D8.1. Human factors in road design. State of the art and empirical evidence (EC: contract no.: 50 61 84, 6th framework programme). Available: http://ripcord.bast.de/pdf/RI-TUD-WP8-R1-Human_Factors.pdf. July 2008

Wickens CD (2007) Attention to attention and its applications: a concluding view. In: Kramer AF, Wiegmann D, Kirlik A (eds) Attention. From theory to practice. Oxford University Press, New York, pp 239–249

Wickens CD (2008) Situation awareness: review of mica endsley's 1995 articles on situation awareness theory and measurement. Hum Factors 50(3):397–403

Wickens CD, McCarley JS (2008) Applied attention theory. CRC Press, Boca Raton

Wickens CD, Lee JD, Liu Y, Becker SEG (2004) An introduction to human factors engineering, 2nd edn. Pearson Education, Upper Saddle River

Wickens CD, Hollands JG, Banburry S, Parasuraman R (2013) Engineering psychology and human performance, 4th edn. Pearson, Boston

Wilde GJS (1988) Risk homeostasis theory and traffic accidents: propositions, deductions and discussions of dissension in recent reactions. Ergonomics 31(4):441–468

Wilde GJS (2001) Target risk 2. A new psychology of safety and health: what works? what doesn't? and why. PDE, Toronto

Wilson JR, Rutherford A (1989) Mental models: theory and application in human factors. Hum Factors 31(6):617–634

Young MS, Stanton NA (2002) Attention and automation: new perspectives on mental underload and performance. Theor Issue Ergonom Sci 3(2):178–194

Functional Safety of Driver Assistance Systems and ISO 26262

6

Ulf Wilhelm, Susanne Ebel, and Alexander Weitzel

Contents

U. Wilhelm (✉)
Robert Bosch GmbH, Ditzingen, Germany
e-mail: ulf.wilhelm@de.bosch.com

S. Ebel
Robert Bosch GmbH, Chassis Systems Control, Driver Assistance Systems, Leonberg, Germany
e-mail: susanne.ebel@de.bosch.com

A. Weitzel
Formerly Technische Universität Darmstadt, Darmstadt, Germany

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_6

Abstract

This chapter gives a brief overview of the automotive standard ISO 26262 containing requirements for functional safety in order to avoid or control systematic failures and random HW failures. In this context the hazard analysis and risk assessment for a driver assistance function is shown, and further steps to prevent relevant hazards outgoing from the function are described. Since very important for the driver assistance functions, hazards resulting from functional insufficiencies, which are out of the scope of ISO 26262, are discussed in the last part of this chapter.

1 Objectives of Functional Safety

1.1 Overview

To release a technical product for sale and use, proof that it is safe enough always has to be provided first. In this general safety consideration, the section of correct and safe product function is known as functional safety (Börcsök 2011).

The reference for evaluating whether a product is safe or not is the tolerable risk limit. If the risk associated with the product is below the risk limit, then it can be considered as sufficiently safe. The risk, in turn, is defined in Engineering as the product of damage severity and probability of occurrence (ISO 31000 2009). If persons involved in the process are able to employ defined actions to avoid damages when a fault occurs, then the controllability can be considered as an additional factor.

The risk limit is defined by the current state of the art. In the event of damage, the manufacturer is obligated to prove that at the time the product was placed on the market, it complied with the state of scientific and technical knowledge regarding safety aspects (ProdHaftG 2002). The requirements resulting from this state of technology are often documented in standards. They comprise product characteristics, development methods, and documentation rules that has to be fulfilled by the product and the manufacturer.

Requirements for functional safety of electrical, electronic, and programmable electronic systems in general are covered by the technical standard IEC/EN 61508 (2010). ISO 26262 (2011), which is the standard relevant for the automotive industry, has been derived from the IEC/EN 61508 standard. It contains definitions, guidelines, and methods for development and assessment of functional safety of electrical and electronic (E/E) components.

1.2 Objectives and Structure of ISO 26262

ISO 26262 defines requirements for the development process of safety-critical components and systems of road vehicles. Currently, this only includes passenger cars up to 3.5 t. However, adapted variants for commercial vehicles (Dardar et al. 2012; Teuchert 2012) and motorized two-wheelers (Bachmann and Zauchner 2013; Werkmeister and Englisch 2012) are under development.

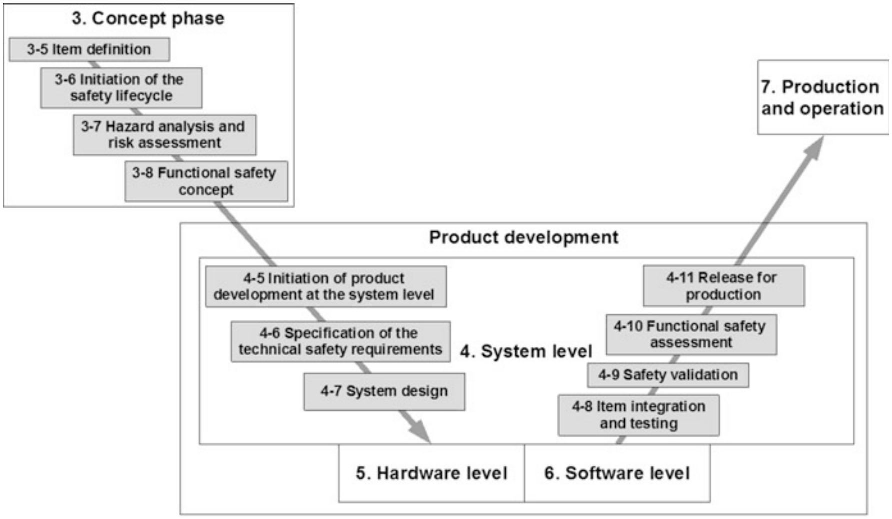


Fig. 1 Development process according to ISO 26262 with focus on system development

The procedure described in ISO 26262 is oriented toward the general V-Model of product development (V-Model 2013). At the beginning of the product life cycle, in the design phase of the system, the hazards that could be caused by a function should be identified and the resulting risks should be quantified. Depending on this risk definition, the safety goals are defined and hereby the requirements for the development methods, quality assurance, and monitoring of the entire product life cycle. The simplified flow of the safety development process as per ISO 26262 with the relevant original chapter headings is shown in Fig. 1.

The method defined by ISO 26262 thus ensures the integration of safety requirements right at the beginning of the development process. Thereby adequate methods are defined depending on product safety criteria. In particular, the quality requirements of the safety concept must be defined even before the product properties are specified in detail.

1.3 Differentiation Between Other Standards and Guidelines

To enable a wide application, even to very different systems, ISO 26262 addresses issues of functional safety on an abstract level. Thus, it inevitably provides less concrete information on the methods and procedures required for application, for example, conversion of controllability estimations and tests for risk assessment. However, the quantification of individual factors influencing the risk directly determines the hazard assessment and, with it, the safety standards to be applied to the system or function. Due to this, the current state of the art regarding objective, generally accepted evaluation methods, and metrics for evaluating the frequency of the driving situation, controllability, and severity of damage should be taken into account during the risk evaluation. In the field of driver assistance systems relying on surround sensor systems, the "Code of Practice" (PReVENT 2009) (from the project "PReVENT") provides a reference. It summarizes the state of the art for evaluation methods and metrics and includes procedures for the evaluation of the human-machine interface. The Code of Practice contains extensive questionnaires to evaluate the interactions of assistance systems and drivers. Depending on the situation, these questions are classified in 18 categories (e.g., predictability, trust, traceability). With the documentation of the state of the art, even if it may be a few years old, the Code of Practice offers definitions and examples for the practical implementation of hazard assessment and risk evaluation.

Due to its cross-functional objective, the ISO 26262 is different from the function-specific standards for driver assistance systems. These standards, such as ISO 15622 for Adaptive Cruise Control (ACC), define functional areas, scopes, minimum requirements, and test methods related to the current driver assistance function.

1.4 Differentiation from Handling of Other Failure Sources

While considering the functional safety of systems according to the specifications of ISO 26262, the distinct limits of fault types need to be defined (a detailed differentiation of the meaning of fault, error, and failure is included in the Glossary of ISO 26262). One limit is the restriction to electrical/electronic (E/E) and programmable systems, mechanical faults, for example, will not be considered as a cause of risks.

Similarly, the standard considers only the functional faults (in the context of a deviation from an explicit specification) of a driver assistance system. This assumes that it is possible to differentiate clearly between a status that is compliant with the specifications, the correct function, and a status that is noncompliant with the specifications, a fault/failure or a malfunction. These faults can clearly be identified and can be reduced to a minimum, for example, with fault detection methods with a very high detection rate, switch-off measures, and degradation mechanisms.

Interventions of driver assistance systems with surround sensor systems that are considered to be incorrect do not only occur due to faults of E/E components or software. Another type of failure occurs even though all components and software function comply with the specifications. Here, the system does not react in a way that is appropriate for the situation because the system specification does not cover all possible cases that can occur while operating a vehicle. The incorrect reaction of the system in the driving situation results, for example, from an incomplete perception of the situation or based on a prediction or model assumptions that are not feasible. According to the current state of knowledge, due to the large number of possible driving situations, systems relying on surround sensors can neither be specified distinctively nor can they be tested such that incorrect interventions occur only outside the specification (Weitzel et al. 2015). A more abstract formulation of the system specification only shifts this problem, as concrete system reactions have to be defined based on the latest available information about the surroundings of the vehicle at the time of technical implementation. Further discussions on this problem are given in Sect. 4 of this chapter.

Even when the technical causes are different for both failure types, the effects that the driver perceives in the vehicle are probably similar (e.g., a vehicle deceleration without any perceivable reason).

The procedure described in ISO 26262 for defining the safety requirements on the basis of an objective risk evaluation can be applied for both fault types. From the point of view of the driver and other people involved in the situation, it is irrelevant if the hazard results from functional faults or inadequacies of the system as long as the effects on the situation are the same or at least similar. In Ebel et al. (2010), this is denoted as a holistic approach, which allows a comparative evaluation of the effects of technical faults as well as of functional inadequacies. However, to what extent should the accepted risk limit and the evaluation metrics (of the resulting permissible fault rates that are to be applied according to ISO 26262) be compared? Which approach is feasible to solve this issue is still discussed among experts. Thus, in Ebel et al. (2010) an accepted risk limit is required at the system level for both fault types, while in (Ross 2014) the functional inadequacies are clearly associated with safety of use.

2 Safety Requirements for Driver Assistance Systems

ISO 26262 requires that during the design phase, the risks associated with the E/E system be identified and the risk potential be evaluated. This is done on the basis of the hazard analysis and risk assessment (H&R), a method whose application is normatively prescribed in Part 3 of ISO 26262 (3–7 in Fig. 1). In the scope of H&R, potential hazards are analyzed at the vehicle level without considering the causes and the detected risks are classified. Here, vehicle level according to (Ross 2014) means the implementation of one or more systems to which ISO 26262 is applied and thus denotes the top most abstraction level in the overall "vehicle system." To prevent the hazards defined in the scope of H&R, ISO 26262 requires the specification of safety goals (3–5 in Fig. 1). These provide the framework for development of the safety concept. The hierarchical structure of safety requirements is shown in Fig. 2. The safety goals are defined at the vehicle level taking into consideration the influencing factors and situations from the environment and subsequently derived as functional safety requirements in the vehicle system for the functions involved in the hazard (3–8 in Fig. 1). This derivation is still "without solution" and thus independent of the concrete implementation. The implementation of one or more functions is only done at the concrete system level, addressed by the technical safety requirements (4–6 in Fig. 1). As opposed to the functional safety requirements, technical safety requirements describe the implementation of the system, and, in the next derivation step, HW and SW safety requirements are given in detail accordingly for implementing the hardware and software (Parts of 5 and 6 in Fig. 1).

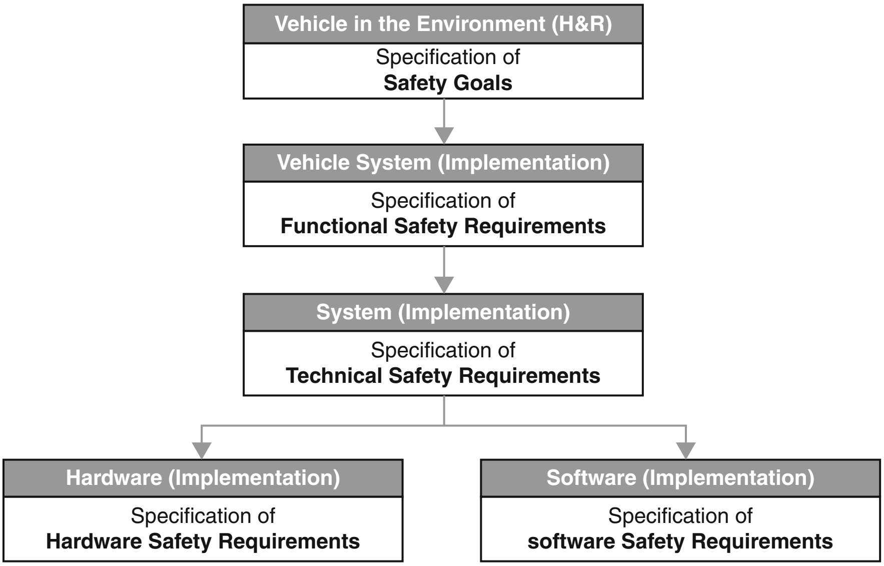


Fig. 2 Hierarchical structure of safety requirements

This safety requirement flow is presented in more detail in the following, using of the "Autonomous Emergency Braking (AEB)" function as an example. The AEB should perform automatic emergency braking in case of an imminent collision with the vehicle in the front. Depending on the extent of intervention, the function is designed for minimizing damage or preventing damage (see ▶ Chap. 46, "Fundamentals of Collision Protection Systems").

2.1 Specification of Safety Goals

2.1.1 Basic Principles

ISO 26262 defines the method for analyzing and classifying hazards normatively. This is done on the basis of the following three risk parameters:

- Exposure: How often do the driving situations occur, in which a driver or other road users can pose a hazard?
- Controllability: How well can the driver or other road users control the hazard in this driving situation so that damage can be avoided?
- Severity: When damage occurs, what is its severity?

Figure 3 shows the relationships between the three risk parameters in a risk diagram. This diagram helps to show how a risk evaluation is carried out. The product risk consists of the factors severity of damage (X-axis) and frequency (Y-axis). With reference to the risk parameters mentioned above, the *severity of*

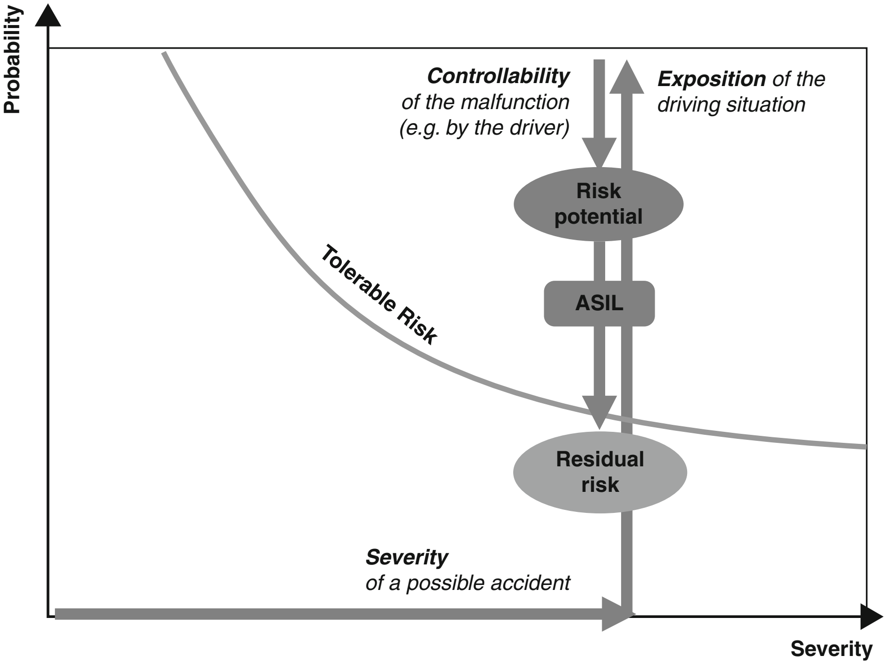


Fig. 3 Risk diagram

damage is evaluated with the parameter of the same name. The frequency results from the frequency of the driving situation and the controllability. The higher the *frequency of the driving situation*, the higher is the resulting risk potential. The risk can be reduced only by means of high *controllability*, which will be present when the road users involved in the critical situation have the possibility to prevent the damage.

The result of the risk evaluation is the risk potential which is evaluated with an ASIL ("Automotive Safety Integrity Level") in the classes QM, ASIL A, ASIL B, ASIL C, and ASIL D. Here, ASIL A corresponds to the lowest and ASIL D corresponds to the highest risk potential. Based on ASIL, the measures for preventing and controlling systematic and random faults are defined normatively so as to reduce the risk potential to such an extent that the remaining risk lies below the tolerable risk. In ASIL D, more methods and actions are to be applied to ensure safety than in ASIL A. However, if a hazard is classified only with QM (Quality Management), then the application of a certified quality management system is sufficient and the further application of ISO 26262 to the development process is not required.

The ASIL classification is defined concretely with the help of evaluation of the individual risk parameters according to Table 1. Here, each risk parameter is classified into three to four classes. Each class corresponds to a different meaning. The annex of Part 3 of ISO 26262 contains informative tables with references and

Table 1 ASIL determination matrix

Severity	Exposure	Controllability		
		C1 (simple)	C2 (normal)	C3 (severe)
S1 (light/medium)	**E1** (very low)	QM	QM	QM
	E2 (low)	QM	QM	QM
	E3 (medium)	QM	QM	A
	E4 (high)	QM	A	B
S2 (severe, likely to survive)	**E1** (very low)	QM	QM	QM
	E2 (low)	QM	QM	A
	E3 (medium)	QM	A	B
	E4 (high)	A	B	C
S3 (risk to life, unlikely to survive)	**E1** (very low)	QM	QM	A
	E2 (low)	QM	A	B
	E3 (medium)	A	B	C
	E4 (high)	B	C	D

examples using the risk parameters in which the severity of hazards of accidents, driving situations, and controllability of the driver have to be evaluated. Driving situations that occur in almost every drive (e.g., a drive through the countryside) are classified as E4 and driving situations that arise at least once in a month (e.g., driving with trailers in traffic jams) are classified as E3. Driving situations that occur even more rarely are then accordingly classified as E2 (e.g., driving on ice and snow) or as E1 for very rare situations (vehicle on chassis dynamometer). This example itself shows that an objective evaluation is not always possible as driving situations can be evaluated differently based on the considered environment. For example, driving on ice and snow in the northern regions of Europe will surely be deemed to occur than in the southern regions.

2.1.2 Safety Goals in the Example of AEB

The AEB function activates the actuator "brake." Based on the actuator function "braking," the following malfunctions can be defined as potential hazards:

(a) Unintentional automatic braking
(b) Unintentional brake boosting during driver braking
(c) No braking in spite of braking request
(d) Too low braking in spite of braking request
(e) Unintentional holding force during standstill

In the first step, the risk potential of the hazards of the "brake function" is defined independently of the driver assistance function. This is done on the basis of H&R with the risk parameters Severity (S), Exposure (E), and Controllability (C) as per Table 2. There, the hazards (a) and (b) are classified as ASIL C. The evaluation is done without considering the safety measures (e.g., deactivation of brakes by the driver or limitation of brake intervention).

Table 2 H&R with the example of actuator function "braking"

No.	Malfunction	Situation	Hazard	S	Reason S	E	Reason E	C	Reason C	ASIL
(a)	Unintentional autonomous braking (assumption: vehicle remains stable)	City driving/ countryside/freeway, heavy traffic with too short safety distance	Rear-end collision by the vehicle behind	2	Accident statistics show light to heavy injuries	4	Heavy traffic in almost all driving situations	3	Controllability difficult for the drivers who are following because safety distance is too short	C
(b)	Unintentional brake boosting (assumption: vehicle remains stable)									

Table 3 Safety goals with the example of actuator function "braking"

No.	Safety goal	Safe state	ASIL
(a)	Avoid unintended braking as it would lead to a hazard	Interrupt autonomous braking	C
(b)	Avoid unintended braking as it would lead to a hazard	Interrupt brake boosting	C

The ASIL that is determined is used for further development of the product as a standard for risk reduction. Here, safety goals are specified for preventing the hazard "rear-end collision." These being at the highest abstraction level (top-level safety requirements) are the starting point for developing the safety concept. Additionally, the "safe state" has to be specified for each safety goal. For the example from Table 2, the safety goal and the safe state are described in Table 3.

Safety goals address all systems that can directly or indirectly access the actuator under consideration and thus have the potential to cause the hazards (a) to (e). The hierarchical structure of safety requirements from Fig. 2 requires the specification of the safety requirements from the vehicle level up to the HW and SW level in more detail. This safety requirement flow ensures the fulfillment of the given safety goals and thus the prevention of the identified hazards.

2.2 Specification of Safety Requirements

2.2.1 Basic Principles

Safety requirements are specified based on the safety goals at every level of the vehicle and system architecture (see Fig. 2). At the vehicle level, the safe behavior of the function, which can potentially damage the safety goal, is specified in the form of functional safety requirements. This is done in all the systems that are involved and is a part of the requirement specification sent to the individual system providers. The functional safety concept (3–8 in Fig. 1) describes the assumptions and solutions with the help of which functional safety requirements were derived from the safety goals.

The requirement specification with the functional safety requirements are created by the party responsible for the vehicle system, this is generally the OEM. The system providers have the responsibility of specifying and implementing the safety concept in such a way that the functional safety requirements from the system are not affected. The technical safety concept contains the documentation of deriving the technical safety requirements from the functional safety requirements. This contains the concrete solution at the system level for preventing the violation of the functional safety requirements and, with it, implicitly for preventing violation of the superordinate safety goal.

2.2.2 Safety Requirements in the Example of AEB

An example of safety requirements for the AEB function at the vehicle and system level is given in Fig. 4. The components of the actuator function of braking are the “Driver Assistance System (DAS)” and the “Electronic Stability Control (ESC) system.” The safety goal “Avoid unintended braking, which leads to a hazard” must then be passed on to both the systems in such a way that it is not violated. This means for the DAS that an unjustified AEB request that could lead to a hazard has to be prevented. By limiting the AEB function in terms of duration and, in part, even in terms of strength, the risk potential that can be reduced as a shorter or weaker braking has a positive influence on the extent of the damage as well as on the controllability. The H&R described in ISO 26262 has its limits when it involves defining this risk reduction in a concrete manner. Usually, because of the rough classification of the risk parameters and the related subjective estimation, it is difficult to evaluate the brake interventions with different brake profiles based on their risk potential. In this case, objectified methods, such as simulations on the basis of equations of motions or use of endurance test data and accident statistics, can help further in defining the ASIL classification of different braking profiles more precisely.

Assume the AEB intervention is limited such that the risk potential can be reduced to an ASIL A. In this case, this ASIL classification, along with the functional safety requirement “Avoid unintended brake request within the AEB specification,” is transmitted to the system FAS. However, to avoid that due to a fault, AEB braking is done outside the specified limits, these limits must also be

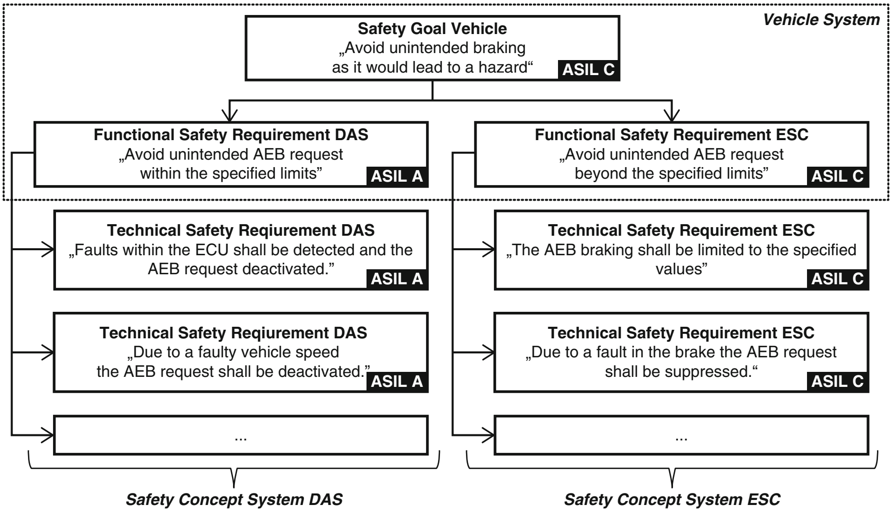


Fig. 4 Hierarchical structure of safety requirements with the example of AEB

secured. Usually, the AEB requirement is limited by the system ESC with an ASIL C derived from the original classification of the respective safety goal.

2.2.3 Decomposition Methods

In addition to the inheritance of safety requirements, in which the ASIL is assigned to at least one derived safety requirement, ISO 26262 also offers the possibility of decomposition. Here, the ASIL of the derived safety requirements can be reduced. However, that is only possible if there is redundancy used. According to ISO 26262-9, Chap. 5.4, a decomposition and with it the reduction of ASIL can only be done if the safety requirement "before decomposition" (initial safety requirement) can be implemented by means of at least two *sufficiently independent* elements or subsystems. Each derived safety requirement must thus be able to fulfill the initial safety requirement "on its own." Here, elements or subsystems are considered as "sufficiently independent" when on the basis of the analysis of dependent faults (see ISO 26262-9, Chap. 7), no *common cause* (CCF, see ISO 26262-1, 1.14) or *cascading failure* (see ISO 26262-1, 1.13) is determined, which would lead to violation of the initial safety requirement. Common cause describes the failure due to a common reason, while a cascading failure causes further failures (Ross 2014).

In Fig. 4, the ASIL is inherited to all the systems involved. Here, the system ESC receives the ASIL of the initial safety goal. The reduction of ASIL in the DAS is not done by decomposition (there is no redundancy here) but by reevaluating the risk potential with limited braking function. Due to the fact that the safety mechanism for ensuring the specified limits is implemented in an independent electronic control unit, according to ISO 26262, it is absolutely permissible to reduce the ASIL to the actual risk potential in the DAS. However, if hypothetically no ASIL C requirement has to be implemented in the ESC system, then it is conceivable to distribute the ASIL C classification to both systems by observing the specified limits even in the DAS. In Part 9, Chap. 5, ISO 26262 offers different options for reducing the ASIL classification. However, here, the initial ASIL should always be given in brackets. Then, for example, ASIL C can be decomposed into an ASIL A (C) and ASIL B (C). For the safety requirements for the AEB function, the change would then result in Fig. 5.

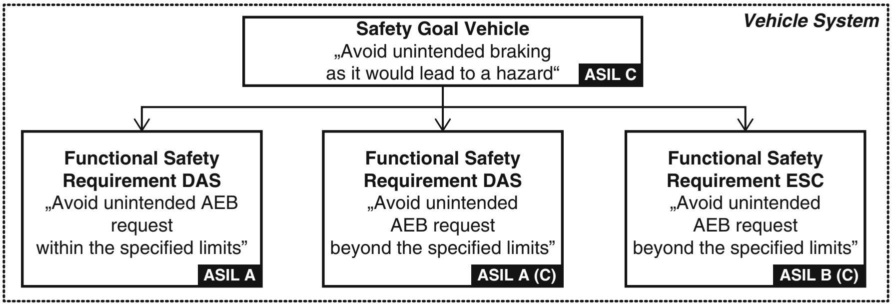


Fig. 5 Application of decomposition with the example of AEB

3 Meeting the Safety Requirements

After the risks resulting from a product are compiled systematically with the help of H&R (3–7 in Fig. 1) and safety requirements are derived from the safety goals, ISO26262 requires that the product development must ensure that these are actually implemented.

According to the basic V-Model (2013), a complete requirement tree must first specify the product properties. This specification at all levels of detail serves as a basis for the right branch of the V-Model, the verification and validation of product properties.

3.1 Traceability of Requirement Levels

If one assumes that the safety goals were validated, then, according to ISO 26262, a product is safe only if it can be proven that the solution fulfills the safety goals. Accordingly, it is not enough to describe precisely which algorithms are implemented or which hardware is used. The solution must be unambiguously linked with the abstractly specified properties, especially the safety goals, for validation purposes. Creating such complete documentation is very challenging in the case of complex systems.

In a first attempt to specify a RADAR-based AEB system, the following requirements might be derived:

(a) The internal resistance of the heating wire of the lens heater must be xx Ohm.
(b) The RADAR sensor must detect reflection points using a fixed-beam-antenna.
(c) AEB must not brake in situations that have no accident risk.
(d) The driver must be able to override the AEB behavior at any time.
(e) A Kalman filter must be used for object tracking.
(f) The vehicle should not cause any accidents due to unauthorized brake interventions.
(g) When activating the brake light switch (BLS), an automatic emergency brake control is interrupted and the deceleration requested by the driver is applied.
(h) Etc.

All requirements given in this example are valid product requirements. However, they are expressed at very different abstraction levels. Requirement (f) corresponds to a general safety goal. Requirements (c) and (d) are abstract descriptions of behavior that do not anticipate the concrete HW and SW solutions as yet. Requirements (b) and (e) describe HW components and SW algorithms without anticipating the implementation details. Requirement (a) in this example is the most concrete requirement, which refers directly to a HW implementation.

After deriving the functional safety concept (3–8 in Fig. 1), requirement (d), for example, was introduced as a safety requirement. The core question that arises here is whether the compliance of the internal resistance according to requirement (a) is

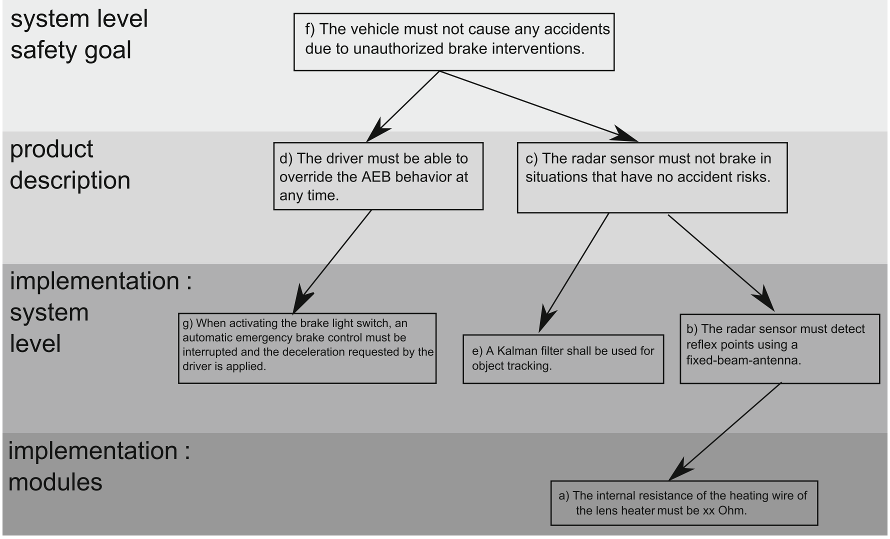


Fig. 6 The requirements can be structured in a requirement tree: This is the basis of traceability

relevant for safety or is motivated "only" by quality aspects. A hierarchical requirement tree that orders the requirements yielding a trace from most abstract requirement to concrete solution definition is necessary to answer such questions (Fig. 6).

With each level, the solution space is limited further. The tree can also be understood as a representation of nested solution spaces (Fig. 7).

In this representation, the benefits of traceability are particularly noticeable: All methods of ISO 26262 are geared toward proving that the concrete requirement and its implementation are a subset of the more abstract specifications. In the subset image, a solution is not allowed if it leaves the solution space spanned by the abstract requirement.

In practice, it often happens that the hierarchy levels are mixed up inadvertently. For example, the developer specifies a concrete algorithm at the level of the behavior description of the DAS. This leads to early, unnecessary limitations in the system design and also makes it more difficult to ensure the traceability, but is insignificant for compliance with ISO 26262. ISO 26262 concentrates on the complete consistency of the requirement tree with reference to the safety requirements only.

To get a better system overview, it is useful to compile all the requirements belonging to one solution space into one system or subsystem description. For example, the requirements (c) and (d) are at the same description level and both describe the behavior of the "feature" AEB. A complete set of such requirements is also called a model: An "output" (system reaction) is always assigned to an "input" (driver activity, environment situation).

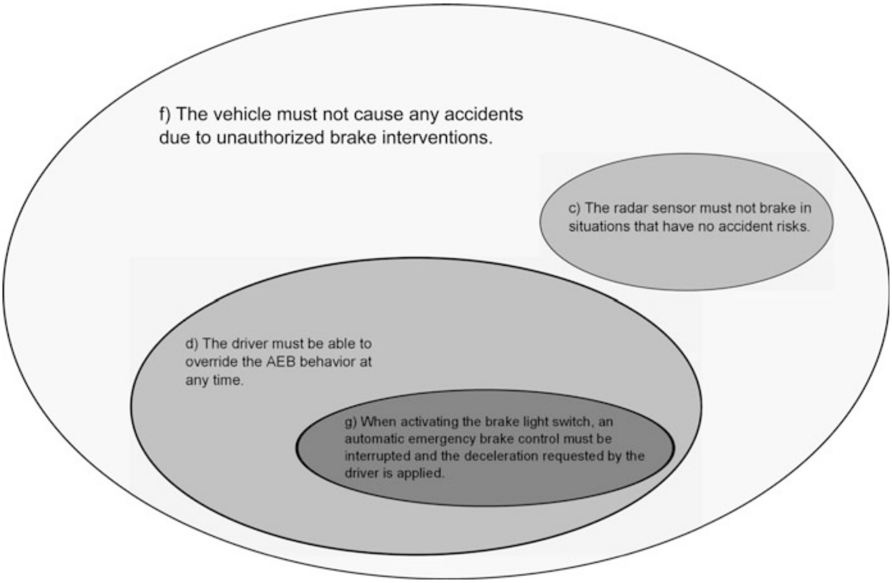


Fig. 7 The requirement tree represented as nested solution spaces. The requirement "The system is not allowed to cause any accidents" defines a solution space that is limited by the concrete solution, "The driver must be able to override the system at any time"

In ISO 26262, four broad hierarchy levels are defined with the corresponding models:

1. Abstract system level specified by means of safety goals (requirement (f))
2. Solution-free product description consisting of functional safety requirements (requirements (c) and (d))
3. Requirements at the implementation level that could be relevant to more modules and do not refer yet to the smallest architectural element: Technical safety requirements (requirements (b), (e), and (g))
4. Implementation requirements relevant to the smallest structural element: HW/SW safety requirements (requirement (a) in this example only at the HW level)

The requirement tree shown here as an example is obviously incomplete insofar as no system can be developed with the help of this simple "specification." Moreover, excessive gaps prohibit an easy understanding of the trace: Is the property of the heating wire really not linked with requirement (f) (no unauthorized brake interventions)? This cannot be answered *clearly* due to a few text lines!

This example demonstrates the challenge posed by a complete, unambiguous, traceable requirement tree for complex driver assistance systems. That is why, while creating the requirement tree, the structuring system architecture is an important element of mastering a safe system design. With the exception of the hierarchy levels required in ISO 26262, it is generally left to the system developer to decide how many derivation steps are to be provided on the way to the smallest structural element of his architecture. Small steps make it easier to trace the individual steps, but let the number of elements and the complexity of the link structure grow very fast.

ISO 26262 demands the complete traceability only with reference to the safety goals derived in H&R. Since a well-documented requirement tree not only makes it possible to track compliance with safety goals but also helps the product quality in general, traceability is a standard requirement in the automotive industry, even without a safety goal to achieve.

In the example shown here, the requirement tree derived from the safety goals is separate from the tree resulting from performance requirements for the AEB function. Often ISO 26262 implicitly assumes that a malfunction at system level is basically always caused by a fault, that is, a deviation from the implementation requirements. In this case, the differentiation between the safety architecture and functional architecture is very clear, as the safety requirements are basically limited to monitoring the specified implementation and reactions to this monitoring. The typical "safety architecture" protects the functional part of the system from implementation faults.

This differentiation is no longer valid if the safety concept also considers the system limits that affect, for example, requirement (c) (AEB should not brake in situations that do not have risk of accidents). A faulty reaction could lead to a faulty safety-relevant behavior, in spite of correct implementation in HW and SW. In this case, all the requirements that result from a safety goal are relevant for safety; explicitly also those that specify the function of the subsystems.

In the current version of ISO 26262, such system limits are explicitly excluded from the scope of the standard: "*ISO 26262 does not address the nominal performance of E/E Systems, even if dedicated functional performance standards exist for these systems*." The challenges posed by it to the application of ISO 26262 for driver assistance systems are described in detail in Spanfelner et al. (2012). In this publication, a reference is made to the fact that in spite of full compliance with ISO 26262, the remaining risk of driver assistance systems can always lie above the tolerable risk (see Fig. 3). Reasons are functional inadequacies that lead to the critical question for safety-relevant driver assistance systems: "How safe is safe enough?" (Ebel et al. 2009).

3.2 Verification

A prerequisite for a valid verification is the complete description of system specification at the defined abstraction levels given in the previous section.

The verification aims to prove that the input/output relationship, which is specified in the models, is implemented by corresponding elements of the product. This is the domain of the requirement-based test.

A correctly derived model, for example, the definition of requirements at the implementation level, is equivalent to the abstract system specification. Accordingly, it would actually be sufficient to verify only the model at the system level in the product. In practice, however, it is not possible to have a complete verification in case of more complex models. The more abstract the specification, the more input/output relations that have to be checked. Since the specification becomes really definite only at the implementation level, a certain degree of completeness can only be practically achieved here.

ISO 26262 tries to alleviate this problem in two ways. Firstly, the verification is supported not only by simple test methods or reviews, but also, depending on the ASIL classification, further, stronger, formalized methods from the state of the art are recommended (e.g., verification of software code by applying the "abstract interpretation"). Going deeper into the different methods would exceed the scope of this overview. Even the ISO 26262 itself only refers to established method descriptions here. Secondly, the verification is repeated at all abstraction levels. A simple SW module can be tested in a more complete manner than a complex, interactive system. Specific integration tests for each integration step identified in the design are also part of the verification at all abstraction levels.

3.3 Validation

The verification basically proves that two equivalent models at the same abstraction levels of specification and implementation actually correspond to each other.

The safety goal derived from the H&R, however, is consciously formulated in an abstract and general manner in such a way that even design faults, that is, models that are derived in a faulty manner, could lead to a violation of these top-level requirements (see Fig. 8).

The phase 4–9 "Safety Validation" is specifically dedicated to the task of ensuring the validity of derivations from abstract requirements. Additionally, the completeness and correctness of the actual safety goals need to be validated. This includes the check of whether the driver can bring the system to a "safe state." The required methods concentrate on system tests on the product level and especially focus on the robustness of the system to "faults," that is, implementation elements that erroneously do not meet the specification.

In practice, according to (Balzert 2008), 55 % of all faults appear in the requirement and design phase itself while defining the abstract requirements for detailed technical models. These faults are especially difficult to detect because all the subsequent design steps and the tests of implementation levels cannot reveal such faults.

ISO 26262 relies on reviews by experts to detect such faults. The prerequisite here is that the experts have a 100 % understanding of the design step as per the state of the art. Correspondingly, ISO 26262 also emphasizes the value of simple designs as a basis for safe systems.

A design step that transforms an abstract system model into a concrete technical model always forms the basis of a derivation model. Only with the help of such a

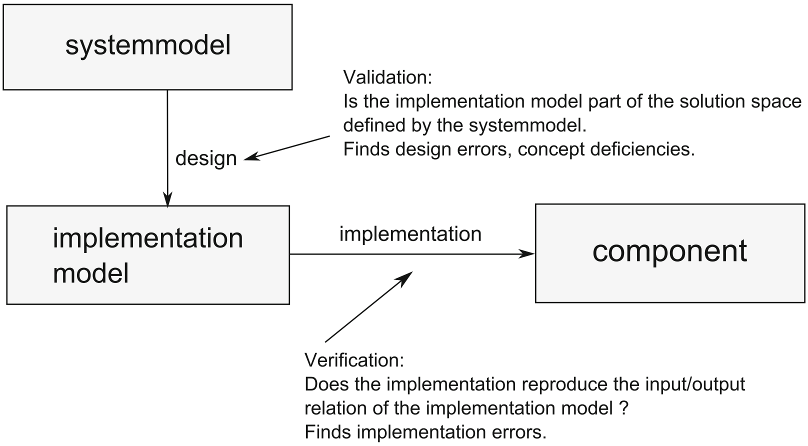


Fig. 8 Representation of the terms – validation and verification. The validation focuses on the correctness of the design step: vertical in the V-Model. The verification ensures the equivalence of implementation and implementation model: horizontal in V-Model

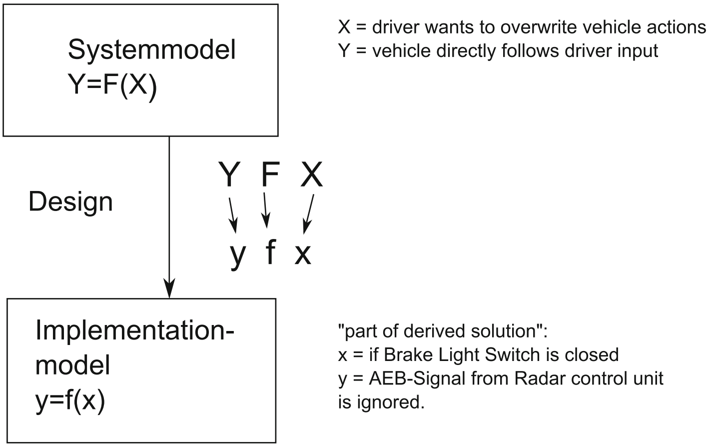


Fig. 9 The design connects the abstract with the concrete model. A "derivation model" forms the basis of the figure

derivation model is it possible to show that the concrete model lies in the solution space of the abstract model. This explanation forms the basis of validation of the concrete model (see Fig. 9).

If, for example, for meeting the requirement (d) (override) the designer selects an electromechanical brake light switch, which is read into an electronic control unit via an A/D converter, then the physical models of the operating mechanism of the switch and of the A/D converter form the basis of this decision. These components are designed with the help of basic derivation models in such a way that they always fulfill their task, as long as the basic hardware components comply with their specifications.

ISO 26262 implicitly assumes that the experts designing the system and those reviewing the system always have access to such a derivation model in the scope of state of the art. Even if this is not required explicitly in the form of documents, a complete requirement tree can be prepared only with the reference to such a derivation model.

4 Limits of ISO 26262

In the previous example, the safety goal "The driver must be able to override the AEB control at any time" could be achieved with a comparatively simple solution with the help of a brake light switch.

The safety goal "Never brake in a situation where there is no risk of an accident" is harder to derive using a complete derivation model. In this case, the safety goal is identical with the system requirement defining the actual function of the system. A "safety architecture" that is separated from the core function is clearly more difficult, if not impossible. Accordingly the traceability must be ensured along the functional path, which could lead to intrinsic gaps in complex, interpreting, and predictive systems.

4.1 Gaps in Traceability

One part of the system function is to predict the behavior of pedestrians in different situations: The system model maps all situations (input) in which the behaviors of the pedestrian and driver do not lead to an accident to the system reaction (output) "Don't react." This leads to the question if there is a derivation model which makes it possible to derive an algorithm which shows the measured initial state of the situation in case of the predicted accident. This model must be able to predict the behavior of all possible pedestrians in all possible situations. This is generally not possible.

The gap in this model lies in the fact that the intention of relevant players in the given situation is not known.

The assumptions for the conventional validation methods, a complete derivation model as per state of the art, are therefore not available.

4.2 Handling the Unknown in the Design Process

Are there comparable "gaps" in conventional systems? The core of the gap is the lack of knowledge of the state of a system element or an environmental "load."

The validation discussed in the previous chapter is based on a complete understanding of the derivation model assuming that the hardware requirements that form the basis of the design should be fulfilled. The hardware properties can be checked roughly after manufacturing and before releasing the product. It is more difficult to

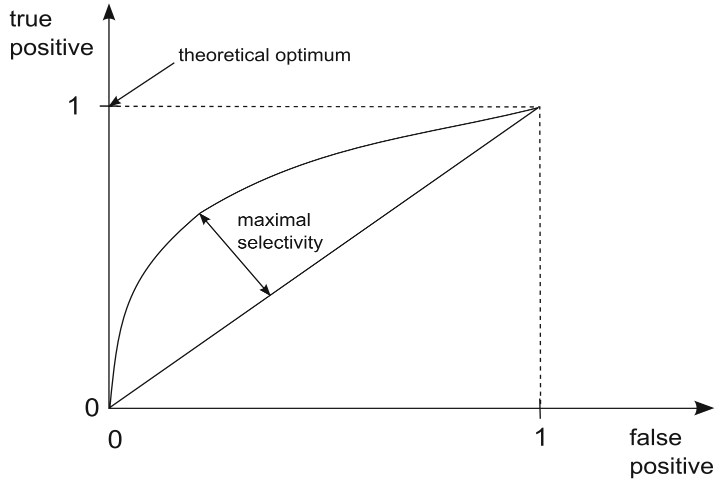


Fig. 10 Detection accuracy of a classifier can be set using parameters. When the fault classifications are reduced, i.e., "false positive" probability, then the detection accuracy is also reduced, i.e., "true positive" probability. On the other hand, increased detection accuracy always goes together with an increased fault classification probability. The gap between the reference lines corresponds to the selectivity and, with it, to the performance of the classifier. The maximum selectivity represents a compromise between the detection accuracy and fault classification probability

model the properties over a longer period and when environmental loads are not precisely known.

A fault of a specific component after a known load leads to a basically predictable, systematic fault. When the fault occurs, the reason and the breakdown process can be explained. However, for an exact prediction, the exact property of the component, possibly at an atomic level, as well as the exact load schema over the lifetime must be known.

As this information according to state of the art cannot be known for all products that are delivered, the individual fault appears incidental. To design the components safely, statistical models are referred to during development. The lack of knowledge about individual properties of the component is thus averaged across a large quantity of components.

Similarly to the statistical material modeling our example, the individual behavior of the pedestrian in a specific situation can be modeled statistically. This results in an algorithm which is often correct across a large number of similar situations ("true positive") and, with a certain probability, wrongly estimates the situation ("false positive"). A wrong estimation that leads to an undesirable behavior that is inconsistent with the specification at the abstract level is referred to in the following as a faulty behavior. The probability of such a faulty behavior is called "residual fault probability."

This residual fault probability can usually be adjusted using model parameters in driver assistance systems (see Fig. 10). In contrast to the release of designs based on complete derivation models, where the design review confirms the absence of relevant faults, a false positive rate of the statistical model is known when the system is released. The system tests are used to confirm the probability of occurrence and with it the conformity with a targeted residual fault probability.

The residual fault probabilities must be defined for designing such a system as part of the system specification. This is also standard for designing HW components. The conventional methods for modeling the fault frequency at the total system level are derived by means of rough estimations either quantitatively (Fault Tree Analysis, FTA) or qualitatively (Failure Mode and Effect Analysis, FMEA).

The gap in traceability of the design that is described here is often called as "functional insuffencies."

4.3 Validation of Systems with Functional Insuffencies

For systems with functional insuffencies, the adherence to an upper limit (Ebel et al. 2009) must be proven for the residual risk inherent in the system. In the simplest case, a statistical "Black-Box Test" should be carried out at the system level. This measures the probability of faulty behavior without knowledge of implementation details using trials on a statistical representative sample. For driver assistance systems this means a lot of test hours under driving conditions that are representative for the function.

Even in case of comparatively "noncritical" functions, depending on the derivation of the acceptable residual risk, residual fault rates of $<10^{-5}$/h have to be proven. Unfortunately, ISO 26262 does not give any concrete steps for deriving such residual fault rates. For critical functions, the "Black-Box Test" can quickly become more expensive than the entire development or, in the time frame of a product development, may not be realistically possible to execute in time. Another special challenge is to prove that the test kilometers that have been driven are representative enough for the situations that can be expected in the field. Which road profiles are to be used under which weather conditions, in which countries, with which drivers (Weitzel et al. 2015)? The statistical models of driver behavior are not always available. It is often necessary to take plausible expert opinions in these cases.

Consequently, the statistical system tests are no replacement for a well-understood design derived as far as possible from robust models. During development there should be a focus on separating elements that can be designed using an adequate model, from elements in which gaps in the derivation model cannot be prevented. In the example of pedestrian protection, the imaging properties of a camera which has to detect the pedestrian should not be modeled statistically. Unique physical models calculate an angle to the object based on pixel mapping. The situation is different when the behavior of the individual pedestrian has to be predicted. Here it is necessary to statistically model the lack of knowledge about the intention of the pedestrian. The resulting algorithm will not estimate the intention correctly in every situation. Residual fault probabilities are the unavoidable consequence.

5 Summary and Outlook

The first part of this chapter shows how ISO 26262 standardizes important procedures for achieving functional safety in the automotive industry. The focus here lies on the prevention of undesired or dangerous system behavior caused by hardware and software faults.

With increasingly complex driver assistance systems, "functional inadequacies," as the reason for unwanted system behavior, are becoming more prominent. In this case, the challenge lies in the limitations of the system design, making it necessary to handle unknowns using statistical models.

One of the assumptions for the validation of such systems with functional insuffencies is the specification of release goals in the form of accepted residual probabilities for faulty behavior of the system. This requires a wide acceptance for the target values in society.

Either tested designs or accepted goals have been established for the hardware. Such a standard still has to be developed in the case of designing new predictive systems for driver assistance systems.

It is difficult to predict if this will be achieved by enhancing ISO 26262 or separate initiatives.

According to current estimates, the application of ISO 26262 itself is still being interpreted differently in individual companies. This is because the standard allows for individual interpretations. This assessment is supported by looking at contributions to the conferences organized specially on the topic "Functional Safety" in Germany, North America, and Asia. For example, the risk potential of identical hazards is classified differently. It is also not yet clear how some requirements and methods are to be implemented or applied in a concrete manner (e.g., conducting a SW safety analysis). For the 2nd edition of ISO 26262, on which work has begun in 2015, it remains a challenge to limit this freedom of interpretation further.

For the safety of the driver assistance systems, a new standard, be it in the scope of the 2nd edition of ISO 26262 or otherwise, should standardize the development goals, not the solutions. Driver assistance is still at the beginning of developments that are necessary for exploiting its full potential. An early rigidity in defined "operationally tested" solutions "motivated" by safety concerns could have the consequence that the benefits of the relevant system cannot be developed further. The original intention to provide more safety in this case might lead to the opposite effect: less protection and safety in traffic!

References

Bachmann V, Zauchner H (2013) Erste Erfahrungen mit dem Automotive-Standard ISO 26262 und Ausblick auf die Adaptierung für Motorräder (Initial experiences with the Automotive Standard ISO 26262 and view on adaptation for motorcycles), Vortrag (Presentation) TU Darmstadt, 23 May 2013

Balzert H (2008) Lehrbuch der Softwaretechnik – Softwaremanagement (Textbook on software technology – software management), 2nd edn. Spektrum Verlag, p 487

Börcsök J (2011) Funktionale Sicherheit (Functional safety), 3rd edn. VDE-Verlag, Berlin/Offenbach

Dardar R, Gallina B, Johnsen A et al (2012) Industrial experiences of building a safety case in compliance with ISO 26262. In: IEEE 23rd international symposium on software reliability engineering workshops (ISSREW), Dallas, 27–30 Nov 2012, pp 349–354

Ebel S, Wilhelm U, Grimm A et al (2009) Wie sicher ist sicher genug? (How safe is safe enough?) Anforderungen an die funktionale Unzulänglichkeit von Fahrerassistenzsystemen in Anlehnung an das gesellschaftlich akzeptierte Risiko (Requirements of functional insuffencies of driver assistance system based on the commercially accepted risk). In: 6th Workshop Fahrerassistenzsysteme (Workshop of driver assistance systems), Löwenstein, 28–30 Sept 2009

Ebel S, Wilhelm U, Grimm A et al (2010) Ganzheitliche Absicherung von Fahrerassistenzsystemen in Anlehnung an ISO 26262 (Complete safety of driver assistance systems in compliance with ISO 26262). In: Integrierte Sicherheit und Fahrerassistenzsysteme (Integrated safety and driver assistance systems), 26th. VDI/VW-Gemeinschaftstagung (community meeting), Wolfsburg, 6–7 Oct 2010, pp 393–405

IEC/EN 61508 (2010) Functional safety of electrical/electronic/programmable electronic safety-related systems (E/E/PES), 2nd edn: International Electrotechnical Commission, Geneva

ISO 26262 (2011) International standard road vehicles – functional safety, 2012

ISO 31000 (2009) Richtlinien und Prinzipien zur Implementierung des Risikomanagements (Risk management – principles and guidelines, provides principles, framework and a process for managing risk): International Organization for Standardization, Geneva

PReVENT (2009) Code of practice for the design and evaluation of ADAS, European Automobile Manufacturers' Association (ACEA), Brussels

ProdHaftG (2002) § 3 para 3 Gesetz über die Haftung für fehlerhafte Produkte; Ausfertigungsdatum: (Law on liability for defective products, date of issue) 15.12.1989, Last changed by Article 9 para 3 G v. 19.7.2002 I 2674; also refer BGH judgement 16th June 2009 – VI ZR 107/08

Ross HL (2014) Funktionale Sicherheit im Automobil (Functional safety in automobiles). Carl-Hanser Verlag, München/Wien

Spanfelner B, Richter D, Ebel S et al (2012) Challenges in applying the ISO 26262 for driver assistance systems. In: 5 Tagung Fahrerassistenz (5th Conference Driver assistance systems), München, 15–16 May 2012

Teuchert S (2012) ISO 26262 – Fluch oder Segen? (Curse or blessing?) ATZelektronik, 6 edn, Springer-Verlag, Heidelberg New York, vol 7. pp 410–415

V-Model (2013) http://www.cio.bund.de/Web/DE/Architekturen-und-Standards/V-Modell-XT/vmodell_xt_node.html. Accessed 16 July 2013

Weitzel A, Winner H, Cao P, et al. (2015) Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung; Forschungsbericht der Bundesanstalt für Straßenwesen, Bereich Fahrzeugtechnik (Safety strategies for driver assistance systems with surround sensor systems, research report for Federal Highway Research Institute, Automotive Engineering). Verlag neue Wissenschaft, Bremerhaven

Werkmeister K, Englisch H (2012) Die ISO 26262 für Motorrad Erfahrungen bei der Umsetzung bei BMW Motorrad (ISO 26262 for motorcycle, experiences with implementation for BMW motorcycle). In: 9th Internationale Motorradkonferenz Institut für Zweiradsicherheit e.V. (International motorcycle conference institute for motorcycle safety), Cologne, 1–2 Oct 2012

AUTOSAR and Driver Assistance Systems

7

Simon Fürst and Stefan Bunzel

Contents

S. Fürst
BMW Group, München, Germany
e-mail: simon.fuerst@bmw.de

S. Bunzel (✉)
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: stefan.bunzel@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_7

Abstract

Software development for automotive applications has become the most sophisticated and critical activity during the vehicle development. AUTOSAR (AUTomotive Open System ARchitecture) develops a standardized open software architecture for automotive electronic control units (ECUs). AUTOSAR is a partnership of automotive manufacturers, suppliers, and tool and semiconductor vendors.

The focus is on managing the growing complexity in the development of automotive electric/electronic (E/E) architecture, with the aim to enable new technologies and improve development efficiency – without making compromises on quality. The standardization is realized in the three technical areas:

1. Software architecture
2. Development methodology
3. Application interfaces

Besides this focus AUTOSAR's key success factors are its development agreement that controls the legal part of the collaboration between the partners and its lean organization that distinguishes between the "Founding Fathers" (i.e., the core partners) and other partners (premium, development, and associate partners) and thus installs lean processes for fast decisions. As a result of its joint development activities, AUTOSAR has provided several releases since its foundation in 2002: The latest Release 4.2.1 has been delivered on October 31, 2014, and contains more than 218 concepts. By October 2014 more than 180 partner companies successfully share and live AUTOSAR's fundamental principle:

Cooperate on standardization, compete on implementation.

1 The Motivation for AUTOSAR

Over the last decades software development for automotive applications has been gaining more and more importance. Ever more demanding requirements on safety, environmental protection, and comfort have resulted in a sharp increase in the number of electronic systems to be found in vehicles. Increasingly stringent legal requirements on exhaust emissions and safety have also fed the trend, as have the numerous infotainment and driver assistance systems, whose functioning relies on the simultaneous interaction of a variety of different sensors, actuators, and control units. Meanwhile 90 % of all innovations are driven by electronics and software (cf. Von der Beek 2006).

This pace of development and the increasing integration of functions and control units pose a challenge for vehicle manufacturers. In order to manage growing system complexity and rising number of dependencies on the one hand while keeping the costs feasible on the other hand, the basic software and the interfaces to applications and bus systems have to be standardized in a future-proof way.

AUTOSAR (AUTomotive Open System ARchitecture) (cf. AUTOSAR 2015) is exactly working on this standard and has released several versions in the meantime. Mastering complexity is paramount and will be achieved by establishing reusable and exchangeable software modules, e.g., like driver assistance systems, between OEM and suppliers.

2 AUTOSAR Organization

In the light of this, leading automotive manufacturers and suppliers launched the AUTOSAR initiative on July 2003. Today's core partners are Bosch, BMW, Continental, Daimler, Ford, General Motors, PSA, Toyota, and Volkswagen. They are responsible for the controlling, administration, and organization of AUTOSAR. Furthermore, AUTOSAR invites companies to participate and bring in their knowledge and experience. In return they will benefit from the knowledge and experience of all the other partners. There are four types of partnerships: Besides core partners, AUTOSAR offers premium, development, and associate partnerships and last but not least for guests the attendee agreement as shown in Fig. 1.

By sharing their knowledge and by their active participation in the work packages, premium and development partners contribute to the development of concepts and specifications and thus determine the technical standard. All partners have the right to use the AUTOSAR specifications for commercial purposes. Research institutes and companies that can be assigned as service providers can obtain the temporary status of an attendee. Attendees participate in the developing of the AUTOSAR standard, but they do not possess a license for commercially exploiting the AUTOSAR specifications.

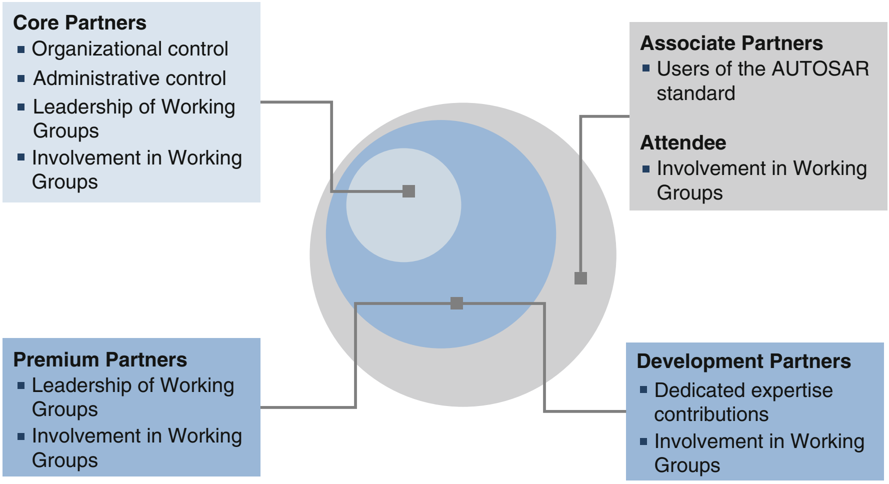


Fig. 1 Structure of AUTOSAR partnership

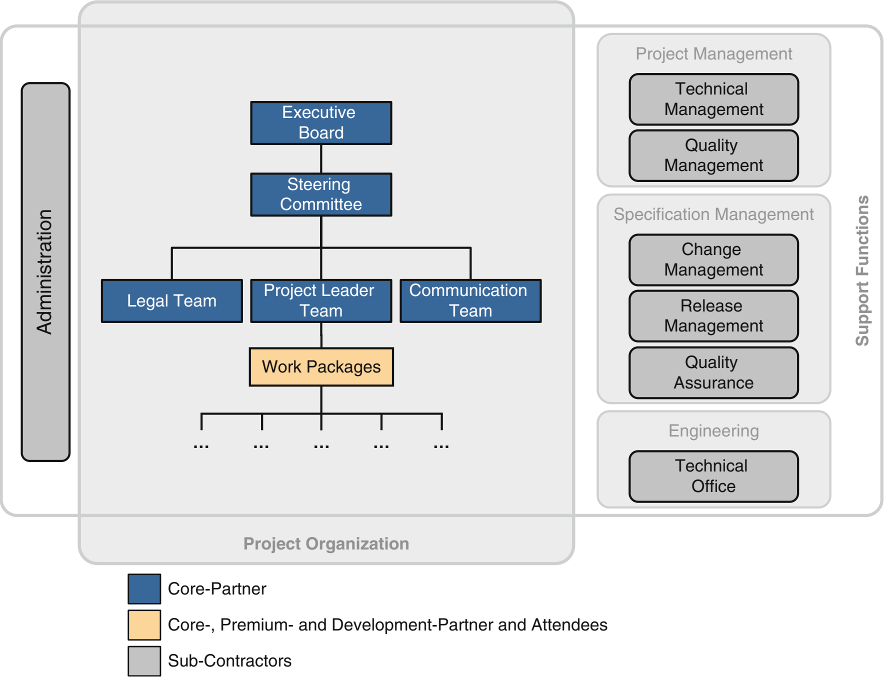


Fig. 2 Organizational structure of AUTOSAR

On the top level of the AUTOSAR organization, there are two boards who are composed of representatives of the core partners. The executive board is AUTOSAR's ultimate decision-making body and defines the overall strategy and roadmap of the partnership. The steering committee is AUTOSAR's top-level acting board. It coordinates the day-to-day nontechnical operation of the cooperation and is commissioned to map out AUTOSAR's long-term objectives.

Technical matters and the coordination of the technical work packages are in the responsibility of the project leader team. Like the executive board and the steering committee, the project leader team is a panel of the core partners.

Work on the AUTOSAR specifications is divided into a number of work packages (WP) which, if necessary, are divided into sub-work packages. Besides the core partners the premium and development partners and the attendees participate in and actively contribute to the work packages.

Figure 2 illustrates the placement and relationship of the different panels and working groups.

The AUTOSAR support functions assist the partners with administration, project and quality management, specification management, and the technical development of the standard.

Technical work was initially limited in time and divided into phases. Due to AUTOSAR's great success story of its phases I–III and its great potential for the future, AUTOSAR now runs open end until cancelled.

3 The Nine Project Objectives of AUTOSAR

Operational work within the partnership is derived from the higher aim to master the complexity of software modules, e.g., driver assistance systems, among OEMs and suppliers. This induces to our so-called nine project objectives:

1. Transferability of software. OEMs and suppliers shall be able to reuse software within the vehicle network. Thus, it is possible to use the same software on different car platforms at different OEMs.
2. Scalability to different vehicle and platform variants. AUTOSAR shall provide mechanisms for developing software systems that can be adapted to different vehicles and vehicle platforms and hardware. That means that AUTOSAR should be configurable so that it can be integrated in different vehicles.
3. Support of different functional domains. AUTOSAR shall enable the reuse of software components for as many functional domains as possible. This includes the exchange of data with non-AUTOSAR systems, e.g., communication to vehicle's infotainment systems.
4. Definition of an open architecture. AUTOSAR's architecture shall be maintainable, adaptable, and extendable. Hence, mistakes can be fixed continuously; future requirements and individual enhancements can be realized.
5. Development of highly dependable systems. Availability, reliability, functional safety, integrity, maintenance, and security shall be realizable by AUTOSAR. By this means, e.g., requirements on functional safety can be taken into account.
6. Sustainable utilization of natural resources. Use of technologies that provide efficient use of natural resources and use of renewable energy shall be supported.
7. Collaboration between various partners. The automotive industry is characterized by the extensive cooperation between partners. AUTOSAR shall support collaboration by defining data exchange formats and an architecture that allows the integration of basic software and applications from different partners.
8. Standardization of basic software functionality of automotive ECUs. The basic software shall be reusable for different functional domains, OEMs, and suppliers. Then the basic software can be offered as a product.
9. Support of applicable automotive international standards and state-of-the-art technologies.

AUTOSAR shall be compatible with existing and relevant international standards. This allows the use of AUTOSAR in current and future vehicle systems. One example is the support of existing and future bus systems such as FlexRay, CAN, Ethernet, etc.

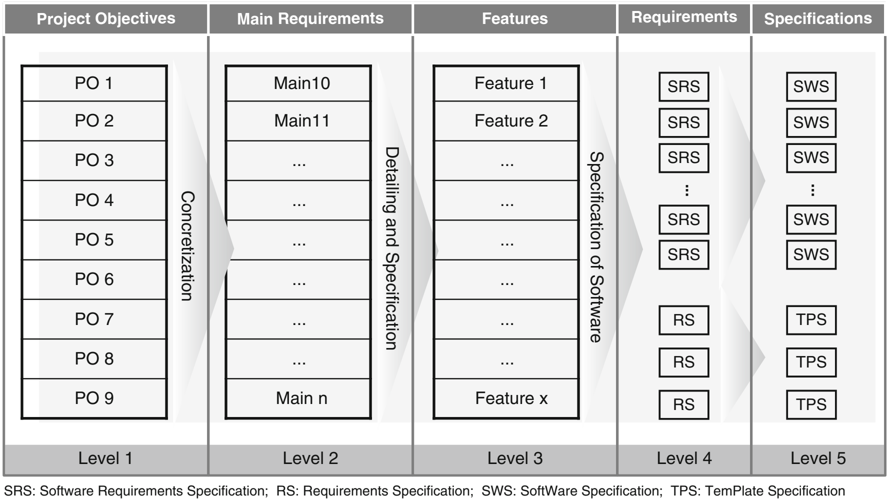


Fig. 3 Correlation between the AUTOSAR documents

These project objectives are detailed in the “AUTOSAR Main Requirements Specification” in high-level system requirements. The following example illustrates this:

The project objective “transferability of software” breaks down into the following main requirements:

- The software architecture of AUTOSAR shall be divided into functional layers.
- AUTOSAR shall provide an isolation layer from the hardware in order to allow for a hardware-independent design of as large as possible software parts.
- AUTOSAR shall permit the free distribution of application software across the vehicles onboard network.
- etc.

The main features and the core functions are derived from the main requirements. From these the “Software Requirements Specifications” (SRS) were derived. The detailing of these SRS results in the “Software Specifications” (SWS). Figure 3 shows their correlation. The “Software Specifications” form the basis for the implementation of the AUTOSAR standard into the software.

4 The Three Areas of the Standardization

The standardization of AUTOSAR is divided into the following three technical areas:

1. Software architecture
2. Development methodology
3. Application interfaces

4.1 Software Architecture

The main concept of the ECU's software architecture (see Fig. 4) lies in splitting up hardware-independent application software and hardware-oriented basic software (BSW) by the software abstraction layer – the Runtime Environment (RTE).

On the one hand this abstraction layer enables the development of OEM-specific and competitive software applications as driver assistance systems. On the other hand it simplifies the standardization of OEM-independent BSW. Furthermore, it is the precondition for the scalability of the ECU's software for different car lines and variants. It provides the possibility to distribute applications across several ECUs and to integrate software modules from different sources (see Sect. 6.4).

The BSW is further divided into the following layers, "services," "ECU abstraction," "microcontroller abstraction," which are detailed in Sect. 6.3. The Runtime Environment abstracts the application layer from the basic software and organizes the data and information traffic between them. This forms the basis for a component-oriented, hardware-independent software structure on the application level, with software modules as independent units.

For example, a function of a driver assistance system is implemented by software modules. These software modules form together the application. The individual software modules communicate directly only with the RTE. So the communication is designed clearly, regardless of whether it takes place within an ECU or exceeds ECU boundaries.

By this independency it is possible to develop software components without having knowledge about the used or planned hardware or rather to distribute the existing software modules between the ECUs.

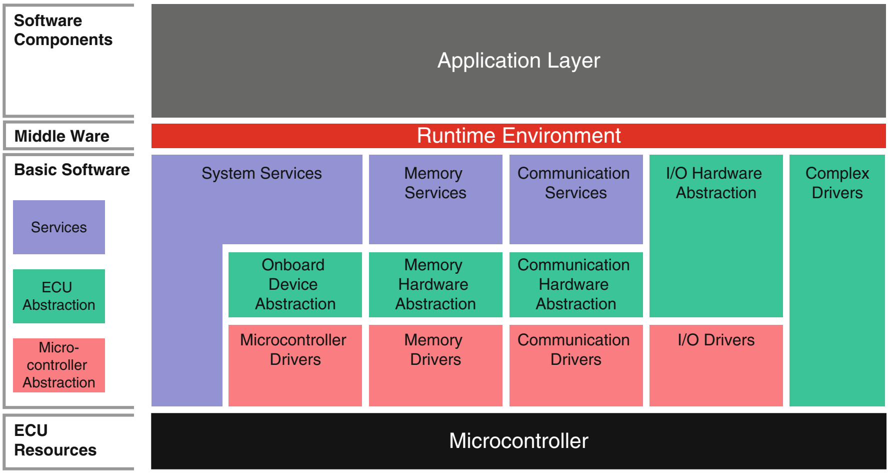


Fig. 4 AUTOSAR software architecture

4.2 Design Methodology

In addition to the software architecture, AUTOSAR also standardizes the methodology of automotive software development facilitating the cooperation of involved partners of modern series projects.

The AUTOSAR design methodology addresses particularly aspects that are necessary to integrate software modules in an ECU and various ECUs into the vehicle communication network with different bus systems. It defines the generic artifacts and the related activities and in particular the dependencies of activities. The design methodology is applied in the development of application software (see Sect. 6.1), the Runtime Environment (see Sect. 6.2), and in the system configuration (see Sect. 6.4).

For information that is produced or can be used in the AUTOSAR design methodology, AUTOSAR has defined a formal data exchange format (AUTOSAR scheme) with semantic constraints. This information is stored as a formal description in AUTOSAR XML (.arxml) files. Many tools use these descriptions for the configuration and the generation of RTE and AUTOSAR BSW. For example, the software component description provides a standardized component model for application software. Or the system description defines the relationship between the pure software layer on the system and the physical system architecture with cross-linked ECU instances. It describes the network topology, the communication for each channel, and the allocation of software modules on various ECUs.

The principle of AUTOSAR design methodology is illustrated in Fig. 5

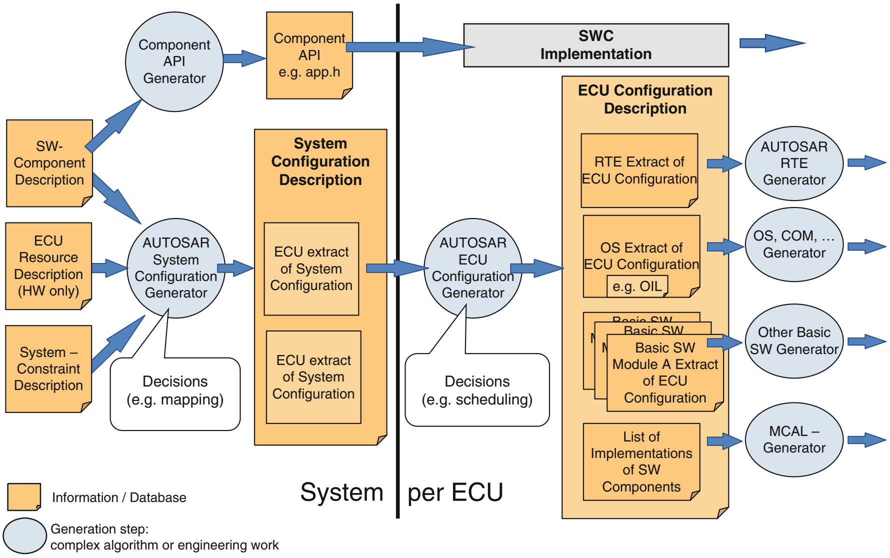


Fig. 5 Principle of AUTOSAR design methodology

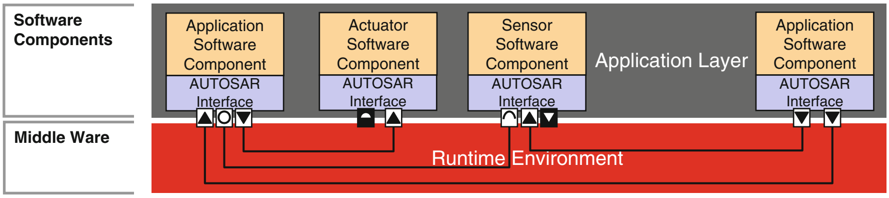


Fig. 6 AUTOSAR application interfaces – the shown symbols of interfaces are described in Fig. 9

Besides the basic ability to describe E/E systems in the automotive industry, there are many aspects that need to be supported by practical exchange formats, such as documentation, requirements traceability, and life cycles of various artifacts.

In addition, the integrated variant management allows OEMs and suppliers to express the basic AUTOSAR product lines and, whenever necessary, to exchange this information with their partners. The common understanding and aligned interpretation of these variants are a key element for the successful cooperation in joint projects.

4.3 Application Interfaces

The linkage of the application modules to the RTE is ensured by application interfaces as illustrated in Fig. 6. AUTOSAR standardizes on the one hand the basic interface mechanism with the syntax and on the other hand the semantics of the application interfaces in the vehicle domains body, interior and comfort, power train, chassis, and passenger and pedestrian protection. There, the focus is on the interface specifications for widely introduced applications to emphasize the reuse and exchange of software modules. Finally, the use of standardized application interfaces has essential importance for the reuse of applications.

Experts from all AUTOSAR partners standardize the interface specifications, e.g., regarding used data types, units, and scaling factors. They allow software designers and developers to use it in the case of extensions or reuse of software modules independently of any specific hardware or ECU.

Applications such as driver assistance systems include differentiating competitive features. Therefore, AUTOSAR does not standardize the internal functional process of an application, such as algorithms, but the information that will be exchanged between the applications. Typical examples of applications are driver assistance systems as described in Part I.

5 System Architecture: The Virtual Function Bus (VFB)

For developing the functional system architecture, AUTOSAR has introduced the concept of the virtual function bus – the VFB. The VFB allows describing the functional interaction between application modules throughout the system, i.e., through the whole vehicle. This description is independent of the actual ECU's architecture and the implemented network. In this way VFB abstracts the applications from the hardware. AUTOSAR describes individual applications as software components (SWCs). The VFB provides both the mechanisms for communication among themselves and the mechanisms for the use of the services of the basic software to the software components (see Fig. 7, upper part). The various mechanisms are represented by so-called ports (c.f. Sect. 6.1).

The functional system architecture is – during the further procedure – mapped on a physical architecture that means on an ECU and network topology. Here, the software components are allocated to the ECUs. In each ECU the function of the VFB is realized by the RTE and the underlying basic software (see Fig. 7, lower part). In order to avoid misunderstanding it should be explicitly noted: AUTOSAR has specified the VFB concept. This concept is implemented in various system architecture tools, which are available on the market.

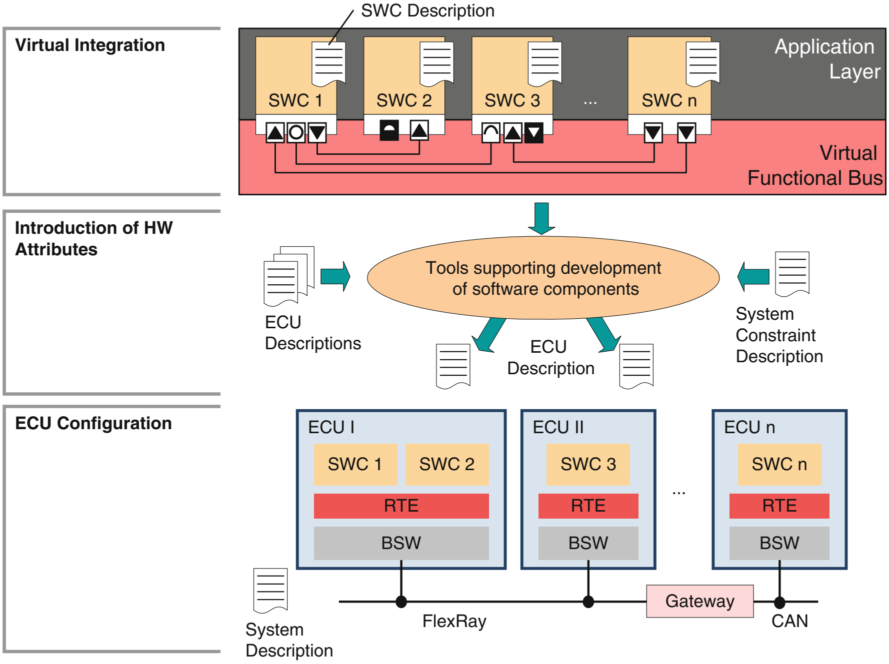


Fig. 7 Overview of AUTOSAR design methodology

6 Software Architecture

6.1 Application Software

The layer model of the AUTOSAR software architecture places application software in the form of software components in the application layer (see Fig. 6). Software components can be grouped to compositions that act externally as a software component again. Through this generic component concept, any nested hierarchies of software components can be realized as a system. The application software can be designed and developed independent from the hardware.

Software components communicate by means of ports that each represents a certain communication mechanism. The most important mechanisms in the communication between applications are "sender-receiver" for communication initiated by the sender of data as well as "client-server" for receiver-initiated communication. Beyond that, there are further ports for process control (external trigger events) or for access to certain parameters (calibration, operation modes, nonvolatile memory). Each port has an interface determining the data types to be communicated. AUTOSAR has defined a precise mapping of the ports in the programming language C. Figure 8 shows the communication path within an ECU and between applications in different ECUs.

A software component is formally described by a specific AUTOSAR description called "Software Component Template" (see Fig. 9). This contains, besides the description of the ports and the interfaces, also the so-called Internal Behavior. This term was established in the early years of AUTOSAR, but unfortunately often

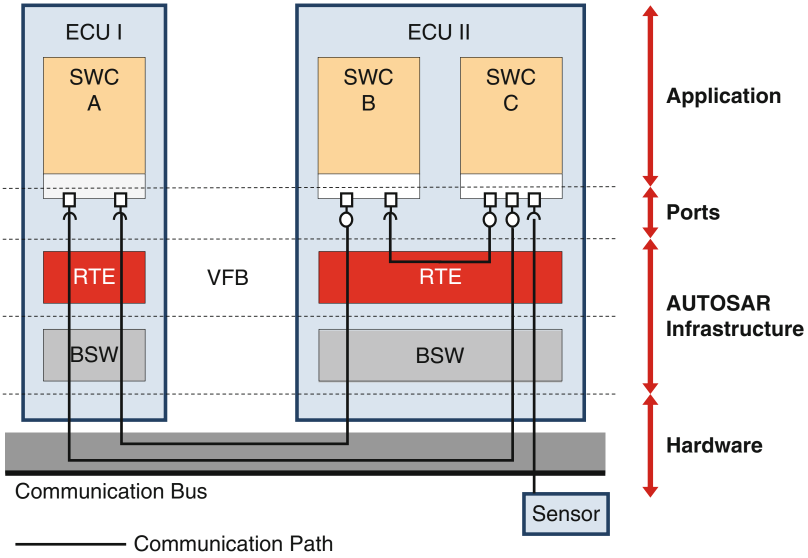


Fig. 8 Communication path within an ECU and between ECUs

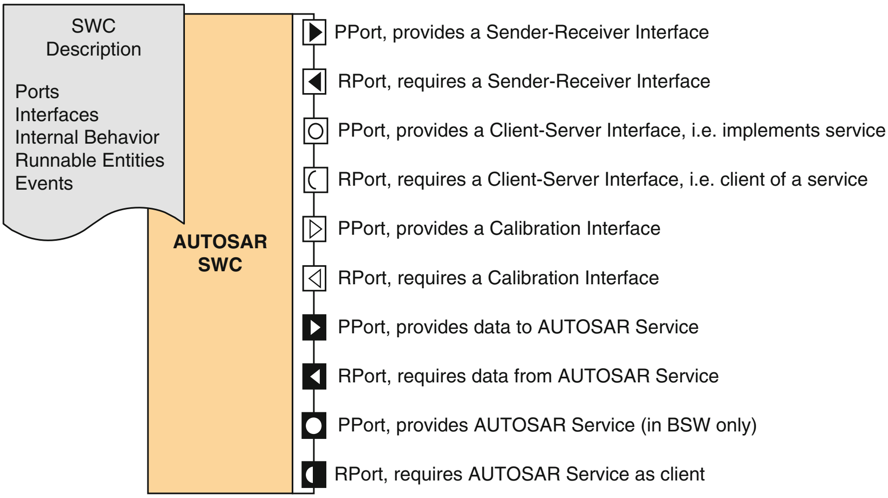


Fig. 9 AUTOSAR's description of components

causes misunderstandings. In the context of AUTOSAR, "Internal Behavior" describes a component with regard to time or event-related process control (events and scheduling). This includes the definition of "Runnable Entities," i.e., the smallest software entities schedulable on events or time by the underlying operation system. The algorithms to be implemented in the component explicitly do not belong to "Internal Behavior."

In practice there are several typical ways to fill in or edit the software component description. Many design tools for model-based development can generate a software component description out of a graphical model and allow editing the corresponding entries. Also, RTE generators (vgl. Sect. 6.2) allow generally the editing of the software component description.

For applications with specific hardware requirements, e.g., as software, which is dependent on certain sensors or actuators, AUTOSAR has provided so-called sensor/actuator software components in which such constraints can be noted in the software component description.

6.2 Runtime Environment RTE

The AUTOSAR Runtime Environment (RTE) abstracts the applications from any implementation details of the basic software and off the hardware of a control device. It represents the runtime implementation of the VFB (vgl. Sect. 5) on a specific ECU. The RTE provides the mechanisms for the communication between the applications and the mechanisms for the access to the services of the basic software. This also includes the provision of data buffering and queuing for communication.

The actual program code of the RTE depends on the applications, their communication, the used services of basic software, and the scheduling. In practice the code is created by the RTE generator according to the information of software component descriptions.

Strictly speaking the RTE is a "middleware" layer technology which enables the relocation of the components of the application layer through a decentralized network.

6.3 Basic Software (BSW)

The basic software provides the applications with all system services and functions through the RTE. Though the functions of the basic software are essential for the applications, these are typically not well noticed by the vehicle user. The basic software is divided further in layers with increasing dependence on the hardware: service layer, ECU abstraction layer, and microcontroller abstraction layer. In turn each of the layers contains individual modules which represent a precisely specified scope of functions. In total AUTOSAR basic software contains around 80 different modules, for which the standard has for each of them a requirement and software specification. Therein the functional behavior of the modules and its interfaces are defined in C, so that two different but standard-compliant implementations of one module are directly interchangeable. The parameterization of the functional behavior of a basic software module and its configuration uses the same formal description mechanism as application components. The configuration descriptions of the basic software modules of a control unit are summarized in the ECU configuration description.

6.3.1 AUTOSAR Services

The layer of services includes system services such as communication services, diagnostic protocols, storage services, management of ECU operating modes, and also the AUTOSAR operating system (OS) as an independent module. The AUTOSAR OS is based on the real-time system standard OSEK/VDX and is extended in some areas, but also restricted in others. It is configured and scaled statically and provides priority-based real-time behavior and treatment of interrupts. During the runtime various protection mechanisms for memory accesses or for the time behavior are available. The AUTOSAR OS is also suitable for small and lower performance microcontrollers, but also meanwhile supports multicore use and the use of multiple memory partitions for both code and data. The modules of services are apart from operating system independent of hardware. These system services are available to the applications via the RTE. Applications cannot directly access to underlying basic software modules. This is reserved for the services to access as part of their function to ECU or microcontroller resources. A service module and its underlying modules are also referred to as functional stacks, e.g., communication stack for FlexRay. Such stacks are sometimes implemented and integrated as one large software unit without the underlying module structure as defined by AUTOSAR. Although this undermines the principle of abstraction and reduces the flexibility, but because of possible higher efficiency and performance of an implementation, handling with functional stacks in AUTOSAR is widespread.

6.3.2 Hardware Abstraction

The layers under the services are used for hardware abstraction. Firstly, the ECU abstraction layer separates the ECU layout (i.e., how the peripheral modules are connected with the microcontroller) from the upper layer. Although this layer is ECU specific, it is independent of the microcontroller. The next level of abstraction is achieved by the microcontroller abstraction layer which includes microcontroller-specific drivers. These drivers are, for example, I/O drivers for digital inputs and outputs, or ADC drivers for converting analog signals to digital values. Thus, standardized hardware is supported directly by the AUTOSAR standard.

The layer of the complex drivers is used for the treatment of special cases, e.g., for controlling complex sensors or actuators with special real-time requirements or with specific electromechanical hardware requirements. Such modules are not standardized as AUTOSAR basic software modules, because here specific expertise and intellectual property of the automobile manufacturer or supplier would be required. However, the complex drivers and the standardized modules have to meet the requirements for interface mechanisms in the AUTOSAR basic software.

6.4 System Configuration

In the context of AUTOSAR, system refers to a composite of networked control units which may include all the ECUs of a vehicle. The system configuration follows the development of the functional system architecture on the VFB level (Fig. 10). When designing the system configuration, decisions considering the actual, physical architecture of the system are taken. These decisions are mainly related to the system topology, i.e., which control units are available and how they are connected. For each control unit, there is a description of the resources regarding processor architecture, processor capacity, memory, interfaces, and peripherals or signaling methods. The description of the network topology ranges from the bus system to the communication matrix of individual channels. Additionally, it encloses the determination of which application software component should run on which control unit. All this information is registered in the system description. In practice this is done either by means of system architecture design tools as well as for the VFB design – then we also speak of system generator – or by means of the configuration tools for the basic software modules.

The system configuration succeeds with the further configuration of the individual control units and finally with the software integration, which is independent for each ECU, i.e. it can run parallel if required. Thereto all relevant information for a certain control unit is copied into the ECU description out of the system

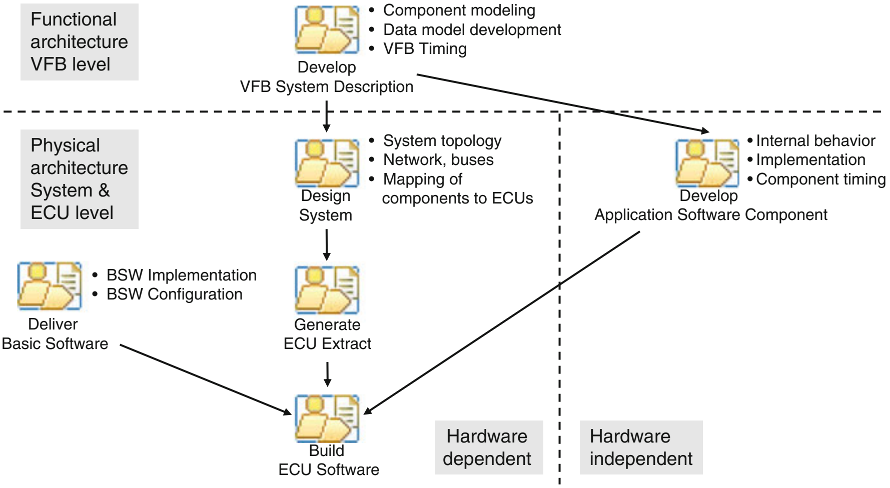


Fig. 10 From functional architecture to executable software

configuration. This is named as an ECU extract of the system description. The ECU description aggregates also the configuration descriptions of each basic software module. Many parameters of the basic software configuration result directly from the system description or the descriptions of software components. The remaining free parameters are set by using the basic software configuration tools. After the configuration step nearly for all basic software modules, the code that belongs to the configuration is produced by a generator – as it is within the RTE.

The implementation of the application software components – i.e., the creation of algorithms and coding – can be done completely parallel to the system configuration, as this step is independent of the hardware. Ultimately, the entire code of the basic software with the RTE code and the code of all application software components are integrated into the ECU software for each control unit.

7 Effect on the Characteristics of Driver Assistance Systems

The AUTOSAR standard was designed for the development of the software of the entire electrical/electronic system in a vehicle. The only exception are infotainment systems, as their software design is very close to that of consumer electronics and thus many of their basic mechanisms such as dynamic memory management are used.

Thus, driver assistance systems are in core focus of the AUTOSAR standard. Some of the specific mechanisms which the AUTOSAR standard provides for driver assistance systems will be discussed in the following.

7.1 Development of Distributed Real-Time Systems

Driver assistance systems are often characterized by complex algorithms, high safety requirements, and a wide variety of sensors. They are typically realized as a distributed real-time system, as, e.g., sensor data is being preprocessed in one ECU and the control algorithm is processed on a different ECU or because monitoring or diagnostic functions are implemented on a separate ECU. AUTOSAR simplifies the development of distributed real-time systems with respect to different aspects:

- The implementation of application software – the coding of source code – is decoupled from the design of functional and physical architecture as well as from the system configuration.
- The AUTOSAR design methodology allows a flexible distribution of application software on electronic control units. The system configuration and thus the distribution of the software components to electronic control units are static, i.e., the distribution cannot be changed during the runtime, but is relatively easy during the design process.
- For software components requirements for response and execution times can be considered. So the conditions for individual ECUs or for the entire system can be defined, as well as for other available resources. This information is available throughout the development process and will be considered by appropriate software development tools. For example, for an application of lane departure warning system, a maximum time delay of 200 ms can be defined and broken down to individual resource units.
- The standardized exchange formats for descriptions simplify the development of a distributed system by several development partners.

7.2 AUTOSAR Mechanisms for Functional Safety (ISO 26262)

The basic approach regarding functional safety for a vehicle is described in the standard ISO 26262 published on November 2011 (cf. ISO 2011). The ISO 26262 is a system development standard which defines the process and technical requirements for the development of electric/electronic systems in the vehicle. Starting from a description of the system and a derived hazard analysis and risk assessment (see ▶ Chap. 6, “Functional Safety of Driver Assistance Systems and ISO 26262”), respective safety mechanisms are defined. They are specified in a functional and a related technical safety concept during the development of the system. From the technical safety concept, requirements on the software of the system are derived in the context of systematic design. Part 6 of ISO 26262 contains the corresponding requirements for the development of safety-related software. Since the AUTOSAR standard defines a development methodology and software infrastructure, only partial aspects of a technical safety concept can be implemented by it. That is why AUTOSAR according to the nomenclature of ISO 26262 is also referred to as

"Safety Element out of Context" (SEooC) (see ISO 2011, part 10, Sect. 9). This means that the specific safety requirements in a development project must be in congruence with the assumptions about the technical safety concept made in the AUTOSAR standard.

During the development of AUTOSAR, typical technical safety concepts and the technical requirements of ISO 26262 were analyzed, and appropriate safety mechanisms have been derived and integrated into the AUTOSAR standard. Especially to be emphasized are:

- Memory partitioning
- End-to-end protection
- Program flow monitoring
- Defensive behavior

During the specification of the technical safety concept of a safety-related driver assistance system, the safety mechanisms provided by AUTOSAR need to be considered appropriately in order to avoid redundant efforts in the development and verification/ validation of an ECU.

7.2.1 Memory Partitioning

Memory partitioning creates protective boundaries in memory for one or more software components that reside an OS application. These software components are organized as logical partitions on application level, as shown in Fig. 11. They have limited write access to memory – including the main memory and nonvolatile memory – as well as memory-mapped hardware. In addition, software applications running in a memory partition are executed in the application mode of the CPU. This means that they have limited access to special control registers of the CPU and cannot execute special commands which are limited to the supervisor mode.

If a software component is running in an application partition and tries to write into an illegal memory area, then this is detected by a so-called hardware-based memory protection unit (MPU) and the respective write access is blocked. This leads to an interruption of the execution and the application partition, which has caused the error. The respective application partition will be ended in a controlled manner by the operating system and the RTE. If accordingly configured, all software components in that partition get started afterwards. The same mechanisms come into effect if a software component located in an application partition tries to change CPU registers which may only be changed in supervisor mode.

7.2.2 End-to-End Protection

The concept of an end-to-end safety protocol assumes that safety-related data exchange shall be safeguarded during the term against the effects of errors in the communication path between two software components that communicate through a physical bus. These are, e.g., random hardware errors such as corrupted registers in the sending or receiving network controller, interferences by electromagnetic

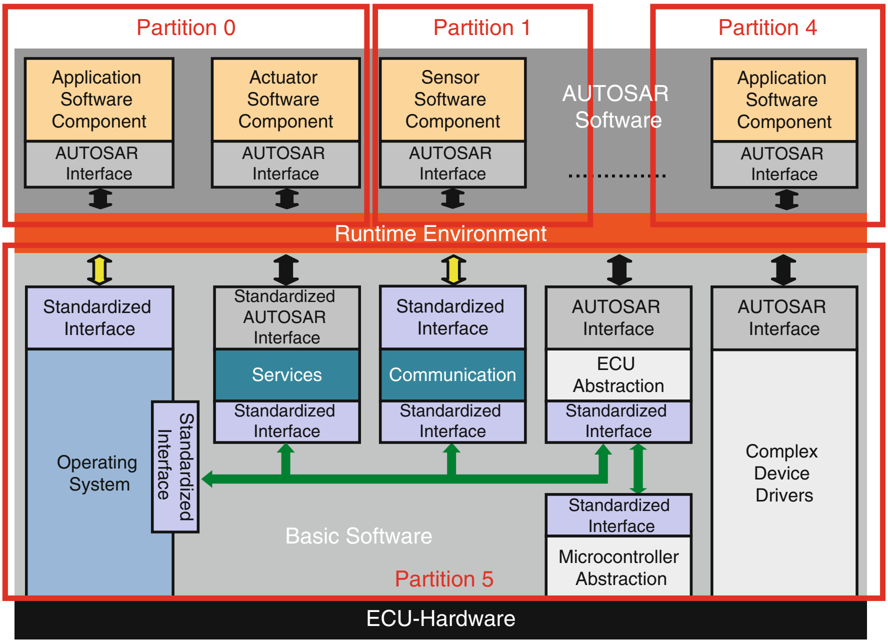


Fig. 11 AUTOSAR memory partitioning

waves, and systematic errors in the software implementing the communication of the virtual function bus, for example, in the RTE or in the network stack.

The end-to-end safety protocol can discover these errors in the communication path and handle them during runtime. The end-to-end library of AUTOSAR provides the corresponding mechanisms which fulfill requirements for safety-related communication up to the Automotive Safety Integrity Level D (ASIL). Their underlying mechanisms are checksums (CRC), message IDs, and alive counters.

The respective algorithms and state machines of assurance mechanisms are defined in the specification of the so-called end-to-end library of AUTOSAR. Figure 12 shows the operation of the communication between two software components using the end-to-end library.

7.2.3 Program Flow Monitoring

The monitoring of the program flow aims at discovering faults in the control procedures of the application software. A faulty program flow takes place when one or more program instructions are either processed in the wrong order, not in time, or not executed at all. Error in the control sequences can arise from systematic software errors or random and systematic hardware failures. They can lead to data corruption, program aborts, and ultimately to violation of safety goals.

The monitoring of the program flow addresses two types of monitoring: firstly, monitoring the behavior over time called “temporal program flow monitoring” and,

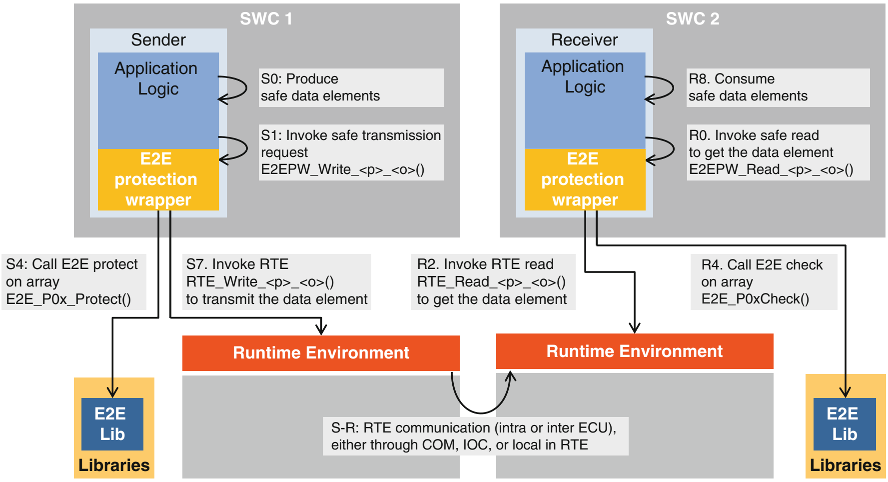


Fig. 12 Functioning of the end-to-end assurance

secondly, the monitoring of the logical order of execution of program sections called “logical program flow monitoring.”

The core function of monitoring the program execution is provided by the AUTOSAR basic software module Watchdog Manager. This monitors the so-called monitored software components, which are logical monitored units. Depending on the safety relevance of the software components and the safety requirements of the whole system, a monitored unit can be a group of software components or an executable software module within a component. The monitored software module calls at predetermined points the watchdog manager with a control code. From this it can determine whether the monitored software modules were executed in the correct order and within the valid time windows. In case of a violation of the predetermined timing constraints, specified predefined safety mechanisms are executed. This can lead up to an immediate reset of the ECU.

7.2.4 Defensive Behavior

Defensive behavior of software aims to prevent reproduction and propagation of errors in the software. It is a nonfunctional characteristic of the software and is generally achieved by appropriate programming guidelines and code patterns. Thus, the defensive behavior of the basic software of AUTOSAR is a characteristic that is not set by the standard itself, but that is ensured by the implementers of the standard. A typical measure is, for example, assurance against corruption of safety-related data of a software module. The corresponding data will be protected with a security code immediately before being transmitted from software component to another. When decoding the data in the receiving software component, the security code is used to restore the data again. If it is not identical to the previously stored one, the data has been modified in the meantime without authorization and the software component can trigger an appropriate error handling.

7.3 Virtualization in the Functional Validation

Distributed development of software components by suppliers and by automobile manufacturers, parallelizing the development of basic software and software components, and the following integration into an ECU by the car manufacturer or by another supplier have its price: The cost of testing to validate the functioning of all components increases rapidly. The functioning of the software components itself, their interaction with the RTE and basic software, the effectiveness of safety mechanisms as described in Sect. 7.2, and compliance requirements on the runtime of driver assistance systems (c.f. Sect. 7.1) are some examples of mechanisms and functions that need to be validated at the earliest point in time. The later the errors are found, the more work intensive their solutions and the higher the risk for the development project. So the question is: How can the function, e.g., of a lane departure warning system be validated early in the operating environment?

The virtual validation provides an effective possibility. On a host environment the AUTOSAR basic software is integrated. The software implementations from different manufacturers can be integrated including customer-specific adaptations. Using specific extensions the behavior of the basic software is similar to the behavior in a control unit with an integrated microprocessor.

An entirely virtual validation platform can either be implemented purely by software or with connected hardware components. In the latter case interface converters are attached to the validation host computer to ensure the connection to the vehicle interfaces. Furthermore, the basic software contains the hardware-compatible “microcontroller abstraction layer” with the related hardware drivers.

During the development of software components, on the one hand their functions can be tested and validated before integrating the software onto the ECU. On the other hand, the development tools can be used which are directly available on the host operating system, such as the debugger, making the workflow more efficient. The standardized basic software further allows the use of tools for the AUTOSAR design methodology, as described in Sect. 4.2.

This system ensures the functions of software components in an early development phase, taking into account existing software and ECUs of series vehicles and new technologies of, e.g., sensors. The hardware of the ECU under development does not need to be available by that point in time.

Certainly, this is possible with the car manufacturer-specific software. But only AUTOSAR enables consistent porting and thus reuse of test cases across manufacturers for other projects. Due to the standardized application interfaces (c.f. Sect. 4.3), the implementation of a virtual validation platform is modular and structured so that it is effectively realizable.

The virtual validation provides essential advantages when introducing methods of agile software development (c.f. Sims and Johnson 2012). Hereby, the software functions are developed in relatively short iterations, and in ideal case a tested and functioning software is delivered after each iteration. For complex systems being developed with various partners, this can only be achieved by a virtual validation with reasonable cost. The almost always used continuous integration in agile development, connected with continuous validation and software delivery between partners, can be realized much more efficiently using a virtual validation platform.

7.4 Mastery of Complexity and Development Time Reduction

From a technical point of view, we are facing increasingly complex systems by both the increasing number of, e.g., driver assistance systems and the increasing degree of interconnectivity. This evokes not only risks in the development but also in the validation of software systems. The described modular structure of AUTOSAR allows reusing the main parts of the software. Furthermore, the development mechanisms described in Sect. 7.1 allow the integration of a large number of software applications. When developing new applications, an ECU does not need to be redeveloped from scratch. New applications can be integrated into an existing architecture by generating and configuring of the new applications. This has a significant impact on quality and development time.

From an organizational point of view, a new complexity arises now. Development projects involve a variety of suppliers and address a large number of interfaces between the functional areas. The development of complex software architectures results in extensive multilateral discussions and negotiations to clarify the technical interfaces between the deliverables and the project-specific interfaces between suppliers.

The AUTOSAR design methodology helps to simplify this. Due to the standardized exchange formats based on XML and the standardized application interfaces, content and interaction are defined.

In addition, there is the possibility to establish fully integrated tool chains. The elements of the design methodology and the AUTOSAR specifications itself will be taken up by toolmakers. Currently, there are various development tools on the market which enable to create an integrated development environment (c.f. AUTOSAR 2013).

7.5 Transition to Flexible Collaborative and Shared Development

Along with the introduction of standardized basic software, there arises fundamental change in the business and working models. Car manufacturers no longer use their own ECU basic software, but microprocessor manufacturers, suppliers, and software suppliers develop and test the AUTOSAR basic software.

The application software communicates with the basic software via standardized interfaces. In this way software applications can be shared by multiple car manufacturers, for instance, software applications for windows or central locking – to mention two examples here. This means that the supplier of one of these applications does not develop customized products, but provide several customers with the same product. These software applications are therefore reusable and by its further spread more stable.

Specific competition-related applications will continue to be manufacturer specific. They can be used through use of the mentioned standardized interfaces for different models of one manufacturer.

The development of these competition-related applications will be done by the automobile manufacturers on their own, by suppliers, or jointly with suppliers. This is simplified by the already described formal descriptions of the software components, the control unit, and the overall system. They are available in a standardized exchange format and allow a tool-supported integration of the software.

The integration work shifts to the software level and thus to the automotive manufacturer, who has to design the architecture and to configure the control units as well as the entire vehicle system, in order to ensure an optimal reuse.

Considering not only driver assistance systems alone but also their current and future networking with each other, vehicle intra-connectivity, vehicle to vehicle, vehicle with traffic systems, or Internet-based services, it becomes obvious that innovation speed is a crucial factor for the successful implementation of such systems in vehicles. The speed of innovation is mainly determined by software reuse and joint development.

8 Summary

Since the beginning of its development 10 years ago in 2002, AUTOSAR has established itself in the automotive industry as the global standard for software infrastructure and system description with a continuous design process and standardized exchange formats to be used by all participating development partners. Starting with the introduction of AUTOSAR Release 4.x in 2009, all proprietary solutions in the non-differentiating software of electronic control units in the vehicle got systematically replaced. Since then the work of AUTOSAR transformed from backward into forward standardization with uniquely supporting new technologies like MultiCore and Ethernet for the automotive industry.

A determining reason for the success of AUTOSAR is the fundamental principle of partnership which has, as of October 2014, more than 180 partner companies:

Cooperate on standardization, compete on implementation.

The main result of the AUTOSAR partnership is therefore in the specification of the AUTOSAR standard and its implementation is subject to free competition.

Another fundamental change that was introduced with AUTOSAR is the paradigm shift for the user, away from implementing the software towards configuration and generation of software. This allows very fast implementation in software by appropriate tools from the system descriptions of AUTOSAR, and thus, an unrivalled degree of abstraction in the development of software for electronic control units can be achieved. This level of abstraction together with the independence of the specific hardware allows a new level of software reuse and enables a focus on the development of new and innovative customer features that currently mainly arise in the field of driver assistance systems.

References

AUTOSAR (2013) AUTOSAR – the worldwide automotive standard for E/E systems. ATZ extra: 10 years AUTOSAR edn, 2013–03, pp 5–12

AUTOSAR (2015) Homepage. http://www.autosar.org. Accessed 13 Mar 2015

ISO (2011) ISO 26262 Road vehicles – functional safety, parts 1–10, 1st edn. ISO, Geneva

Sims C, Johnson HL (2012) Scrum: a breathtakingly brief and agile introduction. Dymaxicon, Foster City

Von der Beek M (2006) Development of logical and technical architectures for automotive systems. Springer, Berlin/Heidelberg

Part II

Virtual Development and Test Environment for DAS

Virtual Integration in the Development Process of ADAS

8

Stephan Hakuli and Markus Krug

Contents

Abstract

In spite of the complexity and large variety of cars, electronic stability programs can still be tested in elaborate real test drives. However, due to the high system complexity, the complexity of the test cases, and the required scope of tests, this is not economically feasible for advanced driver assistance systems with environment perception. The repeatability of the tests, even under the exact same testing conditions and procedure, in practice does not exist due to various

---

S. Hakuli (✉)
Continental Engineering Services GmbH, Frankfurt am Main, Germany
e-mail: handbook.adas@hakuli.net

M. Krug
Department of Automotive Engineering, Hochschule München, München, Germany
e-mail: mkrug@hm.edu

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_8

potential and occasionally unknown or disregarded influences. Therefore, the test results are not reproducible – firstly, because functionally relevant characteristics can depend on the interaction between multiple road users, and secondly, because they can be subject to complex interactions between boundary conditions such as the blinding effect of a low sun and its reflection on a wet road at a certain angle. The features of currently used advanced driver assistance systems (ADAS) access environmental information that is collected by several different sensors and processed to obtain a representation of the environment. To serve their purpose, these functions utilize different actuators and components of the human-machine interface. This architectural distribution of assistance functions to different control units and vehicle components results in a strong interconnection that must be considered during testing and that drives up the costs of testing. This chapter will highlight the advantages resulting from virtual integration and describe its functionality and limitations.

1 Consistent Testing and Evaluation in Virtual Test Drives

The central idea of virtual test driving is the transfer of real test drives to the virtual world in the most realistic way possible. The idea has a primary aim to benefit from the characteristic advantages of simulation with respect to reproducibility, flexibility, and reduction of efforts and to have a possibility for testing and evaluating specifications and solutions derived from them at an early stage of the vehicle development process. The use of suitable simulation processes enables an efficient conception, development, and application of vehicles and vehicle components. The tests shorten and bridge the time to availability of real vehicle prototypes. As with real test drives and the reliability of real experimental results, the use of simulation techniques is an optimization task in which it is necessary to weigh the efforts for modeling, parameterization, and simulation against the efficiency gains from it.

The virtual test drive, like its real counterpart, consists of several components. The main role is played by a virtual vehicle prototype, whose components are integrated, depending on the progress in the development process, as models, software code, or hardware. Due to defined interfaces between the subcomponents of the virtual vehicle, which integration stage the other corresponding components are in is irrelevant for each component.

The virtual vehicle, including all its functions, is operated by a virtual driver, whose driving strategy is parameterized by means of a behavior model and who can perform both open-loop and closed-loop maneuvers and therefore is instructed as in real test drives with maneuver step descriptions that he performs in sequence depending on either time, position on the route, or a triggering event. As a consistent enhancement of the signal-based testing, this maneuver step-based execution of the test with a configurable driver behavior is a prerequisite for the transferability of the simulation results to the results of the real test drives that it is based on.

Table 1 Gradual transition from virtual to real world

	MiL	SiL	ECU HiL	System HiL	Chassis dynamometer	ViL	Test drive
Function code	V	R	R	R	R	R	R
Control unit	V	V	R	R	R	R	R
System	V	V	V	R	R	R	R
Vehicle	V	V	V	V	R	R	R
Driver	V	V	V	V	V/R	V/R	R
Driving dynamics	V	V	V	V	V	R	R
Driving experience	V	V	V	V	V	R	R
Road	V	V	V	V	V	R	R
Traffic/ environment	V	V	V	V	V	V	R

V virtual, *R* real

The virtual driver drives the virtual test vehicle on a virtual road in a virtual environment that maps real routes and their properties and in which he interacts with virtual road users.

While all elements of the test drive in the early concept phase are virtual, the development process subsequently passes through different integration stages, as illustrated in Table 1. Here, a gradual exchange of virtual components with the associated real test components takes places until the point at which a completely real test drive on a real road with real drivers and other real road users has entirely replaced the simulation.

However, the virtual test drive is not only the mapping of a test configuration to the virtual world but also includes the transfer of the evaluation and assessment methodology from the real test drive to the simulation. The continuous fulfillment of specifications at system level does not guarantee the desired behavior of the entire vehicle and, thus, also does not ensure the validity of a product with regard to its ability to meet the product goals, which is not checked until the release test. The goal must therefore be to test partial solutions against the associated specifications during the development process and in the same context be able to test the entire vehicle concept for its ability to meet the desired properties of the overall solution. Therefore, the actual added value of the virtual test drive is achieved by consistently transferring the maneuvers and the corresponding evaluation criteria from real vehicle testing to all earlier development phases along the V-model as described in Sect. 3. In the ideal case, design decisions can be verified in purely virtual tests for their ability to meet the target properties of the entire vehicle. In the subsequent course of development, this approach helps to avoid unnecessary disruptions in the testing and evaluation chain during the integration stages and makes sure that the test results can be compared across the different stages.

The use of virtual vehicle prototypes in virtual test drives, therefore, makes design decisions assessable in the overall vehicle context and contributes to the system and component specifications. This is described in Sect. 2 as being able to be reviewed for their suitability to achieve the target properties of the vehicle even before the development of the first real prototype. Section 4 takes a closer look at this aspect using an example of a function development along the V-model through all integration steps.

2 Efficient Collaboration Between Manufacturer and Supplier by Means of an Integration and Testing Platform

The virtual test drive, based on virtual prototype vehicles as well as catalogs of maneuvers for release testing that are mapped in the simulation together with the associated evaluation criteria, may also contribute to efficiency gains in the collaboration of vehicle manufacturers and system suppliers. The supplier benefits during development and application of a component from the presence of a virtual vehicle prototype, since he becomes more independent of real and usually rare vehicle prototypes for the assessment of the target objectives of the entire vehicle. Maneuver catalogs on the supplier side mainly focus on the verification and validation of contractually guaranteed system properties. By additionally applying maneuver catalogs as provided by the manufacturer, the supplier can examine his development results for their likely ability to meet the defined objectives in the overall vehicle context, even before this component is integrated into a real prototype and becomes verifiable in real-world test drives. For the manufacturer, exchanging maneuver catalogs with his system suppliers is likely to result in saved time and a reduced number of iteration steps and can be seen as an effective means of front loading in the development process.

Such collaboration requires the manufacturers and suppliers to use the same integration and test environment. Both vehicle and component models are exchanged, and release maneuver catalogs are implemented in the simulation with an integrated evaluation procedure according to defined evaluation criteria both at vehicle and component level. The manufacturer can provide the vehicle data sets in encrypted form as a “black box”, which can be used by the supplier in simulations but cannot be seen in detail or changed and can have an expiration date. The module that is developed by the supplier is thereby exempted from encryption and can be integrated and tested in a virtual vehicle. In return, the supplier can share component models and control software in encrypted form and, for example, via the FMI/FMU mechanism (FMI Development Group 2014; Schneider et al. 2014) with the manufacturer, regardless of the authoring tool that was used. This allows the vehicle manufacturer to plan with more mature vehicle components, which have been validated in the context of the entire vehicle, with a shorter delivery time and whose delivery is associated with fewer expensive and time-consuming iteration steps until they are finalized.

3 In-the-Loop Methods and Virtual Integration in the V-Model

Driver assistance functions are heavily based on software. Therefore, it makes sense to use the V-model (V-Model 2015), known from software engineering, or its further development, the “V-model XT” (V-Model XT 2015), as a development process for driver assistance functions. The V-model is basically a chronological development process. The process is not plotted linearly against the time axis but in the shape of the letter V instead. One therefore refers to a descending and an ascending branch. The descending branch contains the steps of the task analysis. The specifications for the components that are to be developed incrementally result from this analysis. It is essential that the overall product requirements (often called customer requirements) are first analyzed and then transferred to a logical architecture. After this, the development of a technical architecture follows that is subsequently decomposed into systems and components and specified. Parallel to each of these steps, test case specifications are developed to be later used to review the development. The last step of the descending branch at the same time marks the first step of the ascending branch. It includes the actual implementation, or more specifically the development of the specified components. The ascending branch includes all testing and integration steps of the single component, the entire system, as well as acceptance testing with the customer. It thus represents the integration and testing in the development process. Each step on the descending branch has a connection to a step in the ascending branch. The connection corresponds to the verification of the subsystem created in the process step with the associated specification. The test cases used in each step are those which have been developed during the specification phase of the descending branch. The last step contains the validation, which is the verification of conformity to all customer requirements and acceptance tests. Figure 1 shows the general development process according to the V-model.

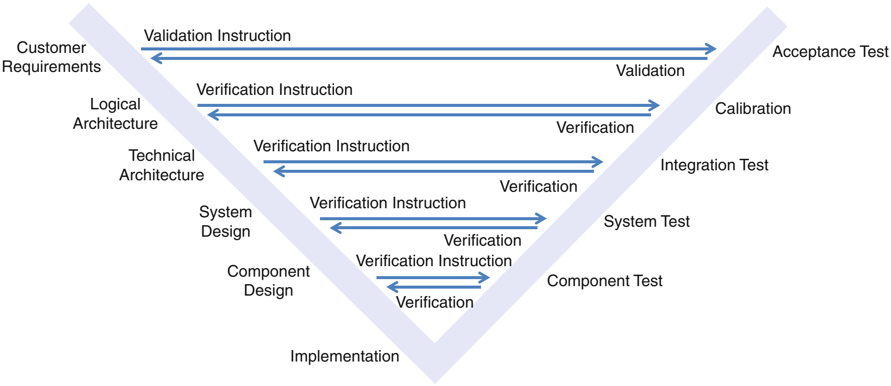


Fig. 1 Development process according to the V-model

The V-model is characterized by its good comprehensibility and the connection between product development and quality management. The connection is achieved by the use of the test cases from the specification phase during the integration phase. A significant disadvantage of the V-model is, however, that checking the solution against the corresponding specifications is not feasible before the corresponding step of integration is reached. This becomes most relevant for driver assistance functions when validating the overall functionality that was derived from the initial customer requirement. The associated validation is not possible until the last step of the development process. If the specification of the customer requirements was incomplete or even incorrect, all subsequent specifications and their implementation are affected. With the V-model, this cannot be formally determined until the last stage of development. In the automotive industry, there are typically 3 years of development and often development costs of several million euros in between. A required change is very likely to increase the development costs and prolong the development time.

To reduce this risk, appropriate methodological additions need to be made during the development process with the aim of being able to make a sufficiently strong statement about the quality of the development at an early stage (Winner 2013; Palm et al. 2013). This statement must be further substantiated along the development process to progressively reach a strong assessment. In order to gain sufficient flexibility, a continuous adaptation of the specifications should also be possible without significantly changing the previously created content. The methods used for this approach are mainly taken from the repertoire of the development of embedded mechatronic systems. Considered are SiL, MiL, and HiL methods (Schäuffele and Zurawka 2013).

The systematic approach of these methods corresponds to the coupling of the models or real components of each development step to a reproduction of its real environment with the aim of obtaining an assessable system. This replica is provided by a simulation environment in virtual form into which the real models or mechatronic systems are embedded. Since there are no real components until the implementation phase in the process, the simulation environment must be able to provide a virtual integration. Figure 2 shows the location of each method within the V-model.

The Model-in-the-Loop method (MiL) allows for a confirmation of the specification of customer requirements up until the step of the logical architecture. In this methodology, algorithms are created that functionally match the development objective. However, they do not yet have a reference to the hardware of the target system. These algorithms are usually created in the form of model-based software. In order to validate these models, they are integrated into a simulation environment and tested by means of virtual test driving. This means that all necessary components (environment, road, driving dynamics, powertrain, sensors, driver model, etc.) are provided as modules. The created model is added to the simulation environment to make the new function verifiable. The integration of models into a virtual prototype and, thus, into the context of the entire vehicle results in specific requirement specifications and, therefore, more specific test instructions, which can

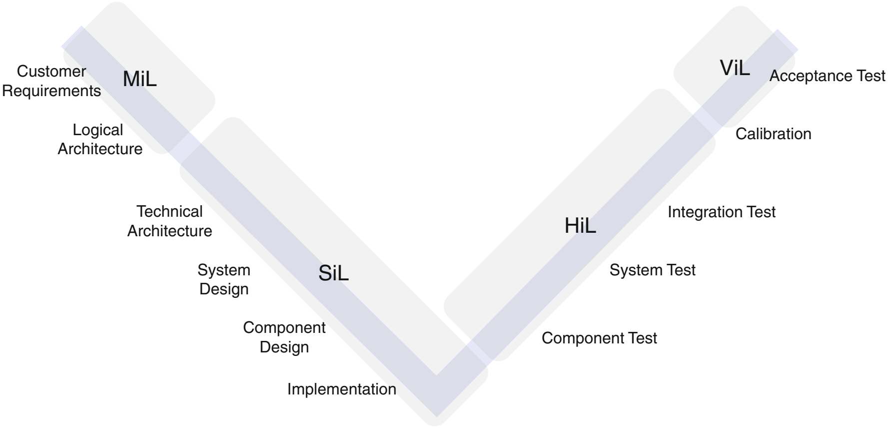


Fig. 2 In-the-loop methods in the V-model

avoid potential surprises during the validation process. Ideally, for this purpose the simulation environment is coupled with a driving simulator, in which test drivers can already test the newly developed driver assistance system. Depending on the level of detail of this method, a substantiated statement about customer acceptance is already possible. This step precedes the corresponding step in the classic V-model and, thus, allows for a significant reduction of the development risk.

The Software-in-the-Loop method (SiL) allows for an assurance up to the level of the individual components. This is achieved by transferring the previously created models into a simulation environment that is very similar to the technical characteristics of the target system in terms of computing power, real-time behavior, or resolution accuracy but is still hardware independent (Martinus et al. 2013). Therefore, the SiL method offers the possibility to check the specifications of the individual components of a system prior to its implementation and adjust them if necessary.

If the development process is complemented according to the V-model by the MiL and SiL methods in a powerful simulation environment, a virtual integration of the entire system is created as a result at the end of the descending branch. Thus, before starting the integration phase, a virtual prototype is available that enables a complete testing and verification of each component and its functionality, individually and with regard to its interfaces. With this virtual prototype, a virtual test drive as described in Sect. 1 is possible. This, in turn, results in the possibility to test the effects of the tolerances of each component on the customer function. Since this can be automated and often performed faster than real time, this virtual prototype is a very powerful tool to examine the individual specifications and the whole system. Moreover, with the appropriate configuration of the virtual prototype, the failure or misuse of single components and the influence on the overall system and its functionality can be tested. The possibility to do so does not make similar tests on the real component superfluous once they are available. However, it is much more flexible, faster, and cheaper. In addition, the findings can be integrated into the specification of the individual components.

The described procedure is not limited to the functional area of driver assistance systems. However, this domain is predestined for such an approach for the following reasons:

- High degree of interaction with the domains of human-machine interface, powertrain, vehicle dynamics
- High demands on functional safety and its assurance with increasing degree of automation
- High development risk due to new sensor concepts and algorithms
- High development risk due to little experience according customer acceptance

The third in-the-loop method is used to transfer the developed models from the SiL environment to the real components or be replaced by them respectively. The method is referred to as Hardware-in-the-Loop (HiL). In distributed systems, this stage is typically performed in several steps. First, the individual components are tested independently against their respective specifications. Here, a simulation environment is used that provides the interfaces of the components that are to be tested. Once all components are verified with this method, they are partially integrated using the same method to also verify their interaction. At the end of this stage, the entire system exists in real components and is tested against its specification up to the level of the logical architecture.

It is important that the test scenarios that were used during the creation of the virtual prototype can be applied again. On the one hand, this lowers the costs. On the other hand, the results of the real and the virtual components can be compared directly. In case of a discrepancy, this facilitates the identification of errors. If this makes a change to the virtual prototype model necessary, the impact can be assessed by means of a virtual test drive beforehand.

Vehicle-in-the-Loop (ViL) is a newer method for usefully complementing and enhancing the development of advanced driver assistance systems with the V-model. It addresses the need of many driver assistance functions for a complex test drive and a high standard of functional safety. This group of driver assistance functions will progress in importance and size. A major reason for this is the growing number of vehicle variants that offer driver assistance functions and which must remain safe even with the ever-increasing degree of automation and network integration. The ViL method allows the operation of the real test vehicle in a virtual environment. The coupling between the vehicle and the virtual environment can be done in two ways. One way is by creating an interface to the available environment sensors and, thus, replacing the real sensors. At this interface, the simulation environment is feeding simulated sensor signals, which correspond to the sensor response from a real environment. This offers an advantage whenever real sensors cannot be stimulated by artificial signals with reasonable effort. Otherwise, the real sensors can be maintained and stimulated artificially, as is feasible, for example, for ultrasonic sensors, which are exposed to artificially generated response signals from an ultrasound transducer (Pfeffer et al. 2013). In both variants, the real test vehicle responds to attributes and events of the virtual environment. This way, critical driving maneuvers with obstacles or objects on a collision course can be tested reliably and reproducibly. The created interface can also be used to generate the sensor signals as they would occur due to a changed position in a vehicle variant or due to different tolerances. This method therefore offers the possibility to test these variants or tolerances with a single test vehicle. In addition to the considerably more safe test operation, this allows efficient testing and application of advanced driver assistance systems. This results in a substantial economic gain with respect to the test drive when it comes to driver assistance systems.

Driver assistance functions are rarely developed from scratch. Typically, a functionality that uses various components already exists. On this basis, a new feature is added. In such cases, it is very helpful if the existing system is already available as a virtual prototype as described. Based on it, the specification of the new functionality can be efficiently integrated into the existing structure of the virtual prototype consisting of MiL and SiL components and be tested. The new real component can also use the existing HiL infrastructure. This approach acts as a very useful quality measure, because the changes to the already tested system are easy to understand and examine. The available test cases can be used together with the existing base system. In addition, it opens up an economic potential, since some development activities can be transferred with the existing infrastructure. The reuse of this infrastructure is also an investment protection.

For complex driver assistance functions, the associated development process according to the V-model cannot be regarded as a sole process. Driver assistance functions have strong interactions with functions from other domains of the vehicle. This interaction requires a domain-independent concept with respect to integration and testing. Today, this is typically accomplished by the fact that in the ascending branch of the V-model so-called synchronization points are agreed between developers from the different domains. These synchronization points represent the integration of all functions in the vehicle with a defined partial functionality. Typical partial functionalities are the availabilities of all system or customer functions or their application. The development processes of the individual domains of the vehicle may differ in terms of the procedure leading to a synchronization point. Also, the scope of implemented partial functionality can be quite different from domain to domain. In the future, this principle is to be expected also in the development process of the descending branch of the V-model. The described virtual integration in particular enables the availability of a virtual prototype at any time during the development process. Only the level of detail differs for different points in time. Therefore, coordinating the integration of functions is already useful in the descending branch of the V-model.

The great potential of the virtual development and integration of driver assistance functions will not be able to completely replace the real test drive. This is due to the fact that some test scenarios are first discovered in real test drives, since any arbitrary situations can occur. If relevant, these test scenarios can then be transferred into the virtual test drive as long as the mapping of the relevant events and mechanisms is feasible. Besides, the subjective assessment of driver assistance features is an aspect that cannot be fully transferred to the virtual test drive.

4 Virtual Integration in the Development Process

The following contains an example which is to show how the virtual integration is part of the development process according to the V-model. The aspect is divided into the specification and the integration phase. This corresponds to a division between the descending and ascending branch of the V-model.

4.1 Specification by Means of Virtual Integration

The following example of a customer requirement in the application area of parking/maneuvering illustrates the extension of the development process according to the V-model through virtual integration. First, a description is given for each step of the development process according to the V-model. The description is rather short and serves only for the understanding of the method. In addition, the particular activity for virtual integration and the resulting added value compared to the classic process of the V-model is given in three separate sections for each step.

4.1.1 Customer Requirement

V-Model The customer requirement is formulated as prevention of damage to the sides of the vehicle caused by collision with a stationary object during a parking maneuver. The maximum driving speed is 10 kph. Beyond this limit the function is inactive. Typical maneuvers are described verbally and defined as test cases for the future development.

Virtual Integration In the simulation environment, the maneuvers, previously defined as test cases, are configured as virtual test drives to make the customer requirement more transparent. The sensors of the virtual test vehicle have ideal behavior with respect to environment perception. The simulation of maneuvers gives developers an important indication of the fulfillment of customer requirements and of noteworthy details of the function. If a driving simulator is available, the customer has the opportunity to experience his initial verbally formulated request and to further specify it if necessary. If there is no driving simulator available, the normal, ideally photorealistic animation of the simulation gives the customer a first impression and the possibility to adapt his requirements based on this.

Result The virtual integration in this step allows a much more transparent discussion about the formulation of the customer requirements. The developers thus get a first impression of the range of functions that require special attention. This also allows a first estimation on the specification of the required subcomponents and their feasibility. Most important in this step is that the customer function can at least be visualized in order for the customer and the developer to have a common basis for discussion. This considerably reduces the risk of misunderstandings and resulting mistakes during the development process.

4.1.2 Logical Architecture

V-Model The logical architecture may be formulated as follows: Detection of stationary objects using the ultrasonic sensors attached to the front and rear bumpers. After the detection, the objects are further localized when exiting the visual range of the outermost sensors by means of object tracking, based on the movement of the vehicle. If the localization, based on the current speed and the current steering angle, yields an approach of an object to the side of the vehicle, a warning is issued. This warning is given by an acoustic signal. Corresponding test cases are formulated to stimulate this event.

Virtual Integration In the virtual integration of this step, the previously simulated scenarios are specified and updated. This refers, for example, to the adjustment of the simulated sensors according to the characteristics of an ultrasonic sensor and the integration of an algorithm for object tracking, including the output of a warning within the simulation. The test cases that were developed for this step in the V-model are simulated.

Result At the end of this development step, the use of virtual integration results in a tested logical architecture of the customer requirement. This allows for a statement whether the customer requirement can be realized in terms of data flow and functional logic within the existing capabilities.

4.1.3 Technical Architecture

V-Model The technical architecture basically consists of the functional parts *Perception*, *Processing,* and *Output*. The scope of the perception refers to the detection of objects. For this purpose, the existing ultrasonic sensors and their interface to the vehicle bus are to be used. For processing, an additional control unit is integrated into the existing control unit network. The scope of information output is realized via a message on the same vehicle bus on which the sensor information is received. The processing of the message occurs as an acoustic warning over the infotainment system of the vehicle. The interfaces that are used for this purpose and not changed in the following development steps are described in their technical details. In addition, test cases are formulated in this development step, which are primarily used to check the interfaces between the defined functional parts.

Virtual Integration Based on the simulation models created earlier, further detailing of the specifications results from the partition of the previously modeled customer functions into the described functional parts. The interfaces between the functional parts are adapted to the actual technical conditions with regard to timing behavior and available bandwidth, and the test cases that have been formulated for this step are conducted again in a simulation.

Result As a result of this step, the virtual integration made the assessment of the impacts on the customer function possible, which result from their integration into an existing control unit network.

4.1.4 System Design

V-Model In this example, the system design focuses on the definition of the software architecture for implementing the customer function. Here, the required functionality is separated into different tasks, and their interfaces are defined. This is followed by assigning the tasks to the involved control units. The description of the task is formulated as a black-box description. The necessary sensors or actuators are specified in a similar form. The test cases defined in this step mainly relate to a testing of the interfaces of the individual components. The interfaces can either be integrated into the control unit via a vehicle bus or via a direct hardware connection.

Virtual Integration In this step's virtual integration, the overall functionality that is already included into the simulation is divided into subfunctions. The subfunctions correspond to the tasks defined in the system design or to the model refinement of sensors and actuators from an ideal to a real behavior. This is done by respective modeling within the simulation environment. If an automatic code generation is planned for the tasks, it will depend on the standards arising from the used program. The previously formulated tests of the interfaces are performed in a simulation.

Result The system design is verified at the level of the component interface using virtual integration. This is a significant advantage over the traditional approach of the V-model, which would not allow this verification in this early development phase. A later change to the component interface comes with a significant change in its design. This change may also lead to changes in the system design.

4.1.5 Design of Components

V-Model In this step, the black-box description of the system design can be converted into a detailed specification of the component. This specification includes the internal flow of data and controls for each task. Eventually, there is a white-box description for each task. The test cases defined for this focus on the test of the algorithms for each respective task.

Virtual Integration If necessary, an adaptation of the functionality previously created on system level is performed in the virtual integration. Often, no significant adjustment is necessary as the component design is the result of the previous stages of development, all of which were already implemented in the virtual integration. The defined test cases are also conducted in a simulation.

Result Thanks to the virtual integration, tested virtual components are available as a result of the function specification. This results in an advantage for the following implementation step as there is already an adequate certainty on the correctness of the specification that is implemented.

Interim Conclusion The customer function used in the example was developed by the author using the described virtual integration. An integration and testing platform (IPG Automotive GmbH) was used, as was an authoring tool (MathWorks 2015) for the actual function development. In fact, numerous smaller and larger specification errors were detected and fixed at an early stage. The errors were mainly related to missing specifications or incorrect assumptions. For example, the geometric extent of the tracked objects, the necessary temporal behavior between object recognition and object tracking, or the behavior in the case of multiple simultaneous warnings were either initially undefined or incorrectly estimated. These errors may be considered as quite typical and can only be identified during the last two steps of the traditional development process according to the V-model. Therefore, the benefits of virtual integration could clearly be shown in a real project.

4.1.6 Implementation

V-Model For each component the implementation is carried out.

Virtual Integration Since all components are already present, there is no dedicated activity necessary for the virtual integration of this development step. In the case that automatic code generation is used, it is applied at this point.

Result If automatic code generation is used, the previously created virtual components can be used directly. This results in a significant advantage in terms of the quality and efficiency of this process step.

4.2 Integration Using Virtual Integration

The following is a description of how the considered customer function is progressively integrated based on the V-model. The integration benefits from the previous work with a simplification of the process at higher quality.

4.2.1 Component Test

V-Model With the help of the HiL method in a white-box approach, each individual component is verified with regard to the behavior according to its specification.

Virtual Integration The same test cases as in the component design step are used. For this, it is necessary that the simulation environment can communicate with the corresponding I/O measurement systems. In contrast to the testing of the component design, the interface to the component in this step is stimulated with real signals or read out respectively. This approach leads to results of the component tests that are directly comparable between the descending and the ascending branch of the V-model.

Result The reusability of test cases and the comparability of test results between the virtual and the real implementation of each component enable efficient examination and evaluation of the reasons for the observed deviations.

4.2.2 System Test

V-Model The number of jointly tested components is gradually increased until all components of the tested system are integrated. The HiL method and the test cases from the system design are also applied.

Virtual Integration Similar to the previous step, the simulation environment can be used to carry out the test cases. Due to the modular overall structure, resulting from the V-model, the number of real components can be gradually increased.

Result The modular overall structure and its mapping as a virtual integration allow a specific and reproducible system test. It is also possible to carry out variations in the order of the system test.

4.2.3 Integration Test

V-Model During the integration test, the new customer functionality is combined with the overall system for the first time. The simulated scope is reduced under certain circumstances to the driver and the environment. For the customer function that was used in the example, the integration is carried out with the infotainment system, and the test cases that were used in the technical architecture process step are applied. The HiL method is applied here too. If there is a high interaction with several functional domains in the vehicle, it may also be useful to already introduce the Vehicle-in-the-Loop method in this step.

Virtual Integration For the simulation environment, in which the virtual integration has been performed, this step is only a further reduction of the available simulation. The reduction primarily affects the functional areas of the vehicle that interact with the new customer function. The applied in-the-loop method is rather insignificant for the virtual integration, since the virtual integration is not tied to a test or a test vehicle. This is also true for the applied test cases, which do not differ between the different in-the-loop methods.

Result The customer function is verified for the entire vehicle. If the ViL method is used, the function would have already been experienced. The virtual integration allows for a specific and progressive integration test.

4.2.4 Application

V-Model The customer function is applied to the entire real system. In the example that was used, this refers, among other things, to the warning distance between the object and one's own vehicle. Ideally, the ViL method is used for this step.

Virtual Integration In connection with the ViL method, the real vehicle can be integrated into a virtual environment in this process step. This allows applying test variants, resulting from the vehicle and the test cases, much more efficiently.

Result The use of virtual integration in combination with the ViL method allows an efficient application of the customer function. The efficiency and reproducibility of the necessary test cases can, therefore, be increased considerably.

4.2.5 Acceptance Test

V-Model In the final development step of the V-model, the customer acceptance for the new function is tested. This is ideally done with a real vehicle in a real environment. The ViL method can still be installed in order to present the customer with options or alternatives that would otherwise mean a physical change to the target vehicle.

Virtual Integration The virtual integration plays only a minor role in this step and might as well be completely omitted.

Result Since the test cases used in this stage of the development have already been tested at the beginning of the development with the help of virtual integration, the biggest disadvantage of the V-model, namely, the validation of the customer requirement at a late point in time, is largely eliminated.

5 Limits of the Virtual Integration

The described virtual integration is not yet used in every development project, despite the proven benefits. The main reasons for this are the high requirements to the simulation environment regarding the necessary level of details with reliable real-time capability at the same time. This mainly concerns the simulation of the environment sensors and the actual environment.

5.1 Simulation of Environment Sensors

A prerequisite for the reasonable use of virtual integration is the valid modeling of the environment and the environment sensors within the simulation. Oversimplifications violate the validity criteria, which leads to the results from the descending branch not being able to be transferred to the ascending branch of the V-model. This corresponds to the optimization problem introduced in Sect. 1. If the mapping of complex physical effects, which are relevant to the function, is too costly or not possible in real time, it can be an exclusion criterion for the application of virtual integration. Approaches for improvement are topics of ongoing discussions (Schick and Schmidt 2012; Roth et al. 2012). However, to this day, no approach has proven to be completely purposeful regarding the validity of the model environment, because either the level of detail or the necessary computation time does not meet the requirements.

5.2 Simulation of Environment

Parallel to the requirement to appropriately simulate the environment sensors, there is also the need to realistically represent the environment within the simulation. This is necessary to be able to implement the interaction of sensors and their observed environment during the virtual integration. Simulating the environment is also of interest for other industry sectors. For this reason, many ideas and initiatives exist that are dedicated to the goal of creating a uniform and adequate specification (OpenDrive 2014; Infrastructure for Spatial Information in Europe 2014). A uniform standard or a de facto standard has not yet emerged, as the current activities still lack either the necessary completeness or the acceptance of a broad range of users.

The complexity of the simulated environment is driven by the number of features contained therein and the demands on the quality and level of detail. Static and dynamic objects in a simulated environment can occur in almost any combination and have any interaction. The resulting amount of scenarios is as endless as the number of everyday situations on the road. However, there are limited resources for the development of a customer function and for the simulation used for this purpose, which requires a reduction to a finite number of scenarios. Therefore, it will be necessary in certain circumstances to define a catalog of scenarios that are relevant for advanced driver assistance systems with a standardized representation of the environment that meets the requirements of the simulation. The scenarios for such a catalog must be chosen so that as many similar scenarios as possible are covered by them. The reduction to a finite number of scenarios at first appears to be a very strong restriction and questionable in terms of functional safety. However, other industry areas and other automotive domains have shown that this approach can lead to an increase in efficiency.

6 Conclusion

Virtual integration is in general not a completely new process in the development of functions in vehicles. It uses established process models and methods and enhances them by using the metaphor of the virtual test driving as described in Sect. 1. Therefore, virtual integration provides a tool for the development of complex, safety-critical, and highly cross-linked functionalities of the vehicle. Advanced driver assistance systems most often have these properties and, thus, benefit greatly from virtual integration.

In order to perform a virtual integration, a powerful and flexible simulation environment is necessary. The necessary properties of this simulation environment go well beyond the simulation of physical behavior and also include the connections to different real components. Therefore, it makes more sense, as explained in Sect. 1, to speak of an integration environment or an integration platform for virtual test driving. The simulation of the physical behavior is a task of the integration platform, while it also includes the type of testing and the possibilities of an efficient collaboration of vehicle manufacturers and system suppliers (Sect. 2).

Limits of the virtual integration are mainly reached in the simulation of environment sensors and the environment. The current level of detail is not yet sufficient to represent the real world and the required sensors for all applications, mainly due to the limitations of available computing power. Despite these limitations, using virtual integration and the in-the-loop methods described in Sect. 3, advanced driver assistance systems can be developed much more efficiently and with lower risks, as it was shown in the example of the developed function in Sect. 4.

A complete virtualization of the function development of driver assistance systems should, despite all the advantages of virtual integration based on virtual test driving, be expected neither in the near nor in the distant future. The permanent increase of test cases and necessary test depth, which are classified as relevant for the real test drive, and the subjective impression of test persons are only two of the reasons to keep real test driving as an integral part of the process in addition to virtual test driving.

References

FMI Development Group: FMI – the functional mock-up interface. https://www.fmi-standard.org. Aug 2014
Infrastructure for Spatial Information in Europe. http://inspire.ec.europa.eu. Aug 2014
IPG Automotive GmbH: CarMaker. http://www.ipg.de. Apr 2015
Martinus M, Deicke M, Folie M (2013) Virtual test driving – hardware independent integration of series software. ATZ Elektronik 8:16–21
MathWorks: MATLAB. http://www.mathworks.de. Apr 2015

OpenDrive. http://www.opendrive.org/. July 2014

Palm H, Holzmann J, Schneider S-A, Koegeler H-M (2013) The future of car design – systems engineering based optimisation. ATZ Automob Z 115:42–47

Pfeffer R et al (2013) Evaluation of a new method for reproducible real-time testing of ADAS ECUs, using the in-vehicle-simulation approach. In: IAVSD symposium, Qingdao

Roth E et al (2012) ADAS testing using OptiX. NVIDIA GTC, San Jose

Schäuffele J, Zurawka T (2013) Automotive software engineering, 5th edn. Springer, Wiesbaden

Schick B, Schmidt S (2012) Evaluation of video-based driver assistance systems with sensor data fusion by using virtual test driving. FISITA World Automotive Congress, Beijing

Schneider S-A, Frimberger J, Folie M (2014) Reduced validation effort for dynamic light functions. ATZ Elektronik 9:16–20

V-Model XT. http://v-modell.iabg.de/v-modell-xt-html-english/index.html. Apr 2015

V-Model. https://en.wikipedia.org/wiki/V-Model. Apr 2015

Winner H (2013) Challenges of automotive systems engineering for industry and academia. In: Maurer M, Winner H (eds) Automotive systems engineering. Springer, Heidelberg

Dynamic Driving Simulators

9

Hans-Peter Schöner and Bernhard Morys

Contents

H.-P. Schöner (✉) • B. Morys
RD/FFS, 059-X820, Daimler AG, Sindelfingen, Germany
e-mail: hans-peter.schoener@daimler.com; bernhard.morys@daimler.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_9

Abstract
Depending on the application, different concepts for driving simulators have been realized. A quite common dynamic driving simulator concept for professional applications, i.e., a motion platform consisting of a hexapod based on a linear rail, is explained in detail. Using Daimler's dynamic simulator as an example, the essential technological components of driving simulators are explained and the potentials and limitations due to the sensitivity of the human vestibular organ are discussed. Reasons for simulator sickness and how to avoid it complete the part on simulator design.

A second focus is placed on the design of simulator experiments with test persons. A clear goal of the simulation experiment, a good choice of technical and psychological test design, and knowledge about the behavior of test persons help set up effective driving simulator tests. Driver distraction is an essential feature to simulate the complete scope of real accident situations and to assess the behavior of a representative set of test persons.

The factors affecting the validity of simulator experiments compared to real-world experiments are discussed, and some findings on the opportunities and limitations of simulator experiments are presented.

1 General Overview of Driving Simulators

1.1 Applications of Driving Simulators

Driving simulators are used widely in the automotive industry and in automotive research institutions for many diverse applications, in particular for the following usages (in the order of rising requirements with respect to realism of motion rendering):

- Functional vehicle demonstrations, advertising new vehicle functions (demos at auto shows)
- Investigations into cabin, display, and control concepts (assessing reachability, clarity, comprehensibility, etc.)
- Training for drivers (of emergency vehicles, for fuel-saving driving style, race tracks, etc.)
- Accident research (accident reconstructions, driver behavior analyses, etc.)
- Study of driver performance and development of driver models (with respect to tiredness, attention, responsiveness, etc.) as a basis for off-line simulations
- Testing and verification of driver assistance systems (effectiveness, controllability, statistical analyses, etc.)
- Development of chassis systems and vehicle dynamics control systems (analysis of variants, parameter tuning, etc.).

All these applications focus on the interaction of human drivers with a vehicle's technical systems, especially in challenging driving situations (with respect to other traffic participants, obstacles, dangers, etc.) and under variable and adverse driving conditions (road condition, weather, lighting, etc.). Depending on the application, there are numerous technical realizations of driving simulators, from a static flat screen with a simple steering wheel and pedals to large dynamic simulators with perfected immersion technologies to reflect a virtual world that involve motion systems as well as auditory, haptic, and visual environment simulation.

With respect to testing and validation of driver assistance systems, which are in focus here, the main advantages of driving simulators are:

- The precise configuration of situations and the high reproducibility of traffic conditions
- A safe representation of critical situations
- Simple and fast variation of vehicle and environment parameters

Besides simulator experiments, various traffic situations with *real* vehicles need to be performed on test tracks and public roads to address effects that cannot be modeled adequately in a simulator. Together with such real-world tests, driving simulators have become an indispensable tool for the efficient and comprehensive testing of driver assistance systems.

1.2 Concepts for Dynamic Driving Simulators

An overview of the historical development of driving simulators was published in Slob (2008). Even before 1980, Volkswagen realized the first automotive driving simulator with three degrees of freedom for yaw, roll, and pitch motion. The VTI (Swedish National Road and Transport Research Institute) in Linköping (Nordmark et al. 1985) built a system with a motion system that was also limited to three degrees of freedom, although that used roll, pitch, and lateral motion (see Fig. 1a) supplemented with vibration actuators for roll, pitch, longitudinal, and vertical motion. In 1985, Daimler-Benz in Berlin (Breuer and Käding 2006) inaugurated a system based on flight simulator concepts, which was equipped with a hydraulic hexapod (Stewart platform with six degrees of freedom) and was the world's largest motion space available at that time (see Fig. 1b). Today, nearly all the main car manufacturers and several large research institutes possess their own dynamic driving simulator. Depending on the application focus and the budget, different system concepts were selected, although a hexapod supplemented by one or two linear rails is the most frequent design.

If a driving simulator is used as a development tool for driving dynamics investigations, the exact assessment of the lateral dynamic characteristics of the car is of greatest importance. With this in mind, during a revision of the Daimler driving simulator in Berlin in 1993, a rail with 6 m length was supplemented, which allowed precise replication of the vehicle lateral dynamics while performing a lane change maneuver (Käding and Hoffmeyer 1995). For investigations with general test persons, the avoidance of kinetosis (motion sickness) is important. Accurate

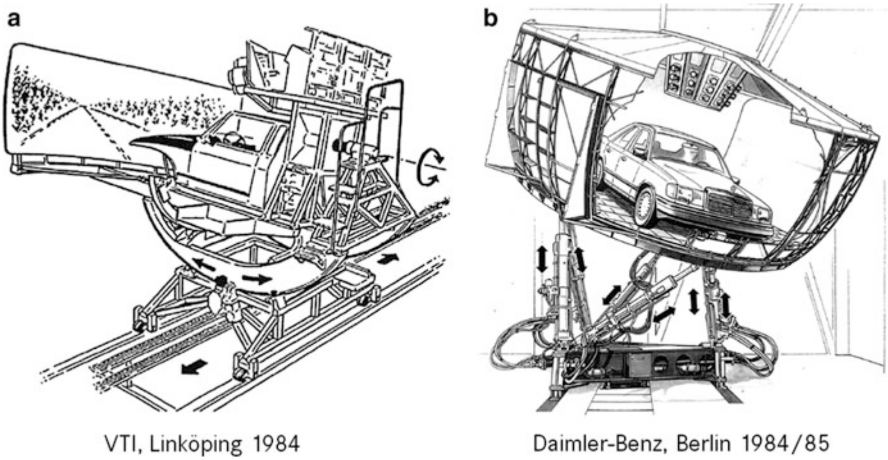


Fig. 1 First dynamic driving simulator concepts: (**a**) VTI in Linköping and (**b**) Daimler-Benz in Berlin

coordination of visual and haptic motion cues is crucial for this. New low-friction hexapod actuators, a digital control system, and an increased field of view of the projection system were integrated into the Daimler simulator in 2004. All these hardware components and a precise coordination of the vision and the motion system contributed to the reduction of kinetosis to under 2 % of the test persons (Käding and Zeeb 2010).

One of the largest challenges for designing the vision and motion system of driving simulators is the realistic representation of turning maneuvers at intersections, which arise generally when driving in city scenarios. For an accurate motion perception, a large motion space must first be available, which measures approximately the same size as the real-world maneuver. For this purpose, in 2006 Toyota built the largest driving simulator (see Fig. 2) with a motion space of 20 × 35 m (hexapod on X-Y slides, nearly identical to the National Advanced Driving Simulator (NADS) at the University of Iowa from 2000; Murano et al. 2009). Second, the picture system should be able to show a laterally moving scene during fast yawing motion without steps and delays to cause a consistent motion perception in the test persons.

The necessary mechanical size and weight for the correct representation of turning maneuvers poses limitations to the system dynamics; thus such a large system is less suitable for investigating fast-driving dynamics maneuvers. High cost and technical limitations of such large mechanical systems are not acceptable for many users, so there are several new solutions under investigation to produce a good motion perception with an alternative (cheaper) motion system. The following completely different designs should be mentioned here as examples (see Fig. 3):

Fig. 2 Toyota’s driving simulator in Higashi-Fuji (Murano et al. 2009)

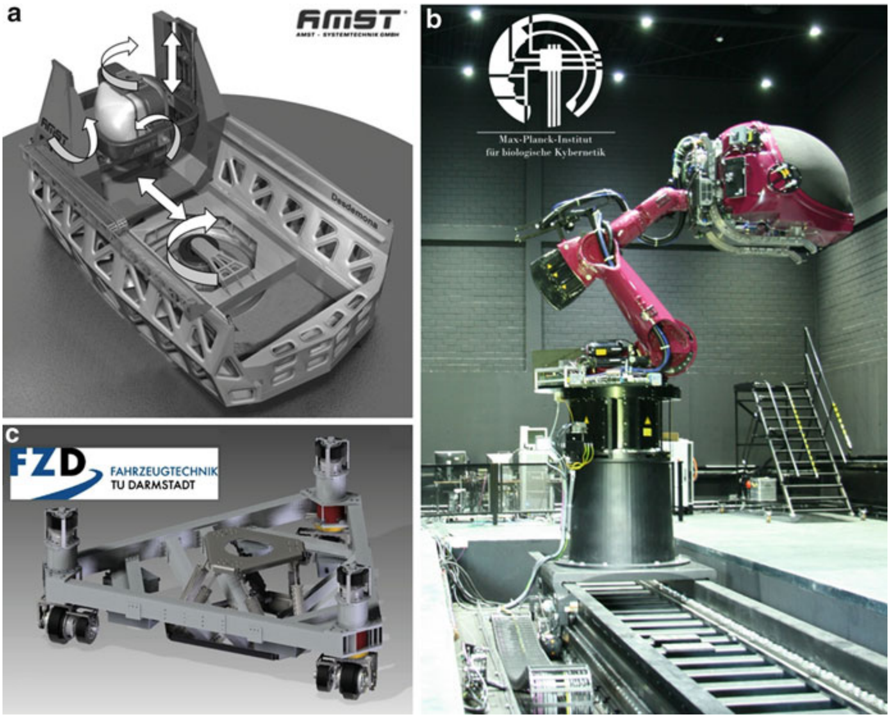


Fig. 3 Driving simulators with alternative motion system concepts

- “Desdemona” system by AMST with realization at TNO (Wentink et al. 2008), which makes the necessary acceleration forces available in six axes on the basis of a large centrifuge and several other nested rotation platforms.
- Robot arm system by Kuka with realization at the Max Planck Institute in Tuebingen (Nieuwenhuizen and Bülthoff 2013). Here the motion system is based on a production robot, which is available in large quantities and thus at relatively low cost. An additional long motion axis is provided by a rail system.
- The concept of a “Wheeled Mobile Driving Simulator” is pursued in a research phase at the University of Darmstadt (Betz et al. 2012). If the free-driving system could be used on a sufficiently large area, the necessary motion space would be available at relatively small cost.

2 Daimler’s Dynamic Simulator as an Example of a Driving Simulator Design

In 2010, Daimler put its new dynamic driving simulator into operation (Zeeb 2010) (see Fig. 4). It was designed on the basis of 30 years of experience with its predecessor, Daimler’s first-generation simulator in Berlin. The detailed motion system design was aimed at performance assessment of chassis concepts and driving dynamics control systems. Variability and close to real-world overall driving experience were the secondary design goals with respect to evaluating driver assistance systems with a large number of test persons.

Fig. 4 Daimler’s dynamic simulator in Sindelfingen: motion system

2.1 Motion System

The driving simulator is based on a motion system with a 12.5 m long linear rail with electrically propelled slider and an electromechanical hexapod on the slider. On top of the hexapod, there is the lightweight carbon fiber dome, which contains the vehicle cabin. A rotating platform allows the vehicle inside the dome to be turned by 90°, so that the linear rail can be configured to represent longitudinal and transverse motion. Accelerations in all directions can reach up to 10 m/s^2. The motion system can therefore provide precise acceleration forces in all six degrees of freedom to the balance (vestibular) organ of a test person.

The design of the dynamics of the motion system is based on the requirements for vehicle dynamics investigations up to the limits of tire-road contact. Because of this goal, a second linear rail for improved simulation of turning maneuvers in city driving was omitted due to the restrictions in dynamics it would have implicated. This requires avoiding turning scenarios in the conception of simulator experiments, although in practice this has not been a major restriction for the arising test tasks.

The entire motion system was realized with the objective of as little friction as possible since the friction forces need to be compensated in order to achieve a precise motion perception. The longitudinal rail is equipped with an air-bearing system, which requires highly accurate manufacturing and installation of the guide rails. Friction reduction and compensation were also considered very carefully in design and control of the electromechanical hexapod actuators.

2.2 Driver's Environment

Investigations with test persons, which are needed for the conception and verification of driver assistance systems, particularly require the test persons to feel completely as if they are in a real driving situation; they should immerse themselves in the virtual environment. Therefore, the entrance to the dome of the driving simulator is arranged in such a way that the test person does not actually see the driving simulator technology. Instead, the person entering the dome finds a complete real vehicle standing on a road, with a virtual scene already projected all around him (see Fig. 5). All control elements of the vehicle and all other visible and audible items are designed to be identical or at least similar to the real experience. The test person sets the vehicle on the road in motion with the same procedure (seatbelt, key turn, hand brake, etc.) as in a real-life vehicle. The instructions given to the test persons are also chosen carefully to imply the experience of driving of a real vehicle and not a simulator.

The driver actually sits in a real cabin taken from a vehicle after all unnecessary components have been removed. In the cabin, some additional actuators have been installed, which imitate the feel of the pedals and steering wheel realistically, according to the speed and other vehicle conditions. Different passenger car cabins

Fig. 5 Daimler's dynamic simulator in Sindelfingen: vehicle in the dome

and truck cabins are used depending on the type of investigation. The cabins can be changed quickly using a standardized mechanical and electrical connection system.

2.3 Visual System

While the motion system provides the correct feeling of acceleration forces to the test person, the image system is responsible for the impression of speed and continuous movement. Eight Liquid Crystal on Silicon (LCOS) projectors with Quad XGA (QXGA) resolution (2,048*1,536 pixel) on the dome's ceiling project a 360° scene for the driver onto the inner surface of the spherical dome. The two outside mirrors of the vehicle cabin are also replaced by displays, which provide the correct rear view as seen from the position of the driver. In combination with traffic simulation software and 3D picture rendering software, a close-to-reality representation of driving situations and driving maneuvers is provided. Day and night scenarios as well as numerous weather conditions can be simulated.

Resolution, shading, and reflections are important properties of the picture system for realistic representation. There has been substantial progress in these aspects of rendering virtual worlds in simulators in recent years. For night driving conditions, correct simulation of vehicle head lamps and street lights and their resulting illumination are a challenge. For highway driving, the other traffic may be represented by vehicles only, simulated as moving, but rigid 3D objects with simple models for driver behavior. For inner-city scenarios, naturally moving objects like pedestrians, cyclists, traffic lights with changing light signals, and objects moved by the wind are increasingly important. For testing driver assistance systems, it is essential to control the position and timing of all road users in relation to the test person's vehicle, according to the needs of the specific experiment.

2.4 Sound System

The sound of driving plays another substantial role to enable test persons to immerse themselves in the virtual world. Engine and driving noises are therefore represented correctly by a sound system, dependent on engine power, engine speed, and driving speed. Uneven road surfaces should also be reflected in the noise. To this end, sound samples from road driving and test stands are adapted and mixed as necessary for a given driving condition (Krebber and Sottek 2000). Without such noise cues, it is much harder for test persons to control the speed of the vehicle. Spectral shifts (Doppler effect) produced by driving past other vehicles must be represented correctly to make the test persons perceive the traffic situations as realistically as possible.

2.5 Models of Vehicle Dynamics and of the Scenario

The entire vehicle motion must be simulated in the driving simulator in a mathematical model that has real-time capability since the driver is a substantial member of the closed-loop control. The motion perception must match the vehicle, so different simulation model parameters describe the behavior of a specific vehicle. Since simulations of commercial vehicles require complex and very different vehicle models (including trailers, articulated busses, etc.), a flexible interface for different real-time simulation models is realized within Daimler. Vehicle dynamics control systems are tied in via the same interface. For the integration of road surface models and static environment models (road scenes), open standards (e.g., "Open Drive," Dupuis et al. 2010) are used.

Simulating driver assistance systems requires an extended interface, which also covers the representation and behavior of other road users. First, the surrounding traffic must be simulated. This requires behavioral models of the road users for all relevant situations. The output signals of sensors (radars, cameras, etc.), which have to be derived from the actual road scene and traffic scenario, must also be provided. To this end, suitable simulation models of how the sensors perceive the scene are necessary.

2.6 Representing Motion in a Limited Motion Space

Due to the physical relations (the stroke is the double integral of acceleration and thus of the perceived forces), a longitudinal motion, which can be very long for a real vehicle, cannot be reproduced in a simulator without trade-offs. Here it helps

Table 1 Sensitivity thresholds (approximate values, Zacharias 1978)

Motion	Direction/axis	Acceleration	Speed	Frequency range with highest sensitivity
Linear	Longitudinal	0.17 m/s^2	–	Approx. 1 Hz
	Lateral	0.17 m/s^2	–	Approx. 1 Hz
	Vertical	0.28 m/s^2	–	Approx. 1 Hz
Rotational	Roll	4–5°/s^2	Approx. 3.0°/s	Approx. 1–10 Hz
	Pitch	4–5°/s^2	Approx. 3.6°/s	Approx. 1–10 Hz
	Yaw	4–5°/s^2	Approx. 2.6°/s	Approx. 1–10 Hz

that the human vestibular organ possesses perceptional thresholds, below which a motion cannot be perceived (Zacharias 1978; see Table 1). Due to this fact, in a driving simulator, a continuous longitudinal acceleration or deceleration can be represented by introducing a tilting of the dome with a rotational acceleration level below the perceptional thresholds. Subsequently, a component of the gravitational force acts in a longitudinal direction of the vehicle and simulates longitudinal acceleration or deceleration. With the visual system providing a consistent motion illusion, the driver in the simulator will not perceive the trick. This procedure is called “tilt coordination.”

In addition, humans cannot determine the absolute value of accelerations accurately. For this reason, the acceleration forces can be scaled (usually with a scale factor between 0.6 and 1) without degrading the motion perception too much. In total, tilt coordination and scaling enable motion to be simulated in the motion space of the simulator to be represented such that after a dynamic maneuver, the driving simulator returns to the center of its motion space. The filter that controls such a motion conversion is called a “wash-out filter” (Zacharias 1978).

It may be noted that the perceptional thresholds according to Table 1 are strongly dependent on whether the test person is concentrating on the motion perception, whether the test person is diverted by other tasks, or whether he is even involved in controlling the motion like an active driver (Nesti et al. 2012). For detailed evaluation of chassis systems with respect to driving dynamics, tilt coordination and scaling should be omitted completely because vehicle dynamics test drivers are quite sensitive and attentive. For this reason, the lateral motion during lane change or slalom maneuvers can be represented without any modifications in Daimler’s dynamic simulator. For experiments with test persons for driver assistance systems, scale factors of 0.8 have shown acceptable results.

2.7 Kinetosis (Simulator Illness)

Simulator illness is related to travel sickness or sea sickness and is characterized by visual symptoms and disorientation, cold sweat, and, in extreme cases, nausea

(Johnson 2005). Sensory conflict theory sees the reasons for kinetosis in incompatible sensory inputs from the vestibular and visual systems (Reason 1978). Around 5–10 % of all humans are very sensitive, while 5–15 % are quite insensitive to kinetosis. It is documented that women are generally more sensitive to kinetosis than men (Flanagan et al. 2005), trained pilots are more sensitive than untrained pilots, and young people are more sensitive than older people (Reason and Brand 1975). For a detailed analysis and prophylaxis of kinetosis, see Schlender (2008). Mental factors and active preparation for the motion perception seem to play a key role; the placebo effect in people with kinetosis is relatively high at 45 % as reported in Schmäl and Stoll (2000).

The only way to avoid kinetosis effectively is the precise coordination of sensory impressions to the test persons, in particular acceleration forces and visual motion cues. Most people become increasingly insensitive to false motion cues over time, so certain imperfections in driving simulators can be accepted. A high yaw rate with false cues seems to be the most critical motion, however. In experiments with inexperienced test persons, such driving situations that cannot be rendered well by the motion system should be avoided by the test design.

2.8 Preparatory Simulators

For the efficient use of Daimler's dynamic simulator, complex experiments are prepared in two static simulators at the same time as ongoing investigations in the dynamic simulator. They are based on identical hardware and software, but they are not equipped with a motion system. The environment is projected here with up to six channels around the vehicle cabin. In these simulators, scenarios can be optimized and a suitable operational sequence is verified before the experiments move to the real test in the dynamic simulator. They are also suitable for complete investigations if motion perception has only a minor effect, e.g., in certain assessments of new man-machine interface concepts.

3 Design of Simulator Experiments

3.1 Goal of Test Person Experiments

During the development process of driver assistance systems, the development engineers regularly test new functions and systems (as "experts"). Additional tests with typical customers/drivers ("test persons") are conducted at different times in the process to gain insights into the effectiveness and acceptance of these functions and/or systems and their usability for subsequent customers and users. This leads to the following investigation goals of test person experiments in driving simulators:

- Driver behavior when using new vehicle systems, such as driver assistance systems, particularly in critical traffic situations
- Controllability of system limitations
- Optimization of innovative user-machine interfaces
- Evaluation of the customer's benefit
- Analysis of the acceptance and usage of new systems

In comparison to tests on testing sites and/or in road traffic, simulator experiments offer the following advantages:

- No risk for drivers, other road users, or the environment
- High reproducibility of the test situation
- Use of the element of surprise
- Quick variation of driving conditions as well as vehicle and environment parameters

Possible disadvantages of driving simulator experiments are a limited mental presence (lack of realistic feeling) of the driver in the simulator, a reduced perception of danger and consequences in critical traffic situations, and the high expenditure for these experiments. Due to the geographical location of the simulator, variation within the group of test persons is usually limited to the immediate surroundings, so cultural differences in the driving behavior of drivers from other regions cannot be analyzed in detail. In driving simulator experiments, the focus is also on analyzing a well-defined and reproducible situation, while testing in road traffic aims to analyze a large number of different situations with a high variance of parameters.

3.2 Specification of Experiments

At the beginning, the objectives of the experiment have to be precisely defined and, if necessary, prioritized. This step is often underestimated in practice. The following aspects are to be described in this phase:

- What are the objectives of the experiment? Which of the abovementioned investigation goals is addressed?
- Which technical system and/or which interacting procedure is examined?
- Which aspect of the system is to be examined (daily use, critical traffic situation, system limitations, system failure, etc.)?
- How many and which system configurations and/or parameterizations are to be examined?
- To which reference basis does the investigation refer (e.g., predecessor system)?
- Which values are to be measured/determined (reaction time, brake/steering behavior, distance from vehicle in front, etc.) during the experiment?
- What is the road layout like (urban roads, country roads, highways, etc.)?
- Which group of test persons is to be examined?

From the objectives of the experiment, specific hypotheses must be derived and operationalized. These reflect the expected test result and have to be confirmed or disproved by the experiment.

A typical drive of a test person lasts about 30–45 min. and consists of the following three phases:

1. In the acclimatization phase of approximately 5 min., the test person adapts to the new vehicle and to driving in the simulator.
2. In the routine drive of approximately 20–40 min., the test person develops confidence in the virtual environment and learns to use the system to be tested in its regular function mode. Furthermore, the test person's mood changes from being in a test situation to being in a routine drive on the road.
3. In the final critical situation, the reaction of the test person is measured and later analyzed.

Before, during, and after the test drive, the test persons are interviewed according to the experiment's goal.

Special attention is paid to the design of the critical situation, which will be analyzed to validate the hypotheses, in addition to the results of the questioning. The critical situation must be a conceivable situation in daily traffic and is derived from usage of the system to be tested and its limitations. Furthermore, the criticality of the situation must be clearly defined between the extremes *extremely easy to control* and *completely uncontrollable* in such a way that the experiment delivers insights into typical usage of the system. The natural behavior of a test person in a critical situation can be evaluated only once on a test drive. Having gone once through a critical situation, the behavior of the test person in similar situations changes from normal due to learning effects and is no longer representative.

The test person must become familiar with the system being tested (with respect to function, human-machine-interface, and limitations) before using it in a critical situation. This is achieved by informing the test person about the system before driving and during the learning phase with practical experience of the system during the routine drive under harmless conditions without reducing the element of surprise in the critical situation.

In Figs. 6 and 7, two typical critical situations are shown: The critical situation 1 represented in Fig. 6 was used to investigate the PRE-SAFE® Brake system with pedestrian recognition. During an urban drive, the test person is distracted from the road ahead by a person at the left side of the road who is gesticulating actively (not visible in Fig. 6). At this precise moment, a pedestrian who was previously hidden by the van at the right side of the road runs unexpectedly into the road and stops directly in front of the test person's vehicle. The driver brakes with a certain delay due to the natural reaction time and due to the more or less intense distraction by the gesticulating person on the left. The assistance system usually recognizes the situation before the driver and initiates braking automatically. The accident rate and severity are compared for peer groups with and without the PRE-SAFE® Brake system.

Fig. 6 Critical situation 1: “pedestrian running into the road”

Fig. 7 Critical situation 2: “vehicle crossing from the right”

The critical situation 2 in Fig. 7 was developed for investigation of the BAS PLUS brake assistant with the Cross-Traffic Assist. The driver drives along a priority road in the city. Again, he is distracted by an event on the left side of the road (not visible in Fig. 7). The vehicle crossing from the right ignores the right of way of the test person’s vehicle and enters the intersection. The accident rate and severity are again compared for test person groups with and without the BAS PLUS system.

Subsequently, the following technical aspects need to be defined considering the goal of the investigation:

- Is a dynamic or a fixed driving simulator needed?
- If a dynamic driving simulator is used, is the vehicle cabin oriented along or transverse to the longitudinal axis of motion?
- Which vehicle cabin should be used?

- Is the modeling quality of the vehicle dynamics model sufficient for the goal of the experiment or are refinements necessary?
- Are hardware extensions needed in the cabin (controllers, displays, user control elements, measurement equipment, etc.)?
- Can the road layout be provided by existing road elements or are further developments necessary?
- What is the surrounding traffic like (vehicles, pedestrians, bikers, etc.)? Can it be realized by existing elements or are further developments necessary?
- Are all relevant traffic maneuvers available for the learning phase for the driver and/or for the critical situation or are further developments necessary?
- Which data and/or video streams have to be recorded?

3.3 Preparation of the Experiment

Each driving simulator experiment is based on stringent project management including definition of the responsibilities, the availability of the simulator, and a clear time line for the preparation and execution phases.

Operational preparation begins with the lengthy development of new components like road elements or traffic maneuvers that do not already exist in the tool box of standard elements. When these developments have been completed and all new components tested, all components including hardware extensions are integrated into the cabin in a preparatory simulator. The test sequences are then optimized, first regarding technical aspects and then in terms of answering the questions of the experiment. Here, especially the learning phases during the routine drive and the critical situation are to be optimized. Then the test sequences for the different systems, system configurations, and/or parameterizations are derived and tested.

Now the vehicle cabin is moved from the preparatory to the dynamic driving simulator in which the experiment is performed. In a preliminary test with a smaller number of test persons, the suitability of the test sequence for reaching the investigation goal is finally verified and a time slot for final detail optimizations is reserved in the experiment's schedule.

Parallel to the technical preparation described above, the questioning concept is developed, suitable test person groups are selected and invited, and the investigators are instructed.

3.4 Artificial Distractions

Accidents result frequently from a lack of attention and/or driver distraction and an unexpected critical traffic situation that occurs simultaneously. To analyze the reaction of a driver in such situations and/or to quantify the benefit of an assistance system, the situation in the simulator must be represented reproducibly. Part of this involves applying reproducible artificial distractions. For this purpose, animated

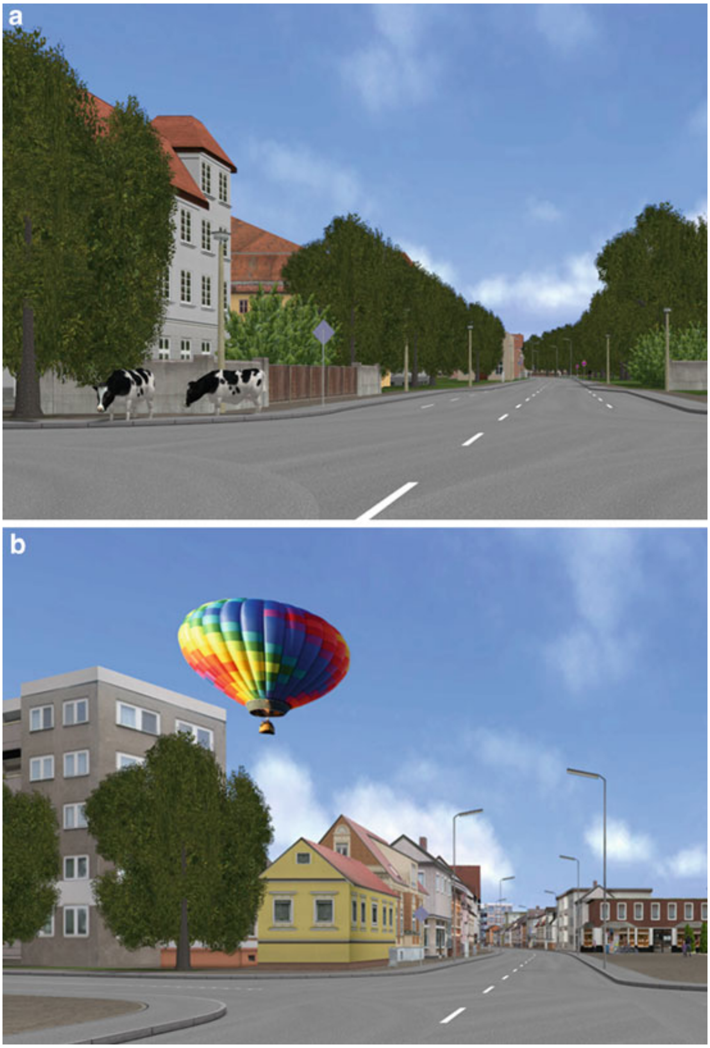


Fig. 8 Distraction of a test person: (**a**) unfenced animals at the side of the road and (**b**) hot-air balloon

graphic elements are used which are conceivable in real road traffic and attract the driver's attention, e.g., unexpected animals such as unfenced cows at the side of the road or a hot-air balloon directly over the houses on the left side of the road (see Fig. 8). These graphic elements should appear unexpectedly and must occur at precisely the same time as a critical traffic situation, such as a vehicle coming onto the intersection from a road on the right without right of way (see Fig. 7). Besides graphical elements outside the vehicle, near-reality control tasks in the vehicle (finding a telephone number in a list, sending a text message, etc.) and unnatural distractions (pressing a key upon appearance of a signal in the field of vision, etc.) may also be used depending on the task under investigation.

3.5 Learning Effects

Drivers are usually familiar with the assistance systems in their vehicle before they find themselves in an (infrequent) critical traffic situation. This must be considered when evaluating the benefit of an assistance system. In driving simulator experiments, however, the test persons are confronted with future systems that are unknown to them. A key aspect of the experiment design is to emulate the learning process during the short test drive, such as becoming familiar with the system, its functions, and its limitations. System descriptions and briefings by the investigators are used before the test drive as well as experiencing the system during the routine drive. The aim is to complete the emulated learning process before the critical situation arises. Wrongly supplied information and overemphasizing the system limitations can have a relevant impact on the test result. With too little information and learning on the system, the test person does not resemble a driver who is familiar with the system. Too much information about system limitations leads to an unrealistically high level of attention of the test person, who then waits for critical system behavior to occur.

3.6 Test Person Group Selection

The selection criteria and number of test persons have a large influence on the valid interpretation of the test results and particularly on their transferability to the use of a system in a real vehicle. The relevant selection criteria of the test persons and their distribution within the peer group (see Table 2) must therefore be defined under consideration of the experiment's goal and hypotheses. The peer groups of test

Table 2 Fictitious example of selection criteria for test persons

Internal/external test persons	Internal test persons, but none from R&D department
Age distribution	30 % under 40 years
	40 % between 40 and 60 years
	30 % above 60 years
Gender distribution	60 % male
	40 % female
Vehicle models/classes/ manufacturers	No restrictions
. . .	. . .
Additional	Driver with average driving experience, annual mileage about 15,000 km, and driving license 5–15 years

persons that experience the different systems, system configurations, or parameterizations, as well as the reference system (control group) must be identical regarding distribution of their relevant characteristics and be representative for the user group to be examined.

To compare different systems, system configurations, or parameterizations between each other and relative to a basis, 30–50 test persons per group are appropriate. The release of a system requires a much larger number of more than 100 test persons (Bubb 2003, Weitzel and Winner 2012; see also ▶ Chap. 12, "User-Oriented Evaluation of Driver Assistance Systems").

3.7 Evaluation of Experiments with Test Persons

The hypotheses derived from the experiment goal are the basis of the analysis (Sect. 3.2). The hypotheses can focus on purely objective data, e.g., the fuel consumption determined in the simulation (hypothesis: The consumption will be reduced with a new energy-saving driving style program), on purely subjective data (hypothesis: When purchasing a new vehicle, men focus more on engine power than women), or on a combination of objective and subjective data (hypothesis: The probability of purchasing a new energy-saving program depends on the reduced consumption that can be achieved).

During the experiment, the objective data is taken as measurements from the simulation, and the subjective data is acquired through interviews and questionnaires from the test persons. For further analysis, programs like MATLAB are frequently used for the objective data and SPSS for the subjective data. SPSS then offers the possibility to integrate the results from MATLAB. The hypotheses are validated with the help of statistical tests (e.g., hypothesis tests or correlations).

An investigation of the BAS PLUS brake assistant is described in detail in ▶ Chap. 12, "User-Oriented Evaluation of Driver Assistance Systems" of this book.

4 Transferability of Results

4.1 Validation of Driving Simulators

In literature, the distinction between absolute and relative validity as suggested by Blaauw is commonly used (Blaauw 1982). Absolute validity is the extent to which the results of the simulation agree exactly with real data in numerical terms. Relative validity describes the extent to which altering a factor in a simulation has the same effect as in a real study, even if the numerical data does not agree exactly. For the investigation of driver behavior, absolute validity is not inevitably necessary, although relative validity is important and often sufficient. Most validation studies come to the conclusion that for the driving simulators under examination, relative validity can be assumed regarding the most important driving parameters (e.g., speed or lateral position), but absolute validity cannot (Mullen et al. 2011). For this reason, new systems are compared with an existing reference system for comparison (e.g., predecessor system) in all studies.

With regard to validation studies, it must be noted that validation can only refer to an exactly specified situation and to only one simulator. Generalizations about situations or simulators are generally of limited validity.

For validation, the data of a drive in the simulator can be compared with the data of a naturalistic drive on a real road (no investigation situation) and/or an instructed drive on a real road (investigation situation, e.g., with an investigator in the vehicle). Interviews with test persons can also elucidate differences.

Another validation approach is to analyze the transferability of learning effects between a simulator and a real road drive; i.e., can driving abilities learned in a simulator also be observed subsequently on a real road?

4.2 Realistic Behavior and Risk Perception

In principle, the following factors influence the validity of driving simulators:

- Technical restrictions, e.g., concerning the motion space or the visualization
- Design of the experiment including the instruction of the test person
- Influence by the investigator, e.g., the tendency of test persons to present themselves positively
- Consequences that are different in real traffic, e.g., injuries sustained in accidents or tickets for exceeding the speed limit (risk perception)
- Increased attention needed in the simulator, e.g., for keeping to a lane and for speed control
- Kinetosis

According to statements of test persons, they become accustomed to the simulator and the new vehicle after the aforementioned acclimatization phase of approximately 5 min., so to a large extent, they report and give the impression that they are driving a real vehicle. When the vehicle cabin is installed transversely to the linear axis of motion of the simulator, the driving and guidance behavior are described as being very realistic, and strong accelerating and braking is perceived as being unusual. After a long drive in the simulator, some test persons indicate increased stress on their eyes. On the dynamic driving simulator described above, only few test persons report early signs of kinetosis, and in the fixed simulators used for preparing experiments, this rate is significantly higher.

Validation measurements produce a high level of agreement in terms of reaction times in the simulator and on the road. There is a tendency among test persons to estimate their own speed too slow in the simulator, so they tend to drive somewhat faster than intended (Tenkink and Van der Horst 1991). This contradicts the intention of the test persons to behave correctly in the simulator and, therefore, not to exceed the maximum speed limit. Lane keeping in the simulator is less exact than in reality (Blaauw 1982).

In general, the test persons in the driving simulator of Daimler AG intuitively behave very close to reality, which is demonstrated clearly in potentially dangerous traffic situations in particular. Thus the intended leaving of the road, driving over curbs, or driving through crash barriers practically does not occur. Before they leave the vehicle, the test persons check whether they can open the door safely by looking over their shoulder. To sum up, very good relative validity can be observed for the investigation goals in practice.

5 Summary and Outlook

A driving simulator generally consists of a driver cockpit with typical vehicle control elements like a steering wheel and pedals. The sensory impressions while driving are presented to the driver by visual means, showing the environment, road, and other traffic, and by audible means, producing appropriate vehicle and ambient noise. Dynamic simulators also present acceleration forces to the driver via a motion system. The arrangement of all these components can vary considerably.

Driving simulators in vehicle development are used to analyze interaction of the driver with the vehicle and/or with new vehicle systems, like drive assistance or dynamics control systems. In early development phases, new concepts can be assessed, and in later development phases, the focus may be on optimizing the function and its user interface. For these analyses, selected test persons drive a vehicle with a new system in several road or traffic conditions. The test persons' behavior and reactions are monitored and evaluated by measurements, videos, and questionnaires. The behavior of several peer groups of test persons is usually compared to each other to prove the effectiveness of a function in different variations.

Driving simulators have been established as a valuable tool, especially for assessing driver behavior in reproducible critical situations without safety risks, thus allowing vehicle and environmental parameters to be modified easily and quickly at the same time.

In the automotive industry, a shift to further increase validation activities based on simulations rather than road tests is expected. This requires a professionalization of operating driving simulators, transforming them from exotic research tools into efficient service facilities. Investigation content in the coming years will be affected by issues concerning autonomous driving and integral views of the vehicle, as well as communication and entertainment functions. At universities and research institutes, new concepts of driving simulators are still necessary to find better ways for overcoming the restrictions of current implementations.

The technical advancement of driving simulators will concentrate on avoiding current deficits such as when driving along curves with high yaw rates or limited ability to judge near-field distances, e.g., through the use of 3D visual systems, or on further improving motion perception. New simulator concepts and detailed optimizations of existing concepts will be developed. While there will still be an increasing number of different mechanical simulator concepts, common tools such as those for visual representation of the real world, simulating road users and traffic scenarios, as well as interfaces of the different components will see more standardization.

References

Betz A, Winner H, Ancochea M, Graupner M (2012) Motion analysis of a wheeled mobile driving simulator for urban traffic situations. In: Proceedings of the driving simulation conference, Paris

Blaauw GJ (1982) Driving experience and task demands in simulator and instrumented car: a validation study. Hum Factors 24(4):473–486

Breuer J, Käding W (2006) Contributions of driving simulators to enhance real world safety. In: Proceedings of the driving simulation conference Asia/Pacific, Tsukuba

Bubb H (2003) Wie viele Probanden braucht man für allgemeine Erkenntnisse aus Fahrversuchen? (How many test persons are needed for general investigations in driving tests?) In: Landau K, Winner H: Fahrversuche mit Probanden – Nutzwert und Risiko. Fortschr.-Ber. VDI Reihe 12 Nr. 557. Düsseldorf

Dupuis M, Strobl M, Grezlikowski H (2010) OpenDRIVE 2010 and beyond – status and future of the de facto standard for the description of road networks. In: Proceedings of the driving simulation conference DSC Europe, Paris, pp 231–242

Flanagan MB, May JG, Dobie TG (2005) Sex differences in tolerance to visually-induced motion sickness. Aviat Space Environ Med Aerosp Med Assoc 76:642–646

Johnson DM (2005) Introduction to and review of simulator sickness research. DTIC document

Käding W, Hoffmeyer F (1995) The advanced Daimler-Benz driving simulator. Society of Automobile Engineers, Warrendale. SAE technical paper 950175

Käding W, Zeeb E (2010) 25 years driving simulator research for active safety. In: Proceedings of the international symposium on advanced vehicle control (AVEC 2010), Loughborough

Krebber W, Sottek R (2000) Interactive vehicle interior sound simulation. ISATA ‘00. Automotive & Transportation Technology, Dublin

Mullen N, Charlton J, Devlin A, Bedard M (2011) Simulator validity: behaviors observed on the simulator and on the road. In: Fisher DL, Rizzo M, Caird J, Lee JD (eds) Handbook of driving simulation for engineering, medicine, and psychology. CRC Press, Boca Raton

Murano T, Yonekawa T, Aga M, Nagiri S (2009) Development of high-performance driving simulator. SAE Int J Passeng Cars Mech Syst 2(1):661–669

Nesti A, Masone C, Barnett-Cowan M, Robuffo-Giordano P, Bülthoff HH, Pretto P (2012) Roll rate thresholds and perceived realism in driving simulation. In: Proceedings of the driving simulation conference, Paris

Nieuwenhuizen FM, Bülthoff HH (2013) The MPI CyberMotion simulator: a novel research platform to investigate human control behavior. J Comput Sci Eng 7(2):122–131

Nordmark S, Jansson J, Lidström M, Palmkvist G (1985) A moving base driving simulator with wide angle visual system. In: The TRB conference, session on simulation and instrumentation for the 80s, Washington, DC

Reason J (1978) Motion sickness adaptation: a neural mismatch model. J R Soc Med R Soc Med Press 71:819

Reason JT, Brand JJ (1975) Motion sickness. Academic, New York

Schlender D (2008) Simulatorkrankheit in Fahrsimulatoren (Driving sickness in driving simulators). Zeitschrift für Verkehrssicherheit 54(2):74–80

Schmäl F, Stoll W (2000) Kinetosen. HNO 48:346–356, Springer, (http://www.neuro24.de/schwind.htm)

Slob JJ (2008) State-of-the-art driving simulators, a literature survey. DCT 2008.107, DCT report. Eindhoven University of Technology

Tenkink E, Van der Horst ARA (1991) Effects of road width and curve characteristics on driving speed. Report IZF 1991 C-26. TNO Institute for Perception, Soesterberg

Weitzel A, Winner H (2012) Ansatz zur Kontrollierbarkeitsbewertung von Fahrerassistenzsystemen vor dem Hintergrund der ISO 26262. FAS 2012 – 8. Workshop Fahrerassistenzsysteme, Walting

Wentink M, Pais R, Mayrhofer M, Feenstra Ph, Bles W (2008) First curve driving experiments in the Desdemona simulator. In: Proceedings of the driving simulation conference, Monaco

Zacharias GL (1978) Motion cue models for pilot-vehicle analysis. AMRL-TR-78-2

Zeeb E (2010) Daimler's new full-scale, high-dynamic driving simulator – a technical overview. In: Conference proceedings, driving simulation conference Europe, Paris

Vehicle in the Loop

10

Guy Berg, Verena Nitsch, and Berthold Färber

Contents

Abstract

Modern driver assistance systems are designed to intervene actively during driving in order to avoid an imminent accident or to reduce potentially health-threatening consequences of an accident. Due to the severe risks that are associated with such interventions in critical situations, it is of paramount importance that these systems are evaluated thoroughly in the development stage, with methods that not only demonstrate technical functionality but also

---

G. Berg (✉)
BMW Group, München, Germany
e-mail: mail@guyberg.de

V. Nitsch • B. Färber
Institut für Arbeitswissenschaft, Universität der Bundeswehr München, Neubiberg, Germany
e-mail: verena.nitsch@unibw.de; berthold.faerber@unibw.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_10

take into account the behavior of the driver as he/she interacts with the technology. With increasing complexity and criticality of the driving situation, in which an assistance system is supposed to intervene, it becomes increasingly difficult to test the interaction of system performance and driver behavior reliably and safely. The vehicle in the loop (VIL) combines a virtual visual simulation with the kinesthetic, vestibular, and auditory feedback of a real car. As such, the VIL offers a variety of new options for evaluating driver assistance systems. Thus, the VIL constitutes a viable alternative to established evaluation methods such as field studies and conventional simulators. The VIL was developed on the basis of empirical evaluations. The present article describes this development process and discusses its potential for future development.

1 Motivation

Since the introduction of ESC, it can no longer be disputed that the widespread employment of driver assistance systems can lead to substantial improvements in the safety of car drivers and other traffic participants (cf. ▶ Chap. 39, “Brake-Based Assistance Functions”). Early generation assistance systems such as ESC or ABS mostly supported the driver at the stabilization level (cf. ▶ Chap. 39, “Brake-Based Assistance Functions”). Systems, which support the driver at the maneuvering level, have typically been implemented in the form of comfort-enhancing rather than safety-enhancing systems. Due to improved environment detection technology (cf. ▶ Chap. 15, “Vehicle Dynamics Sensors for DAS”) and situation assessment capabilities, a new generation of active safety assistance functions (cf. chapter ▶ Chap. 46, “Fundamentals of Collision Protection Systems”) has appeared on the market, which intervenes in critical traffic situations at the maneuvering level in order to avoid an imminent accident or reduce potentially health-threatening consequences of an accident. As such, the development of this particular type of driver assistance systems poses new challenges to car manufacturers. It is no longer sufficient to concentrate research and evaluation efforts on the development of new technology such as sensors, control algorithms, and actuators. Due to the severity of potential risks, it is of paramount importance that these systems are evaluated thoroughly in the development stage, with methods that go beyond demonstrating technical functionality and reliability. Instead, the evaluation of the behavior of the driver as he/she interacts with the technology becomes increasingly important in the development process.

Virtual simulations are frequently used evaluation tools, as they allow for efficient and cost-effective testing processes. Even in early stages of development, virtual simulations can be used, for instance, to test new algorithms using prototypical software in the loop (cf. ▶ Chap. 8, “Virtual Integration in the Development Process of ADAS”), while new sensors and actuators can be evaluated using hardware in the loop, thus omitting the need for constructing a real car. Similarly, driving behavior and controllability issues can be examined with driving simulators (driver in the loop) (cf. ▶ Chap. 9, “Dynamic Driving Simulators”).

For final system validation, a new system must be implemented and evaluated in a real vehicle. Testing prototypes in real traffic situations with common drivers is oftentimes not possible due to legal and safety concerns. Consequently, numerous safety-critical assistance functions (e.g., the automatic emergency brake that stops the car in front of pedestrians) are oftentimes evaluated on a testing track with static or dynamic dummy objects. In many cases, it is very difficult to replicate real traffic situations with this method. Moreover, with increasing complexity of the driving situation, in which an assistance system is supposed to intervene, it becomes also increasingly difficult to replicate realistically and to test reliably the interaction of system performance and driver behavior.

The vehicle in the loop aims at addressing this problem by providing a viable alternative testing environment to field studies. Essentially, the VIL places a real vehicle into a simulated environment. Hence, the driver sees an augmented or virtual reality via a visualization medium while receiving vestibular, kinesthetic, and auditory feedback from the interaction with the real car, which he/she drives on a testing track. Thus, the VIL provides a real driving experience combined with the safety and replicability of driving simulators.

2 The VIL Operating Principle

The general operating principle of the VIL is portrayed in Fig. 1, which depicts the information architecture and flow. The required soft- and hardware components and their functions as well as logistic requirements are elaborated in the following.

2.1 The Test Track

In order to evaluate the effects and effectiveness of driver assistance systems using the VIL, a test track is required of which the positions and pathways of lanes are known. The virtual environment with which the VIL interacts needs to be tailored to the test track, so that drivers only pass virtual roads that correspond with real lanes on the test track. In order to embed a real car into the simulation, it is necessary to pinpoint the location of the car at any given time, e.g., using a DGPS reference station. If the testing venue does not feature a DGPS reference station, the necessary correction signals can also be acquired via a commercial satellite reference service. There are no requirements that are specific to the vehicle which is embedded in the simulation. Hence, the VIL can be implemented into virtually any serially produced car.

2.2 The Traffic Simulation

The traffic simulation is a central component of the VIL. The simulation can portray virtually any urban, rural, and highway traffic landscape, as long as the virtual roads

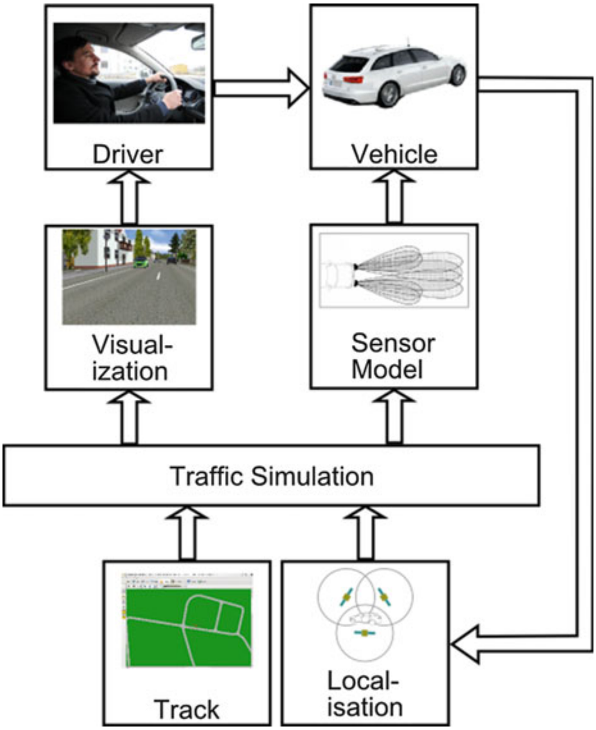


Fig. 1 Functional architecture of the VIL

that the driver is supposed to pass correspond with the real lanes of the utilized test track. Furthermore, numerous traffic situations can be simulated with autonomously operating or programmable nonplayer characters, for instance, to evaluate driver-system interactions during simple pursuit and tailgate maneuvers or complex intersection situations.

2.3 Positioning the Real Car in the Traffic Simulation

For the VIL, it is necessary that the maneuvering control of the virtual ego vehicle, i.e., the vehicle that is supposed to be controlled by the user, is decoupled from the traffic simulation, so that the virtual representation of the driver's car can be controlled independently of the traffic simulation, but is still registered by the simulation as an "active road user." By mounting an inertial sensor platform which is linked with a DGPS to the real car, its position on the test track can be located. Since the pathways of the lanes in the virtual world correspond with those of the real test track, the corresponding position of the virtual car in the virtual world can be calculated, so that the movements of the real car on the test track can be mapped onto the virtual ego car.

Fig. 2 Potential visualization forms in the VIL: (**a**) driver with HMD visualization, (**b**) screen visualization

2.4 Physical Interaction with Virtual Objects

In order to evaluate assistance systems that interact with objects in their immediate environment, the system requires information regarding potentially relevant objects in the traffic simulation. For example, in order to evaluate an automatic emergency brake function, the relative position and velocity of the preceding car in the simulation need to be determined and communicated to the system, so that the function can be triggered when a certain minimum distance between the virtual preceding car and the ego car is reached. For this purpose, the traffic simulation contains virtual equivalents of many common automotive sensors. By positioning virtual vehicles in the traffic simulation that are controlled by the VIL, information is provided to the assistance systems based on sensor models and object lists via an interface. Thus, it is possible to feed a function with the appropriate data on environment detection to the traffic simulation. Furthermore, new assistance functions can be tested in numerous traffic scenarios during early development stages before they are implemented in reality.

2.5 Visualization

While the driver receives kinesthetic, vestibular, and auditory feedback from driving the real car, he/she receives visual feedback from the traffic simulation. Currently, two options are considered for visualization, which should be selected depending on whether the VIL is primarily utilized as a development tool or for the evaluation of driver behavior (cf. Fig. 2):

(a) *Head-Mounted Display (HMD)*
As the name implies, the HMD is a monitor that attaches to the head of the user, so that information is displayed in very close proximity to the eyes (cf. Fig. 2a). Due to this close proximity, the field of view of the user is limited. To ensure a realistic viewing experience, the displayed image must be adjusted according to the driver's head movements. For this purpose, a headtracker must be installed

into the car when using a HMD as visualization medium in the VIL. On the one hand, the HMD provides an immersive experience for the driver. On the other hand, the driver's view of the real world is either partially or fully occluded by the virtual world while he/she is driving. This circumstance necessitates the presence of at least one other person that monitors potential hazards in the real world (e.g., unexpected obstructions of the road) at all times. In addition, while significant improvements have been made, modern HMDs are still rather heavy and tend to restrict drivers in their head movements.

(b) *Stationary Computer Monitor*
Displaying the virtual world of the traffic simulation on a conventional monitor instead of a HMD requires a less elaborate setup, because in this case, no headtracking is needed for the visualization (cf. Fig. 2b). Hence, a particular advantage of this visualization method is that developers can test systems in exactly the same driving situations, without the need for extensive setups or assistance from others. Moreover, since the developer can see the real environment (at least when his or her focus is not on the monitor), it might be possible for several developers to use different sections of the test track in parallel. A similar setup was used, for example, in Schuldt et al. (2014), to assess the implementation of an assistance system for construction sites. On the downside, it has been suggested that the degree to which a driver feels immersed in the virtual world rather than reality might be substantially lower with the monitor than the HMD.

While, to date, there is little conclusive evidence to suggest that fully immersive environments are a necessary prerequisite for simulator validity, it would seem more likely that the driver shows more realistic behavior when feeling immersed in the virtual world. Hence, it would seem that a HMD would be more suitable for the evaluation of realistic driver-system interaction, while a stationary computer screen would likely be sufficient in early stages of development, when the emphasis is placed on evaluating various system parameters in their effects on driving dynamics rather than driving behavior.

3 VIL Development Milestones

The VIL concept was originally developed by Thomas Bock in cooperation with Audi AG (Bock 2008). He used the traffic simulation software "Virtual Test Drive" by Vires (von Neumann-Cosel et al. 2009; Berg 2014). Subsequently, the Carmeq GmbH took over commercial distribution of the VIL, and the Human Factors Institute (IfA – "Institut für Arbeitswissenschaft") at the Universität der Bundeswehr München (UniBW) pursued further development. Unlike Bock who considered the VIL to be primarily a tool for developers with which they can efficiently test different driver assistance systems with regard to functionality, at the UniBW, it was mainly used to investigate driver behavior with assistance systems. In particular, it was hoped to gain more insight into factors that affect the controllability of driver assistance systems. While the traffic simulation software "Virtual Test Drive" is still utilized, numerous changes have been introduced by the UniBW, in particular with regard to the visualization.

3.1 The VIL with Augmented Reality Visualization

Bock decided on using a HMD for visualization, whereby he used an "augmented reality" mode. In this mode, the HMD features semitransparent displays, with which the driver can see the real environment. At the same time, virtual vehicles are displayed in such a way that the real, empty test track in the background is partially occluded by the virtual vehicles in the foreground. Thus, the driver gets the impression that the virtual vehicles are situated in front of him/her in the real environment. In an experiment, Bock demonstrated that the VIL can be considered a valid testing tool for the development of driver assistance systems (Bock 2008). In this experiment, he compared people's driving behavior with the VIL and in reality in various traffic situations. The results indicated that the participants behaved similarly in both testing environments. Furthermore, Bock stated that the participants viewed the VIL positively. Not only did they seem to adapt very quickly to the simulation, but they were also reported to praise the VIL for its ability to portray reality so closely.

On the other hand, Bock also reported a number of problems that were associated with the "augmented reality" setup. For example, with greater sunlight exposure, it became increasingly difficult to see the virtual objects. Furthermore, even tiny errors by the headtracker in the localization of the driver's head led to the misplacement of the virtual objects in the HMD, so that, occasionally, the driver received the impression that the virtual cars were driving through the road or hovered above it.

3.2 The VIL with Virtual Reality Visualization

Due to the problems that were observed with the "augmented reality" mode, it was decided that the second-generation VIL should make use of a "virtual reality" visualization (Starke and Hänsel 2011). With this form of visualization, the HMD is not transparent, so that a completely virtual scene is displayed to the driver and he/she is visually decoupled from reality (cf. Fig. 3).

In order to obtain meaningful results from experimental evaluations of driver-system interactions with the VIL that generalize beyond the experimental setting, it is vital that the VIL promotes realistic driving behavior, i.e., shows behavioral validity. This validity was demonstrated for several traffic scenarios with cross and parallel traffic in a number of controlled experiments in which naïve participants drove in reality and with the VIL (with "virtual reality" visualization) (Berg 2014; Berg et al. 2011; Karl et al. 2013; Sieber et al. 2013).

Fig. 3 Virtual reality representation in the VIL

3.2.1 Simulator Sickness

While the "virtual reality" visualization solved numerous problems that had been associated with the previous "augmented reality" mode, several issues still needed to be addressed in order to further enhance the VIL driving experience and thus ensure valid evaluation results over a variety of traffic scenarios. For instance, it was found that the virtual reality seemed to precipitate the onset of simulation sickness. Simulation sickness is a phenomenon, which is reported to occur with all virtual reality-based simulations (cf. ▶ Chap. 9, "Dynamic Driving Simulators"). There are many theories regarding the development of simulator sickness, but none of them can satisfactorily explain the pathogenesis and observed symptoms (Verburg et al. 2002). It is believed that specific properties of HMDs, such as aperture angle and ocularity (e.g., monocular/binocular/bi-ocular displays), among others, might have an effect on the onset of simulator sickness and the severity of the experienced symptoms.

In an effort to determine the best possible HMD configuration for the VIL, another experiment had been conducted to examine the influence of different HMD configurations on perception, driving behavior, and simulator sickness (Berg 2014). In this experiment, the NVIS nVisor SX111 was tested, which features a visual aperture angle of 102° on the horizontal axis and 64° on the vertical axis, as well as the NVIS nVisor ST50 with a considerably smaller aperture angle of 40° horizontally and 32° vertically. The former only provides a stereovision option, while with the latter, both mono and stereo display options were tested. In the experiment, participants were randomly assigned to one of three HMD display conditions (SX111 stereo/ST50 mono/ST50 stereo), and each participant drove in four different traffic scenarios which were designed to highlight in particular depth perception and field of view. The results indicated that neither perception nor behavior of the drivers differed significantly between the three HMD configuration conditions. Furthermore, no statistically significant differences were found with regard to simulator sickness. However, a tendency was observed for participants to report more symptoms with the HMD SX111, which features stereo displays and a larger aperture angle. This HMD was also heavier and more cumbersome to use compared to the smaller ST50. This was reflected in participants' ratings of the HMD configurations, in which the SX111 received the lowest ratings. Since the two HMDs differed in more than one respect, it is difficult to draw any conclusions from this experiment regarding possible factors that contribute to simulation sickness. However, the results suggest that ocularity might be less important than the aperture angle or overall wearing comfort. Based on these studies, the VIL at the UniBW was equipped with the ST50.

3.2.2 Latency

The conversion from "augmented reality" to "virtual reality" also uncovered a distinctly perceptible time lag between drivers' actual head movements and those that occur in the virtual world. Since this type of latency is suspected to have a large influence in the development of simulator sickness and also directly affects driving behavior, it should be avoided or at least minimized. An analysis of the system indicated that the observed latency could be partially attributed to the employed method of headtracking, which required at least 70 ms in order to register a head movement. Another causal factor seemed to be the traffic simulation software, which required 50 ms in order to display the head movement in the virtual reality. With the addition of further latency times due to the visualization medium, a total system latency of at least 150 ms had been identified. In order to reduce this latency, the previous headtracking method, which had only utilized a single, optical headtracker (~70 ms at 55 Hz) for the assessment of the head position and alignment, was supplemented with an inertial measurement unit (IMU). Unlike the optical tracker, the IMU has the advantage, which the rotation rate of the object, at which it is attached, can be measured without substantial delay and at a high frequency (~2 ms at 100–512 Hz). However, since an IMU only measures the rotation rate, it does not provide information about the current absolute position of the object. By fusing both data sources, the disadvantages of the individual sensors can be compensated and a stable, absolute head orientation in x-, y-, and z-direction can be calculated with a latency of less than 10 ms. In order to compensate the latency that is caused by the simulation software, the virtual head rotation is linearly extrapolated based on the actual head orientation and rotation rate, so that the latency due to the display of rotation movements of the head is further reduced. Since mostly head rotation and fewer translatory movements occur during driving, the position of the head is still measured via the slower optical tracker. A first blind, randomized experimental evaluation indicated that participants preferred the new tracking method over the previous one (Berg 2014).

The validity of the improved VIL was evaluated in another controlled experiment. This time, an urban traffic scenario was simulated in which participants were instructed to navigate through passageways that were marked with delineator posts. These passageways were systematically varied in their width, and participants were randomly assigned to a VIL or reality experimental condition. The results showed that VIL drivers tended to judge the lane width as more critical and significantly reduced their speed when passing through them compared to drivers in the reality condition. However, overall the drivers in both conditions displayed comparable behavioral tendencies. Hence, behavioral validity was demonstrated for the VIL with the new configuration settings in this particular scenario (Rüger et al. 2014; Purucker et al. 2014).

3.3 The VIL with Video-See-Through Visualization

With these improvements, the "virtual reality" approach seems overall more suited to the validation of advanced driver assistance systems than the "augmented reality" visualization. In particular, the susceptibility of the superimposed virtual objects to bright exterior lighting limits the use of augmented reality. On the other hand, the "virtual reality" approach has the distinct disadvantage that the entire view of the vehicle's interior is missing. With the current state of the art, rendering a virtual cockpit with arms and hands of the driver, the steering wheel movements, and dashboard displays would take too many resources. However, this level of detail would be necessary, in order to ensure a high degree of immersion into the virtual world.

To address this problem, a new approach is currently being pursued at the UniBW, with which real footage of the interior is integrated with the "virtual reality" display on the basis of a video-see-through "augmented reality" visualization. For this purpose, a video camera, which films the current view of the driver, is mounted to the HMD. The video footage of the windshield view is then segmented and replaced by the current "virtual reality" view of the traffic simulation. The

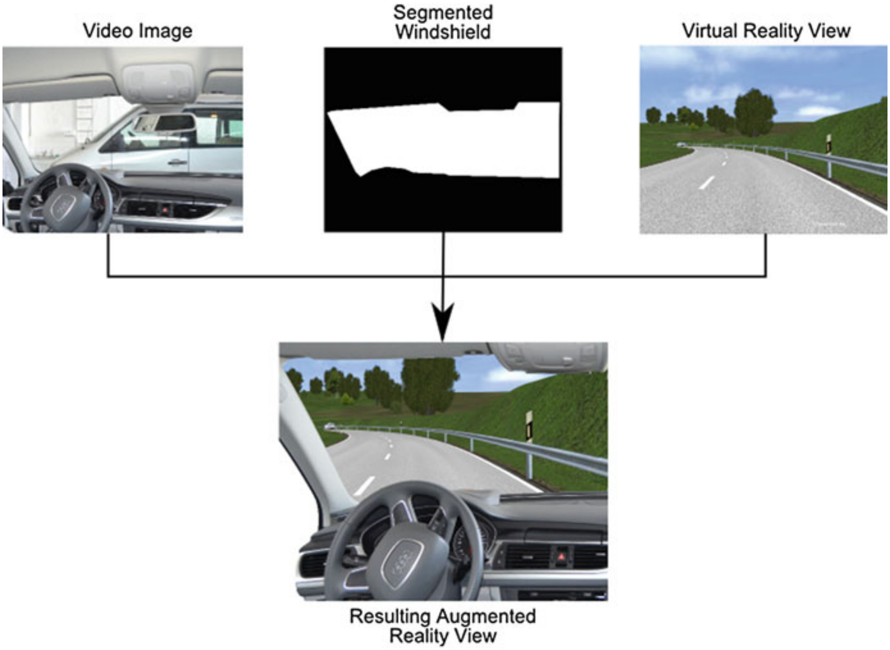


Fig. 4 Principle of augmented reality presentation composed of video and virtual reality image

windshield view is computed based on a 3D model of the car and the current head orientation of the driver. This has the advantage, which the segmentation occurs independently of the video footage, and therefore, the traffic simulation is always displayed in place of the windshield view. The composited image is then displayed to the driver via the HMD (cf. Fig. 4). In this way, the driver perceives the familiar view of the vehicle interior and sees not only the real interior but also his/her own body and movements of arms and hands. As soon as the driver looks outside, however, he/she sees the virtual traffic simulation. While a first proof of concept had already been carried out for this approach, the composited image appeared with some delays, and the segmentation of the windshield view still needs to be improved further (Berg 2014; Berg et al. 2013).

4 Conclusion and Outlook

With the introduction of the VIL concept by Bock, a tool for the development of driver assistance systems has been created which combines the replicability and safety of a driving simulation with the driving experience of reality (Bock 2008). Hence, it constitutes a new option for the cost- and time-efficient development of safety-critical driver assistance systems. Through the continuous development and improvements described in Berg (2014), the potential field of application has been extended to include the evaluation of driver-system interaction. As such, it offers a viable alternative to traditional testing environments such as field studies and conventional static and dynamic driving simulators.

While the driving dynamics experienced in the VIL equal those experienced in reality, there is still considerable room and need for improvement, in particular with regard to the visualization. Even modern HMDs are still too heavy and cumbersome, and therefore they severely affect the driver's comfort level. Furthermore, most HMD displays are still rather sluggish and technically rarely up to date. While the HMD technology had not made much progress since the great VR boom in the 1990s, there have now been substantial advances with the development of smartphone technology. For example, Oculus VR introduced a first HMD prototype, which had solved many of the visualization problems that had afflicted HMDs previously (Kushner 2014).

At the same time, the method "augmented reality" on video-see-through basis will be improved based on previous testing results, so that future users of the VIL can interact with the vehicle interior. Due to continuous advances in display technology, it may also be possible to omit the use of HMDs altogether and replace them, for instance, with contact analogous displays (Jansen 2013).

Arguably, even the most elaborate simulator setup cannot perfectly copy reality and therefore can never fully substitute field studies. However, with continuous development of the VIL, the driving experience becomes increasingly more realistic while at the same time offering the opportunity to investigate naturalistic interactions between driver behavior and safety-critical driver assistance systems in a safe environment.

References

Berg G (2014) Das Vehicle in the Loop – Ein Werkzeug für die Entwicklung und Evaluation von sicherheitskritischen Fahrerassistenzsystemen (Vehicle in the loop – a tool for the development and evaluation of safety driver assistance systems). Dissertation, Universität der Bundeswehr München

Berg G, Karl I, Färber B (2011) Vehicle in the Loop – Validierung der virtuellen Realität (Vehicle in the loop – validating the virtual reality). In: Tagungsband Fahrer im 21. Jahrhundert, pp 143–154. VDI-Verlag, Düsseldorf

Berg G, Millhoff T, Färber B (2013) Vehicle in the Loop – Zurück zur erweiterten Realität mittels video-see-through (Vehicle in the loop – returning to the Augmented Reality via video-see-through). In: Proceedings of Fahrer im 21. Jahrhundert, pp 225–236. VDI-Verlag, Düsseldorf

Bock T (2008) Vehicle in the Loop – Test- und Simulationsumgebung für Fahrerassistenzsysteme (Vehicle in the loop – test and simulation environment for driver assistance systems). Audi Dissertationsreihe, Vol. 10, Vieweg, Göttingen

Jansen A (2013) Augmented Reality in zukünftigen Head-Up Displays – Prototypische kontaktanaloge Navigationsdarstellung im Versuchsfahrzeug (Augmented Reality in future head-up displays – prototypical contact analogue navigation display in a test carrier). In: Proceedings of Fahrer des 21. Jahrhunderts, pp 189–200. VDI-Verlag, Düsseldorf

Karl I, Berg G, Rüger F, Färber B (2013) Driving behavior and simulator sickness while driving the vehicle in the loop: validation of longitudinal driving behavior. Intelligent transportation systems magazine Vol. 5(1), pp 42–57. IEEE

Kushner D (2014) Virtual reality's moment – the oculus rift could finally take VR to mainstream. In: IEEE spectrum, special report: 2014 Top Tech to Watch, IEEE

Purucker C, Rüger F, Neukum A, Färber B (2014) Comparing the perception of critical longitudinal distances between dynamic dricing simulation, test track and vehicle in the loop. In Proceedings of the 5th international conference on applied human factors and ergonomics, pp 421–430. AFHE Conference, Krakow

Rüger F, Purucker C, Schneider N, Neukum A, Färber B (2014) Validierung von Engstellenszenarien und Querdynamik im dynamischen Fahrsimulator und Vehicle in the Loop (Validating narrow passageway scenarios and lateral dynamics in a dynamic driving simulator and the Vehicle in the Loop). In: 9. Workshop Fahrerassistenzsysteme, pp 137–146. Uni-DAS e.V., Walting im Altmühltal

Schuldt F, Lichte B, Maurer M, Scholz S (2014) Systematische Auswertung von Testfällen für Fahrfunktionen im modularen virtuellen Testbaukasten (Systematic analysis of test cases for driving functions with the modular virtual test kit). In: 9. Workshop Fahrerassistenzsysteme, pp 169–179. Uni-DAS e.V., Walting im Altmühltal

Sieber M, Berg G, Karl I, Siedersberger KH, Siegel A, Färber B (2013) Validation of driving behavior in the vehicle in the loop: steering responses in critical situations. In: 16th intelligent transportation systems conference, pp 1101–1106. IEEE-ITSC, Le Hague

Starke A, Hänsel F (2011) Vehicle in the Loop – Fahrerassistenzsysteme mit Virtual Reality im realen Fahrzeug testen, entwickeln und erleben (Vehicle in the loop – testing, developing and experiencing driver assistance systems with virtual reality in a real car). In: AAET – Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel Symposium, Braunschweig

Verburg D, van der Knaap A, Ploeg J (2002) VEHIL: Developing and testing intelligent vehicles. In: Intelligent vehicle symposium, pp 537–544. IEEE, Versailles-France

von Neumann-Cosel K, Dupuis M, Weiss C (2009) Virtual test drive provision of a consistent toolset for [D,H,S,V]-in-the-Loop. In: Proceedings of the driving simulation conference. DSC, Monaco

Part III

Test Methods

Test Methods for Consumer Protection and Legislation for ADAS

11

Patrick Seiniger and Alexander Weitzel

Contents

Abstract

The term test procedure refers to a method that describes how a system has to be tested to identify and assess specific behavior or properties by experiments. This also includes the specification of required tools, equipment, boundary conditions, and evaluation methods.

Test procedures are an essential tool to check whether desired product properties are present, which of course also applies to the development of driver

---

P. Seiniger (✉)
Referat F1, Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany
e-mail: seiniger@bast.de

A. Weitzel
Formerly Technische Universität Darmstadt, Darmstadt, Germany

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_11

assistance systems. In addition to development and release testing that mainly is performed by the vehicle or system manufacturer, there are tests with the purpose of an independent product testing that are conducted by external test organizations. These tests are needed for vehicle type approval (for admission to a specific market), in the context of applying the standard for functional safety (in both cases mainly executed by technical services (being accredited as certification laboratory)) or for customer information purposes (by a test institute for consumer protection).

The focus of this chapter is these "external" test methods. After a taxonomy of test procedures, the differences between legislation (type approval) and consumer testing are highlighted. Typical tests and the associated test setup, tools, and assessment criteria are discussed, and an outlook toward testing in the near and mid-future is given.

The adjacent area of development-related tests is described in this chapter only to differentiate these from consumer and regulatory tests. Assessment methods addressing functional safety (although part of the development process but with a different focus and optionally carried out by external organizations) can be found in ▶ Chap. 6, "Functional Safety of Driver Assistance Systems and ISO 26262" and in the standard (ISO 26262 2012).

Test method requirements for development, consumer, and legislation testing are structured based on a taxonomy of test methods. To clarify the taxonomy, examples are selected and briefly described. Detailed information about some of the examples can be found in ▶ Chaps. 12, "User-Oriented Evaluation of Driver Assistance Systems," ▶ 13, "Evaluation Concept EVITA," and ▶ 14, "Testing with Coordinated Automated Vehicles."

In addition to the evaluation methods and metrics, test tools are the core of a testing method. Requirements of these tools are described and explained. Finally, the challenges posed on test methods by increasingly complex networked driver assistance systems are discussed and possible solutions outlined.

1 Taxonomy of Test Procedures

For the evaluation of driver assistance systems, a wide variety of methods is used. The procedures vary depending on the objective of the test (what kind of system functionality is to be checked against the criteria) and the available stage of development of a system, resulting from the following development process sequence: The system development starts with the general formulation of a proposed system benefit, from which system functionality and thereby system requirements are derived. After that, the system is specified and component development starts. Verification and validation tests give evidence if the requirements are fulfilled and if the planned benefits can be achieved.

1.1 Testing in the Product Design Process

For an overview on test methods, a taxonomy based on the V-model of product development (V-model 2013) is set up. During the development process, the level of abstraction of the respective system properties to be tested first decreases down to component development and then increases again. At the same time, the focus of development tests moves from system design and layout toward the validation of desired properties.

In horizontal levels of the V-model, each requirement of the respective driver assistance system is defining a test method needed for verification. In the left branch of Fig. 1, during the product definition and design, test methods can help in detailing requirements and address feasibility issues. In the right branch, the focus of tests is the verification of the specifications and, at the highest level, the validation of product requirements. Within one of the level of the V-model, therefore the requirements for testing procedures in the right and left branch can be similar or even identical. Legislative requirements or standards and guidelines can be grouped into the taxonomy too. Figure 1 shows the V-model with system levels and assigns standards and guidelines to the appropriate levels.

The highest level shown in Fig. 1 is not part of the V-model. On the left side of this level, the benefit a driver assistance system provides for road safety is evaluated and set into reference with a possible additional risk caused by the system. On the right side, the actual benefit after possible introduction of the system on public roads has to be assessed. The transitions between the levels can be vague. A clear separation is not always possible, especially because a detailed risk assessment often requires a more concrete definition of the system, which demands the definition of requirements.

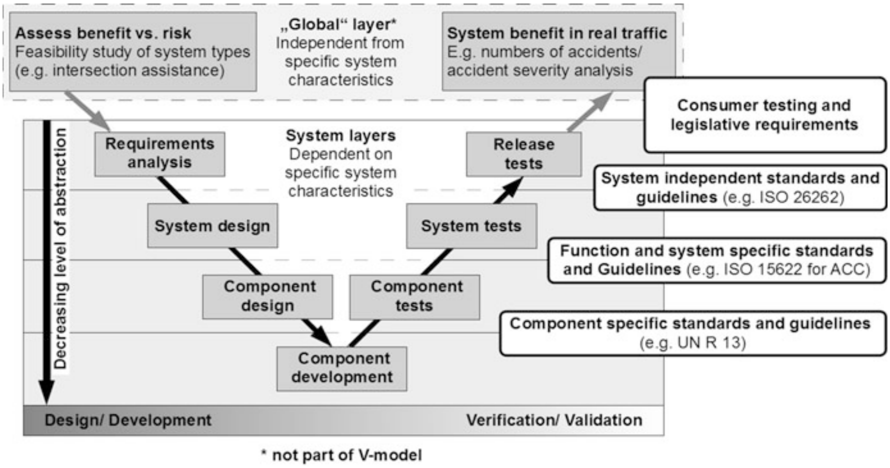


Fig. 1 V-model of the development process

In the early stages of product development, benefit and risk assessments are carried out in order to evaluate the chances of success for a new system. To support these assessments, studies will be carried out to identify the benefit in the case of justified interventions (see ▶ Chap. 13, “Evaluation Concept EVITA”) and the risk of assistance function in case of unjustified interventions (Kobiela 2011; Weitzel 2013). Given the comparable level of abstraction, these test methods can partially also be used for the final validation of system benefits and risks.

In cases in which sufficient evidence cannot be achieved with available objective methods or reasonable effort, expert judgment is a suitable complement (PReVENT 2009; Fach et al. 2010).

1.2 Differentiation by Characteristic Properties

In addition to this general and abstract view based on the product development process, test cases are often distinguished based on characteristic properties. For

Table 1 Classification of test procedures

Property	Possible characteristics	Typical characteristics in legislative and consumer testing
System category	Comfort systems, safety systems	Safety systems
Assessment level	Component, whole vehicle	Whole vehicle
Degree of virtualization	Virtual, partly virtual (X-in-the-loop), real	Real
Test environment	Simulation, lab, test track, real road (closed to other traffic), real road	Test track
Propulsion of the host vehicle	Stationary, self-propelled, towed	Self-propelled
Representation of target object	Passenger car, pedestrian, cyclist, others	
Propulsion of the target object	Stationary, self-propelled, moved externally	
Collision tolerance of the target object	No collision possible, only in emergencies, within appropriate boundaries	Within appropriate boundaries (e.g., limited test speeds)
Control of host vehicle	Subjects, professional test drivers, automated with supervisor, fully automated	Professional test drivers, automated with supervisor
Focus of the assessment	Man-machine interface, biomechanical properties, system characteristics (performance, false positives), whole vehicle performance, false activations	Whole vehicle performance, false activations
Assessment method	Questionnaire, vehicle measurements	Vehicle measurements
Rating method	Comparative rating, absolute rating	Comparative rating, absolute rating

each assessment of a driver assistance system with environment perception, typical use cases representative for driving situations need to be generated. Because these systems often cannot be tested on public roads, an equivalent situation in a test environment must be created. For example, in the case of an emergency braking function, this requires objects that represent stationary or slow-moving vehicles. Distinguishing criteria for the test method in that case would be the type of the target objects, their motion systems, or the degree of virtualization of the tests. A summary of typical criteria describing test methods is shown in Table 1. Mixed forms are also possible.

2 Testing in Legislation and Consumer Protection

In addition to the test methods that are applied during the product development process, there are two main types of test methods: those that are required for type approval and those that are used by consumer organizations for comparative assessment of products. Technical type approval regulations are nowadays almost exclusively regulated on an international basis, at least for high-volume vehicles; see ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems." Test procedures for type approval testing and related pass/fail criteria are transparently communicated at least for the activities of the UN starting from the definition phase (all minutes of the working groups and the majority of the work material are available online at www.unece.org).

The common goal of legislators and consumer protection organizations is to establish requirements that they consider to be necessary for vehicle safety (and possibly the environment) in the development process. Sometimes even system specifications (e.g., night ability of certain emergency braking systems, details of robustness) are set by regulators and consumer organizations. The strategic aspects of this issue are also discussed in ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems." However, both approaches (regulations and consumer testing) differ in their focus.

Technical regulations must be met by all vehicles sold in a specific market and therefore are mandatory for all vehicles. Due to this importance, there are high barriers for the introduction of new regulations, and these regulations typically cover only minimum standards. Usually robust and recognized scientific analyses of system benefit are needed. This often is only possible when technical systems have already been present in the vehicle fleet for some years.

The fulfillment of requirements from consumer protection organizations on the other hand is totally voluntary for the vehicle manufacturer, but it is associated with relatively low barriers for the introduction of new assessment procedures. In general, the test methods used by consumer protection organizations are set up by themselves. By being able to introduce test procedures for systems without the necessity of a robust benefit estimation, consumer protection tests are particularly suitable for promoting technical innovations.

A comparison of the characteristics of test methods is shown in Table 2.

Table 2 Comparison of test procedures for legislative and consumer testing

	Legislative testing	Consumer testing
Assessment result	Pass/fail	Typically gradual result
Purpose of the test procedure	Presence of minimum standards	Promotion of innovations and ensuring general good standards in the field
Required justifications for tests	Typically cost-benefit analyses	None
Decision methods	Vote of contracting parties (=nations)	Typically vote of limited persons in a board of directors

For the test of a variety of driver assistance systems, there are standards that are developed within the International Standardization Organization (e.g., ISO 15622 for ACC 2010). The national responsibility for ISO standards differs; in Germany, for example, the responsibility for vehicle technical standards is at the Association of the Automotive Industry (VDA). Due to this proximity to industry, ISO standards do play a limited role in regulations and consumer protection, for example, by certain test details that are referenced to standards. However, standards are relevant as a definition of the state of the art for product liability issues in judicial decisions.

2.1 Requirements from Legislative Testing

The European Union and many other countries are contracting parties to the “World Forum for Harmonization of Vehicle Regulations” from 1958 (UNECE 1995) which is widely known for the UN-R (UN-Regulations) developed under this agreement (formerly UNECE regulations). Table 3 provides an overview on regulations relevant for driver assistance systems.

It is noticeable that emergency braking systems and lane departure warning systems so far are only required for heavy commercial vehicles and buses. This is due to the severe consequences of rear-end collisions and run-off-road accidents of those vehicles in connection with inattentive drivers.

Accuracy limitations of the respective test procedures are usually not that strict and allow testing without the use of driving robots.

2.2 Requirements from Consumer Testing

An essential feature of tests of consumer protection is that in principle such experiments must be carried out without vehicle manufacturer- or system manufacturer-specific knowledge.

Since the results can significantly influence the success of a vehicle in the market, highly reproducible methods with transparent evaluation criteria are

Table 3 Overview of vehicle regulations relevant for driver assistance systems

Regulation	Title	Test procedure for driver assistance system
UN-R 13h	Uniform provisions concerning the approval of passenger cars with regard to braking	Annex 9: electronic stability control and brake assist systems, as well as minimum requirements for the functional safety of all driver assistance systems that use the brake system as actuator
		It is planned to transfer stability control and brake assist procedures into a new separate regulation
UN-R 79	Uniform provisions concerning the approval of vehicles with regard to steering equipment	Annex 6: special requirements to be applied to the safety aspects of complex electronic vehicle control systems (no test procedure). This defines limits for steering intervention and sets minimum requirements for the functional safety of those driver assistance systems that use the steering as actuator
UN-R 130	Uniform provisions concerning the approval of motor vehicles with regard to the lane departure warning system (LDWS)	Currently only for heavy goods vehicles of category N_2 (vehicle category according to R.E. 3 of the UNECE 2010) above 8 t gross weight, M_3 and N_3 (for $N_2 < 8$ t in the near future)
UN-R 131	Uniform provisions concerning the approval of motor vehicles with regard to the advanced emergency braking systems (AEBS)	Currently only for heavy goods vehicles of category N_2 above 8 t gross weight, M_3 and N_3

requested. The required reproducibility often goes beyond what is technically required (and also beyond the requirements of type approval tests).

Consumer protection tests are conducted worldwide by the various new car assessment programs (NCAPs) according to transparent procedures and assessment criteria. However, sometimes there are tests without prior notice that aim at highlighting a certain safety deficit in the media (e.g., ADAC test for automatic pedestrian braking in 2013 ADAC 2013). An overview of the individual new car assessment programs and their planning with regard to the test of driver assistance systems is found in Table 4.

The test procedure to assess automatic braking car to car in JNCAP, KNCAP, and IIHS is based on Euro NCAP's test procedure (see Euro NCAP 2014), but rating and sometimes test speeds differ. The collision warning test according to US NCAP is significantly different.

The main assessment criterion for the Euro NCAP tests on automatic emergency braking and collision warning is the speed reduction, measured for different test speeds, in a situation where the test vehicle approaches a moving or stationary test

Table 4 Worldwide NCAP tests for driver assistance systems

Name	Region	Initiator	Driver assistance system
ANCAP	Australia	ANCAP, Canberra	Automatic braking car to car
ASEAN NCAP	Southeast Asia		None
C-NCAP	China		None
Euro NCAP	EU-28	Euro NCAP, Brüssels	Automatic braking car to car (since 2014), automatic braking car to pedestrian (planned for 2016), lane departure warning (since 2014), speed assistance systems in several variations (since 2014)
JNCAP	Japan	NASVA, MLIT	Automatic braking car to car and automatic braking car to pedestrian in preparation
KNCAP	Korea		Automatic braking car to car (planned for 2015), several others in preparation
LATIN NCAP	South America	LATIN NCAP, Uruguay	None
US NCAP/ 5-star safety ratings	USA	NHTSA, Washington	Collision warning, lane departure warning
IIHS	USA	Insurance Institute for Highway Safety, Arlington	Automatic braking car to car

target (see Table 5). The test conditions are maintained highly reproducible by the use of driving robots. Test scenarios and parameters have been derived from accidentology.

While driving robots allow a good reproducibility, they could interfere with the vehicle controls and possibly abort the automatic brake intervention. This is taken into account when programming the robots – for instance, speed control mode has to be switched off before automatic braking starts.

Efficiency of a given collision warning system is assessed with a simulated driver input to the brake: a brake robot actuates the brake pedal. The brake robot profile is comparable to a regular driver's brake actuation and tuned toward the specific vehicle to be tested. The preparation of that brake profile requires the measurement of pedal travel d_4 and pedal force F_4 that corresponds to a deceleration of 4 m/s^2 in the vehicle under normal circumstances. During the collision warning test, these values are applied by the robot 1.2 s after the detection of the acoustic warning signal and pedal travel value d_4 is achieved 0.2 s after start of actuation. Pedal force control (desired value F_4) is activated when the force value F_4 is reached. Different test labs use different instruments to detect the acoustic warning; however most of those instruments use a combination of frequency detection and sound level detection.

Table 5 Comparison of automatic braking tests in Euro NCAP and US NCAP

	Euro NCAP (2014)	US NCAP (Forkenbrock 2011)
Scenario "stationary target"	Test speed 10–25 km/h (automatic braking only), 30–50 km/h (automatic and collision warning), 55–80 km/h (collision warning only)	Test speed 72 km/h (collision warning only)
		Pass criterion: warning at TTC ≥2,1 s
Scenario "moving target"	Target moving with 20 km/h, test speed 30–70 km/h (automatic braking), 50–80 km/h (collision warning)	Target moving with 36 km/h, speed 72 km/h (collision warning only)
		Pass criterion: warning at TTC ≥2,0 s
Scenario "braking target"	Initial speed of both vehicles 50 km/h, initial distance 12 and 40 m	Initial speed of both vehicles 72 km/h (only collision warning), distance 30 m
	During the test, the target decelerates with 2 and 6 m/s^2. Assessment with regard to automatic braking and collision warning (8 test runs in total)	During the test, the target decelerates with 3 m/s^2
		Pass criterion: warning at TTC ≥2,4 s

Individual test results (=speed reductions) for each test speed and test scenario (for both automatic braking and collision warning tests) are then combined to a single test result with an assessment curve. This assessment curve defines the relevance of a certain test speed in the real world and has been derived from accidentology, for instance from the German In-Depth Accident Study GIDAS (www.gidas.org).

The assessment criterion of the US NCAP collision warning tests is the warning timing only. This test is simply a pass/fail test that decides whether the collision warning will be recognized in the whole vehicle star rating.

3 Properties of Test Tools

In tests of driver assistance systems, a critical driving situation for the vehicle under test is created using suitable tools to evaluate the vehicle response. The main components are therefore targets for the representation of other vehicles (automatic braking car to car) and persons (automatic braking car to pedestrian) with matching motion systems.

3.1 Target Objects Representing Cars and Propulsion Systems

Today's emergency braking systems are not able to avoid collisions for all test speeds and test parameters, so at some point during testing, collisions between the vehicle to be tested and target objects will occur. Therefore test target robustness against collisions is required to limit test cost and time.

Fig. 2 Euro NCAP vehicle target (EVT) in testing

Such robust test objects are usually made of soft and light materials such as rubber and filled with air to present a soft surface with small and constant contact forces to the vehicle under test.

Rubber has substantially different RADAR reflection characteristics than metal. Appropriate measures to increase the target RADAR cross section have to be used, usually corner reflectors (see ▶ Chap. 17, "Automotive RADAR") or large patches of reflective foil shaped to represent vehicle structures. Compromises are required for other properties as well: optical and IR reflection characteristics are different from metal or glass, only simple shapes are possible, and components like spinning wheels are not fully replicable.

The appropriate propulsion system for the target systems needs to achieve the required reproducibility and should not interfere with the reflection characteristics of the target object.

An optimal target and propulsion system needs to be identified in this trade-off between realistic appearance, collision tolerance, and movement characteristics.

The standard target object for the majority of consumer test procedures (Euro NCAP, IIHS, JNCAP, KNCAP, ANCAP) is currently the "Euro NCAP vehicle target." EVT is shown in Fig. 2.

The EVT is made of foam and air-filled rubber structures and contains an integrated corner reflector to generate a RADAR cross section representative for a typical passenger car (2.5 m^2 RCS at 77 GHz; see Euro NCAP (2014), annex 1). It represents only the rear view of a vehicle and has a depth of approximately 1 m. Therefore it is not appropriate for future intersection or cut-in tests. The RADAR and optical properties are tuned in such a manner that the EVT is robustly classified as relevant target by all current vehicles.

The EVT can be fixed toward a movable rail with 21.4 m length, towed like a trailer. In case of impacts, the target travels along this rail (allowing more room for the vehicle under test to brake) and is stopped at the end of the rail with an industrial damper system; see Fig. 3.

Another example for a target propulsion system is the self-driving soft-crash target used in the test setup of the Daimler AG, called SimCity (see ▶ Chap. 14, "Testing with Coordinated Automated Vehicles"). Here soft-cushion parts are

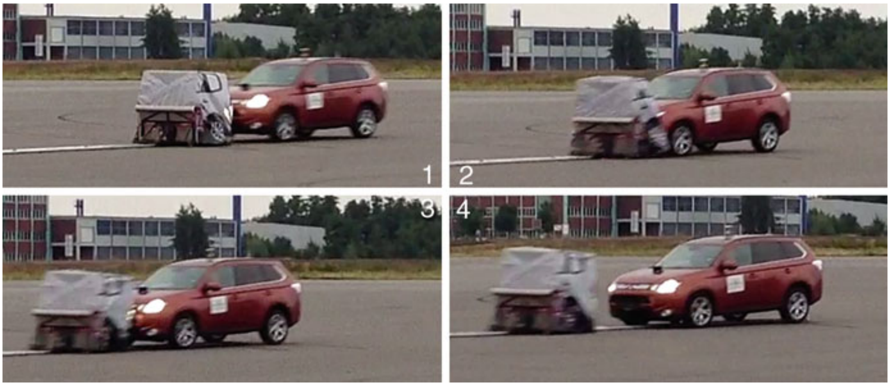


Fig. 3 Impact process for tests conducted with the EVT

Fig. 4 Driving tests with a first prototype of the ASSESSOR and BASt's propulsion system

mounted around a center box which can be remote controlled or programmed. The cushions are responsible for energy absorption as well as for providing realistic optical and RADAR target characteristics.

The European Commission-funded research project ASSESS (www.assess-project.eu) has picked up the SimCity soft-cushion concept and developed the target object ASSESSOR. The RADAR characteristics of the ASSESSOR are adjusted to match those of typical compact-class cars. In opposition to the concept of the Daimler AG, it can be combined with various different propulsion systems, for example, with the remote-controlled race kart as developed by the German Federal Highway Research Institute (BASt) (Seiniger et al. 2011); see Fig. 4.

The target object used by the US National Highway Traffic and Safety Administration, the NHTSA SS_V (Buller et al. 2013), is a CFK reproduction of the rear end of a Ford Fiesta, and its RADAR and other characteristics are designed to match those of a real vehicle. It is not impactable by itself – a separate cushion is needed to achieve emergency collision tolerance. The SS_V can be mounted on a structure similar to the EVT's propulsion system.

A nonimpactable target object is the EVITA target (see ▶ Chap. 13, "Evaluation Concept EVITA"). Experiments can be conducted only up to a TTC greater than 1 s. EVITA is mainly a tool for subjective tests, where the appearance of the target toward the test subject is important. Therefore a replica of a real vehicle rear end is mounted on a trailer. RADAR characteristics are increased with a corner reflector: opposed to sensor and system performance tests, the robust detection of the target is more important than a realistic reproduction of a real vehicle's RCS.

3.2 Target Objects Representing Pedestrians and Propulsion Systems

Emergency braking systems for pedestrian accidents will probably never be able to avoid an accident in all test cases. Using real humans as target object is – also from an ethical standpoint – never an option. A target object (pedestrian dummy) needs to resemble a human being as close as possible with the characteristics needed for typical sensors. Since there will very likely be non-avoided accidents, the target object either needs to be impactable (without causing damage to the vehicle's sensor systems), or there must be a provision to remove the target object from the vehicle's path within fractions of a second. In the second case, the dummy must be able to withstand the relatively high acceleration values generated by the removal motion. Usually these types of dummies are hanging from a massive overhead gantry. In the first case however, the dummy itself must have sufficient shape stability and the abrasion of the clothing needs to be controlled, but the dummy then can be moved with cheap and portable platform-type propulsion systems. The platforms are either remote controlled and self-driving or guided and driven by a belt (in some cases with a rope).

Regardless of the propulsion system, pedestrian dummies can be static (shape is constant, no articulated extremities) or animated (with articulated extremities). While the static dummy is simple, enduring, and easy to build, the animated one is much more realistic, which would generate test results that better match the real-world performance. The static dummy might still be acceptable for optical sensors, but new RADAR sensors are able to use the movement of arms and legs for a much quicker classification (and thus a much better system performance) (Heuel and Rohling 2012). Systems using this feature would be systematically discriminated with a static test setup. Currently, it is still unclear if and how animated dummies will achieve the necessary durability.

Current pedestrian dummy solutions can be arranged in this field made up of collision tolerance, propulsion system, and dummy complexity.

Fig. 5 Pedestrian test system FGS by company 4a, Austria (4a Engineering 2014)

Fig. 6 Pedestrian test system by Applus IDIADA, Spain (IDIADA 2013)

Austria-based company 4a Engineering offers an overhead gantry made mainly of carbon-fiber material that allows for the dummy removal within less than 150 ms while being partly invisible for RADAR sensors; see Fig. 5. Dummy removal times in that low range have a neglectable influence on automatic braking system performance. A relatively high test frequency is possible, and animated dummies are easy to implement with this type of setup. The test setup itself unfortunately is not portable and not appropriate for test labs that do not own a test track of their own.

Spanish test lab and engineering company IDIADA has developed a stationary gantry-type test setup as well (see Fig. 6) which allows for easy implementation of

Fig. 7 Pedestrian test system by Continental Safety Engineering, Germany (Conti 2014)

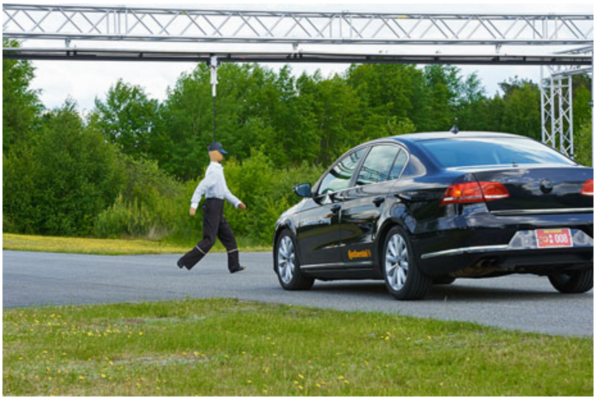


Fig. 8 Propulsion system UFO by company DSD, Austria

an animated dummy. Since dummy removal is not possible, the animated dummy needs to be robust.

Continental Safety Engineering from Germany offers a gantry construction (see Fig. 7) that allows curved dummy trajectories. It is portable, and the time required for setting up the gantry is approximately half a day.

Austrian company DSD is developing a self-propelled platform that uses DGPS and/or odometry to identify its position. This allows very flexible dummy trajectories, but the height of the platform is greater than that of belt-driven platforms; see Fig. 8.

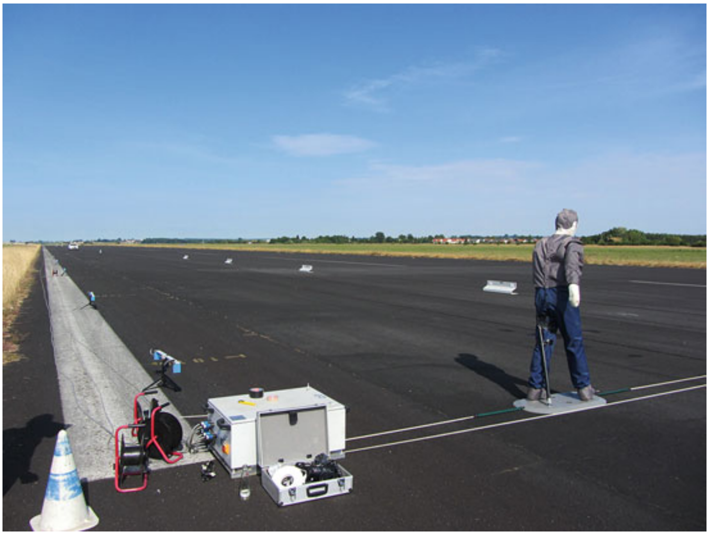


Fig. 9 Propulsion system “surfboard” by company 4a, Austria (4a Engineering 2014)

Belt-driven concepts like the “4a surfboard” (see Fig. 9) achieve much less platform height and better reproducibility than self-propelled platforms but are not as flexible in terms of dummy trajectories.

4 Testing Closer to Reality Versus Necessary Effort

The development of driver assistance systems in recent years shows that the systems are getting increasingly interconnected and support or act autonomously in an increasing share of all driving situations. This has been made possible by an increasing volume of outside information that those systems detect, process, and evaluate in real time during driving. Future test procedures need to be able to provide or at least simulate all relevant outside information during testing.

This increases the effort for the development and use of real and relevant test situations for driver assistance systems. A validation of the test tools is also required: it needs to be proven that these test tools replicate reality sufficiently accurate. For instance, impactable target objects require even today a validation whether they are appropriate for the relevant sensors or whether they offer an appropriate threat impression to the driver in subjective testing. Being appropriate for sensors not only means they are detectable but also whether they are representative for the collective of vehicles, pedestrians, or other objects in real traffic. An example of such a study for RADAR sensors can be found in Marx (2013). If these studies are not available, no connection between test results (in artificial situations) and real road traffic safety will be possible.

The same challenge applies for test scenarios. Their relevance for real traffic needs to be proven or better quantified. The increasing number of sensor types used for a driver assistance function and parameters defining the function therefore increases the effort for scenario definition, test conduction, and validation tremendously.

With an increasing number of situation parameters such as road category, precipitation, and illumination, it needs to be proven as well what share of real traffic and accident situations is addressed.

For safety systems, this estimation is possible by using accident databases, if the available data quality is sufficiently detailed. For partly and highly automated driving, a driving situation collectively addressing all possible influence parameters should be taken into account, which is currently not available. Alternatively an absurd high number of test mileage under all possible conditions is required. However, the effort to achieve this mileage is high and the transferability of results limited; see Weitzel et al. (2015).

5 Outlook: What Could Be Tested in the Future and How?

The development of new test procedures in Euro NCAP typically starts if a substantial benefit for accident numbers is expected from a new technology, and that technology is on the market or about to be introduced. This had been the case for car-to-car automatic emergency braking and pedestrian automatic emergency braking.

Accidents between cars and bicycles and accidents at intersections also occur frequently, so assistance systems to influence these kinds of accidents are candidates for introduction into the vehicle rating when available.

First Cyclist-AEB systems have already been introduced and the development of test procedures for these systems has already started. The development of test procedures for intersection accidents is about to start in 2015.

Both types of tests will require new kinds of tools: pedestrian target objects are not representative for cyclists, and also current car-to-car target systems such as the EVT and its propulsion systems are not applicable to cross traffic because they are either 2D or not sufficiently impactable.

In the near future, car targets are expected to become impactable at much higher speeds, and self-driving, overrunable platforms to move them will be available (ABD 2014); see Fig. 10. This would allow safe testing even of high relative speeds.

Basically, the test methods described here are selective methods to test few operating points of a complex technical system – often the operating points are defined by the application limits of the used tools. Generally, it is assumed that the measured performance of the system is also transferable to other operating points.

Especially with very complex information-processing systems, this assumption could easily be wrong. Therefore, to improve the robustness of systems, it would be useful to consider a much wider range of operating points in the evaluation. A conceivable approach is the increased use of simulations, wherever appropriate, in

Fig. 10 Propulsion and target system “guided soft-crash target” by company Anthony Best Dynamics, UK (ABD 2014)

the assessment of effectiveness. To increase the number of test cases, passive safety uses so-called grid methods: the vehicle manufacturer must provide own measurement values for a large number of test cases, and test labs would check a limited, randomly selected number of these test cases.

References

4a Engineering GmbH, Traboch, Osterreich

ADAC press release test of pedestrian braking assist systems. Published 14 Nov 2013

Anthony Best Dynamics, Ltd (2014) ABD DRI guided soft target vehicle. http://www.abd.uk.com/en/adas_soft_targets/abd_dri_guided_soft_target_vehicl. Accessed 14 Feb 2014

Applus IDIADA Group, Santa Oliva, Spanien

Buller W et al (2013) Radar measurement of NHTSA’s surrogate vehicle “SS_V”. Report No. DOT HS 811 817. NHTSA, Washington, DC

Continental Teves AG & Co. oHG, Frankfurt am Main (2014)

Euro NCAP (2014) AEB test protocol version 1.0. Available at http://www.euroncap.com/technical/protocols.aspx. Accessed 13 Feb 2014

Fach M, Baumann F, Breuer J, May A (2010) Bewertung der Beherrschbarkeit von Aktiven Sicherheits- und Fahrerassistenzsystemen an den Funktionsgrenzen (Evaluation of controllability of active safety and driver assistance systems at the functional limits). In: Integrierte Sicherheit und Fahrerassistenzsysteme 26. VDI/VW-Gemeinschaftstagung, 06–07 Oct 2010, Wolfsburg, pp 425–435

Forkenbrock G (2011) A test track protocol for assessing forward collision warning driver- vehicle interface effectiveness. Report DOS HS 811 501. NHTSA, Washington, DC

Heuel S, Rohling H (2012) Pedestrian classification in automotive radar systems. 19th International Radar Symposion, 23–25 May 2012, Warschau, Polen

ISO 15622 (2010) Adaptive cruise control – performance requirements and test procedures

ISO 26262 (2012) Road vehicles – functional safety

Kobiela F (2011) Fahrerintentionserkennung für autonome Notbremssysteme (Driver intention recognition for automatic emergency braking), 1st edn. VS-Verlag für Sozialwissenschaften, Wiesbaden

Marx B (2013) Bewertungsverfahren für Radareigenschaften von Personenkraftwagenkarosserie (Evaluation of radar characteristics of passenger vehicle bodies). Dissertation TU Darmstadt

PReVENT (2009) Code of practice for the design and evaluation of ADAS; 13.08.2009

Seiniger P et al (2011) Development of a target propulsion system for ASSESS. 20th conference on the Enhancement of the Safety of Vehicles (ESV). June 2011, Washington, DC

United Nations Economic Commission for Europe Inland Transport Committee (1995) Agreement concerning the adoption of uniform technical prescriptions for wheeled vehicles, equipment and parts which can be fitted and/or be used on wheeled vehicles and the conditions for reciprocal recognition of approvals granted on the basis of these prescriptions. Revision 2: Genf. Available at www.unece.org

United Nations Economic Commission for Europe Inland Transport Committee (2010) Revision of the Consolidated Resolution on the Construction of Vehicles (R.E.3). Dokument ECE/TRANS/WP.29/2010/45. Genf

V-Model (2013) Available at http://www.cio.bund.de/DE/Architekturen-und-Standards/V-Modell-XT/vmodell_xt_node.html. Accessed 16 July 2013

Weitzel A (2013) Objektive Bewertung der Kontrollierbarkeit nicht situationsgerechter Reaktionen umfeldsensorbasierter Fahrerassistenzsysteme (Objective controllability evaluation for situation- inappropriate reactions of surrounding sensor based driver assistance systems), Dissertation TU Darmstadt

Weitzel A, Winner H, Cao P, Geyer S, Lotz F, Sefati M (2015) Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung (Safety strategies for driver assistance systems with surrounding sensor systems) Forschungsbericht der Bundesanstalt für Straßenwesen, Bereich Fahrzeugtechnik. NW-Verlag, Bremerhaven

User-Oriented Evaluation of Driver Assistance Systems

12

Jörg Breuer, Christoph von Hugo, Stephan Mücke, and Simon Tattersall

Contents

Abstract

The development of driver assistance systems must be accompanied by user-oriented evaluation procedures in order to achieve high effectiveness and acceptance in everyday use while minimizing unwanted “side effects.” In addition to

---

J. Breuer (✉)
RD/FZ, Daimler AG, Sindelfingen, Germany
e-mail: joerg.breuer@daimler.com

C. von Hugo • S. Mücke • S. Tattersall
RD/FAA, Daimler AG, Sindelfingen, Germany
e-mail: christoph.von_hugo@daimler.com; stephan.muecke@daimler.com; simon.tattersall@daimler.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_12

design recommendations such as the Response Code of Practice, *user-oriented tests* with both expert and nonexpert drivers have to be conducted in various phases of the development process.

This chapter explains the suitability, planning, execution, and analysis of these user-oriented tests in a variety of test environments. It covers the whole range from driving simulator experiments, tests on proving grounds and in real traffic environments, up to and including field validation tests before final sign-off.

Recommendations are given for the experimental design, including the selection of test persons, sample size, test scenarios, and setups, as well as the choice of suitable assessment parameters and criteria. A special section is devoted to the execution and analysis of field validation tests.

Practical examples for all test settings complete the chapter, including from the area of Functional Safety. The evaluation and validation of Mercedes-Benz safety systems for rear-end crash avoidance is a recurrent theme, supplemented by examples of other types of driver assistance system.

Driver assistance systems should support the driver with appropriate information (informative systems), relieve the driver of particular subtasks (comfort systems), or help the driver to deal safely with critical driving situations (safety systems). However, they cannot and should not replace the driver or relieve him/her of the responsibility of safely driving the vehicle.

This definition makes clear that driver assistance systems must prove themselves in interaction with the driver during real-life driving on public roads, in order to achieve high *effectiveness* and *acceptance*, while also minimizing unwanted *side effects*. Therefore, the development of driver assistance systems must be accompanied by *user-oriented evaluation procedures*, complementing technical functional tests at the component and vehicle level, in order to ensure development and testing from the perspective of the customer. In addition to design recommendations, norms, and checklists, tests with both expert and normal drivers, in a variety of test environments, are particularly relevant here.

This chapter focuses on user-oriented evaluation testing including from the area of Functional Safety (ISO 26262 2011, cf. ▶ Chap. 6, “Functional Safety of Driver Assistance Systems and ISO 26262”), whereas tests for functional verification or performance ratings, which generally exclude the influence of the driver (“open loop”), are treated in ▶ Chap. 11, “Test Methods for Consumer Protection and Legislation for ADAS”.

1 Aims of User-Oriented Evaluation

Throughout the development of driver assistance systems, a number of user-oriented requirements must be met and experimentally confirmed. In practice, a formal separation of “development” and “validation” has proven effective, in order to ensure independent evaluation of the systems.

The Response Code of Practice (Response Consortium 2006, cf. Chap "▶ Guidelines for User-Centered Development of DAS") assigns the design and validation of aspects of human-machine interaction (HMI) and controllability to the phases of the development process. In addition to comprehensive checklists, fundamental requirements for validation tests with both expert and nonexpert drivers are described. The goal of CoP is to provide guidelines for establishing and proving the controllability of the driver assistance system, defined as the "likelihood that the driver can cope with driving situations including ADAS-assisted driving, system limits and system failures" (Response Consortium 2006).

Fach et al. (2010) distinguish between faulty operation in the sense of ISO 26262 2011 and fault-free operation of a system, including its so-called functional deficits. This includes interventions which are correct in terms of the technical implementation of the system, but which, due to system limitations such as incomplete environmental perception or incomplete decision-making ability of a technical system, are inappropriate for the particular driving situation.

In addition, Breuer (2007) and Eckstein (2008) define further design principles:

- Rapid familiarization.
- Consistent system behavior, conforming to user expectations (cf. product liability laws).
- Simple operation and clear display concept.
- Safety systems only intervene when a potential accident has been reliably detected.
- Effectiveness during real road use.
- Avoidance of foreseeable misuse.

2 Experimental Design

All tests and experiments should fulfill the fundamental principles of objectivity, reliability, and validity (Bortz 2005). In addition, in particular for field tests, the principles of safety and data protection of the test participants must be considered while also ensuring they are not placed under unreasonable expectations (Laurig and Luttmann 1988).

At the start of any experiment or test, the *experimental goals* must be stated and the *experimental hypotheses* defined. Based on this, an appropriate experimental design can be selected.

2.1 Tests with Experts and Intended Users

Tests with Intended Users To ensure the validity of the results, it is highly recommendable to carry out evaluation tests with participants representing the population of later users of the assistance systems (in contrast to experts involved in the development process). This is especially important for final field validation tests. It also becomes even more important; the stronger interaction between the driver and the system affects the effectiveness and use of the system. The controllability of system interventions and system limitations can also only be evaluated validly with nonexpert drivers. Here, it is important to note that safety has the top priority in such tests. Finally, by definition, studies concerning the way the system is used (including possible misuse), as well as the subjective evaluation of customer benefit and acceptance, can only be carried out with users not biased by involvement in the development process.

Tests with Experts In contrast to the continuous testing carried out by the system developers, tests with experts can be helpful and necessary, in particular for concept evaluation, i.e., very early in the development process. Evaluation should be from the point of view of the "customer," requiring a capacity for abstraction to the needs of the intended user. Tests with experts can also be used in preparation for or to assess the necessity of tests with nonexpert drivers.

For safety-critical tests, e.g., investigations of the controllability of faulty system interventions at high vehicle speeds, tests with expert drivers are indicated for safety reasons. Furthermore, in particular in the area of Functional Safety, tests with experts are required, in order to keep the sample size at a manageable level (see Sect. 2.2).

2.2 Test Person Selection and Sample Size

Test Person Selection The validity of test results and their applicability to the intended user population ("target group") depend decisively on the size and composition of the test sample. Test samples should be *representative* for the individual characteristics relevant for the particular question at hand, i.e., the frequency of properties, abilities, and needs should have the same distribution in the test sample as in the target group.

Customer acceptance depends strongly on the fulfillment of product-related expectations and needs and must be evaluated for the target customer group. For this purpose, the relevant characteristics of the *customers* should be represented as closely as possible in the test sample.

For many safety-related questions, however, the main concern is adaptation of systems to the needs and abilities of humans as drivers. The composition of the test sample should therefore be oriented more toward dimensions of individual performance, such as reaction time or driving experience.

Sample Size If the relevant characteristics are not sufficiently known, it is possible to draw sufficiently large random samples from the total population of users. Due to economic restrictions, however, the test sample will normally be smaller and also often drawn from a particular milieu (e.g., company employees). In this case, it is important that the individual characteristics which are presumed to be relevant are represented with the same frequency in all possible combinations (for formulae and examples, see Bubb 2003).

From practical experience, the Response CoP (Response Consortium 2006) states a minimum number of 20 valid datasets per scenario, in order to provide a "basic indication of validity." It is important that test persons are "naïve," i.e., that they do not have more experience or knowledge of the system than the intended customers.

Considering the controllability levels of ISO 26262 2011, in order to prove that 90 % of the total population can master a particular critical situation ($C = 2$), the binomial distribution with a confidence level of 5 % leads to a minimum sample size of 29 test subjects, all of whom must "pass" the test. Due to practical considerations of sample size, a controllability level of $C = 1$ (99 %) can only be evaluated through expert judgments (Weitzel and Winner 2012).

Representative samples are also difficult to achieve due to the time frame and financial conditions of the development process. Normally, when comparing different system characteristics, a carefully selected test sample of 30–50 subjects will be sufficient, comparative, for example, in age distribution and driving experience, to the target population. However, the final validation of a system requires significantly larger samples of between 100 and 500 subjects.

Within-Subjects vs. Between-Subjects Designs Within-subjects designs, in which each test participant performs in multiple test conditions, are particularly suited for comparing various system design options. An exception, however, is the investigation of critical driving situations, since unprepared behavior can only be observed in the first trial.

On the other hand, if the suitability of a system design option for a particular section of the user population is being determined, a between-subjects design should be selected. Since here both *intraindividual* and *interindividual variations* play a role, larger sample sizes are required in order to achieve statistically significant results.

2.3 Test Scenarios

While field testing is intended to capture user behavior in the real driving environment, tests under experimental conditions require the definition of test scenarios for the aspects under study. These should be based on the probability of occurrence during real driving – although reducing the number of scenarios, for example, to the "worst case," is often productive. The Response CoP (Response Consortium 2006) recommends using the following classification of relevant driving situations, with respect to the driver-vehicle-environment triad:

- Use cases, i.e., the situations in which the assistance system is designed to function
- Nonuse cases, i.e., situations in which the assistance system intervenes where it should not, for example, due to limitations of the available sensors ("false positives")

- System limits, either specified (e.g., speed ranges) or functional deficits
- System failures, such as the system aborting, but also failures in the sense of ISO 26262 2011
- Foreseeable misuse
- Expected system reactions which do not occur (“false negatives”)

Based on a complete “list of potentially safety-critical situations” (Response Consortium 2006), the tests should be focused on the most critical situations.

For tests on proving grounds, it may be necessary, for reasons of safety and reproducibility, to use abstracted testing scenarios, even if this could have an effect on the validity of the results. An example is the use of a “crash target” (such as EVITA, cf. ▶ Chap. 13, “Evaluation Concept EVITA”) instead of a braking lead vehicle. The selection and design of particular test scenarios, while considering general requirements such as the safety of the participants, also requires experience in testing, and trial runs should be carried out with both experts and naïve subjects.

2.4 Assessment Parameters and Criteria

Building on the experimental hypotheses and the selected test scenarios, system-specific parameters must be defined, whose measurement allows a quantitative description of the test results and, in the case of controllability tests, defines a clear criterion for passing or failing. A general distinction can be made between objective measures and subjective assessments. In addition, some studies use complex physiological measures or observations of driver behavior (e.g., eye-movement analysis) in order to quantify driver workload or analyze driver behavior in detail.

Objective Measures When assessing the effectiveness and controllability of driver assistance systems, objective measures of the degree of success in dealing with a driving task or a derived measure, e.g., for the criticality of a driving situation, can be used. Objective measures include:

- Vehicle reactions in relation to the environment:
  Longitudinal dynamics, e.g., collisions with leading or trailing vehicles, speed of collision, and remaining distance and time to collision (TTC), following distance behavior
  Lateral dynamics, e.g., collisions with vehicle to the side, distance and TTC, number and size of lane crossing events, number of traffic cones hit, and measures of variance of lateral position
- Driver actions and reactions, e.g., reaction times and strength of reaction on steering wheel and pedals
- System performance in interaction with driver and environment (field validation), e.g.:
  Availability and duration of use

Frequency and effectiveness of warnings and interventions (use cases)
Frequency of false-positive warnings and interventions

Subjective Measures During tests with both experts and nonexperts, some aspects can only be assessed via self-reports of the subjects' sensations and perceptions. Here, standardized questionnaires, rating scales, and interviews can be used. Areas relevant to driver assistance systems are:

- Physiological states, such as drowsiness: for example, the Karolinska Sleepiness Scale (KSS, Akerstedt and Gillberg 1990) was used as an external criterion during the development of ATTENTION ASSIST, since the early phases of drowsiness cannot be detected reliably with behavioral indicators (e.g., eyelid closures).
- Cognitive states, such as mental workload (NASA-TLX, Hart and Staveland 1988, also see De Waard 1996; Wierwille et al. 1996).
- Acceptance or purchase intention for new driver assistance systems (Schierge 2005).
- Controllability of system failures (Neukum and Krüger 2003).
- Perceived safety, criticality of driving situations (Neukum et al. 2008).

In field validation tests, it has proven useful to allow the driver the possibility to directly evaluate the situation by manually triggering a short measurement, including voice recording.

3 Test Settings

In this section, the advantages and disadvantages of different possible test settings are laid out, as summarized in Table 1. Test settings range from tests in driving simulators to field validation tests, with corresponding increasing closeness to real driving, accompanied by decreasing control over test conditions.

Tests in a Driving Simulator Driving simulator tests are particularly suitable for early evaluation of assistance system concepts or for studying the effectiveness of safety systems in safety-critical situations without endangering the participants. They also allow a high degree of reproducibility and control over experimental conditions. The main disadvantages lie in possible limited validity, e.g., due to limited vehicle dynamics or field of view, reduced perception of danger, effects on driving behavior, simulator sickness, and limited test duration (also see ▶ Chap. 9, "Dynamic Driving Simulators").

Tests on Proving Grounds Tests on a proving ground are well suited to the study of driving situations in which the interaction between the driver and the actual vehicle reaction is essential to the validity of the results, but where endangerment of the participants or others must be ruled out. They allow significant control over the test

Table 1 Advantages of test settings

Test setting	Advantages
Driving simulator	Situations under study can be precisely determined and reproduced
	Can be used in early phases of development, particularly for evaluation of concepts
	Large potential for variation of test scenarios (driving situations) and system parameters
	Test participants can be put in critical driving situations without endangerment
Proving ground	Realistic environment: real vehicle without limitations on vehicle dynamics or field of view
	Perceived criticality of the driving situation is hardly reduced
	Minimal endangerment of test participants or others
	Geographical flexibility (tests in different locations)
Tests on public roads	Realistic driving environment
	Realistic driving task
	Learning curve and use of systems can be analyzed
	Greatest geographical flexibility: tests can be carried out in many different locations
	Validation data with high validity: vehicles and systems are tested under realistic conditions
Field validation tests	Realistic driving environment
	Free driving without additional tasks or limitations
	Usage behavior, learning, and first familiarization effects can be analyzed
	Greatest geographical flexibility: tests can be conducted in target markets
	Previously unknown potentially critical situations may be discovered
	Validation data with the highest validity: vehicle and final system are tested under real-life conditions without the influence of a test supervisor

scenarios to be maintained, although complex traffic situations are very difficult to implement. Examples are studies of controllability within Functional Safety.

Tests on Public Roads Driver assistance systems with a low endangerment potential, such as systems which only provide warnings or which have already reached a high degree of maturity and validation, can be studied via tests on public roads, usually with an accompanying test supervisor. Such tests are very close to real-life driving, but test scenarios can only be controlled to a limited degree. Such tests are also often only possible late in the development of the system and may also require special safety equipment when preproduction vehicles are used.

Field Validation Tests (Also Known as Field Operational Tests) While the effectiveness of safety systems can be evaluated first in driving simulator tests and later possibly with the help of actual accident data, validation tests of possible unwanted side effects must be carried out in the field under conditions as close to real life as possible. For example, minimization of false alarm rates of safety systems can only be achieved using appropriate data from field tests. In addition to proving controllability, variation in the driver, driving situation, and environmental conditions should also ensure that no potentially critical situations remain unaccounted for. Field validation tests, which are normally carried out without an accompanying test supervisor, also allow a detailed analysis of driving behavior with an acceptance of new systems. They form the foundation for final approval of the system.

The large effort and costs required to achieve significant test mileage, as well as to analyze large amounts of data, are some of the challenges of field validation tests.

4 Execution and Analysis of Field Validation Tests

Before the start of a field validation test, the required total mileage and geographical variation are defined. Based on these requirements, a suitably sized vehicle fleet is equipped with the near-series assistance systems and data-logging equipment for recording vehicle bus, ECU, and sensor data, as well as data-logging cameras. The vehicles are then driven on public roads.

One possibility is a so-called customer-oriented driving test – here, volunteer employees are randomly selected and provided with a test vehicle, typically for a period of 1 week. Another possibility is vehicle endurance testing, whereby vehicles are driven in shifts by professional drivers. This is particularly suitable for targeting particular driving situations, such as urban driving.

For the configuration of the data-logging equipment, it has proven effective to record two types of measurement in parallel:

Continuous measurements record several hundred of the most important signals over the whole trip. They allow statistical analyses such as speed profiles, warning rates, etc. Evaluation of comfort systems, which generally assist the driver over longer periods, relies particularly on such measurements.

Trigger measurements cover a shorter period, focusing on specified driving situations, and are automatically triggered when particular conditions are fulfilled. Such conditions are derived from the analysis of potentially critical situations. A button for manual triggering allows situations to be recorded where a system intervention was missing (“false negatives”) and also allows direct evaluation of the situation by the driver. Trigger measurements typically use a ring buffer, allowing data both before and after the relevant event to be recorded (e.g., 40 s before and 20 s after). They may contain several 10,000 signals, videos from additional data-logging cameras, and raw sensor data. This last type of data allows the system behavior to be re-simulated with modified functional software releases, for example, in order to prove the effectiveness of developmental measures.

If the data-logging technology allows, it may also be possible to record all relevant vehicle bus, ECU, and raw sensor data for the complete trip. During automated post-processing, equivalent trigger and continuous measurements are extracted from the full measurement for further analysis. This places greater demands on logging and storage, but has the advantage that a complete re-simulation with modified functional software releases over the whole field validation test is possible.

Field validation tests, in particular when several systems are being tested in parallel, sometimes with total driving distances of more than 1,000,000 km, can produce tens of terabytes of measurement data. In order to successfully administer such large datasets and to provide the validation team and the system developers with the required information, a database-driven measurement data management system has been developed at Mercedes-Benz (Tattersall et al. 2012). The measurement data from the test vehicles is processed automatically and saved on central servers, while meta-information is stored in a database. Access to the database and measurement data proceeds via a special user interface.

The focus of the analysis lies on "events" (occurring at a specific point in time), such as system warnings and interventions, driver actions such as strong braking, or system evaluations by the driver. For each event, a "short video" (ca. 10 s) of the traffic situation around the vehicle is generated, supplemented by information on system states and signal graphs. In this way, the validation team can quickly and conveniently identify and analyze the relevant measurements and events and document the results. The database also serves to organize the workflow between the validation team and the system developers. In addition to the technical functional approvals (verification), documentation of the validation results serves as a foundation for the final series approval and sign-off of the system.

5 Example Studies

In this section, examples are given for tests in all the settings mentioned above. The evaluation and validation of safety systems for rear-end crash avoidance is a recurrent theme, supplemented by examples of other types of assistance systems. Awards granted to these active safety systems in the Euro NCAP Ratings (Euro NCAP 2013) starting 2014 are a testament to the topicality of this topic (see also ▶ Chap. 11, "Test Methods for Consumer Protection and Legislation for ADAS").

5.1 Simulator Studies of the Effectiveness of Safety Systems

Many rear-end collisions could be avoided or their severity reduced if the technical and physical deceleration potential was fully utilized. Accident research has shown that many such accidents are due to human factors. Sometimes the driver reacts quickly but too hesitantly, underestimates the deceleration of the leading vehicle, or reacts too late or not at all due to distraction.

As early as 2005, Mercedes-Benz therefore introduced Brake Assist BAS PLUS. With the aid or RADAR-based distance information, the system provides a collision warning and, if necessary, increases the driver's braking force according to the

Fig. 1 Driver support by COLLISION PREVENTION ASSIST PLUS to avoid or mitigate rear-end collisions

Table 2 Test scenarios for evaluating Brake Assist BAS PLUS (RADAR-based braking support)

Nr	Road type	Speed	Initial time headway to leading vehicle	Scenario
1	Highway	130 km/h	1.45–1.55 s	Driving on left lane, vehicle pulls into lane from right, TTC = 2.0 s
2	Highway	130 km/h	1.45–1.55 s	Leading vehicle brakes for 0.7 s with 1 m/s^2 and then increases deceleration to 8.5 m/s^2
3	Rural road	80 km/h	1.45–1.55 s	Leading vehicle brakes for 1.0 s with 1 m/s^2 and then increases deceleration to 9.0 m/s^2

situation. PRE-SAFE® Brake added automatic partial and full braking (in situations where a collision is unavoidable). Today, these functions are available as COLLISION PREVENTION ASSIST PLUS as standard equipment in almost all Mercedes-Benz passenger vehicles (Fig. 1, also see ▶ Chap. 46, "Fundamentals of Collision Protection Systems").

Collision Warning and Adaptive Braking Support In order to evaluate the effectiveness of BAS PLUS in the concept phase, a driving simulator study with 110 nonexpert participants was carried out. Three typical driving situations, which according to accident statistics lead particularly often to rear-end collisions, had to be mastered (see Table 2). In a between-subjects design, the control group drove with only the pneumatic Brake Assist (BAS), while the experimental group drove with Brake Assist BAS PLUS, including collision warning.

In the study, the combination of collision warning and BAS PLUS reduced the accident rate by up to 75 % in comparison to the control group. Even for participants who reacted too late to avoid an accident, the accident severity was considerably reduced: in comparison with the control group, the collision speed was on average 35 % lower.

Automatic Emergency Braking (PRE-SAFE® Brake) In a similar way, the effectiveness of additional automatic partial braking was evaluated in a driving simulator study with 70 participants. The particular challenge for this test concept lay in generating a test scenario which made it probable that the participants would not react to a collision warning. After several not so successful attempts at distraction by means of manual and cognitive secondary tasks (e.g., changing a CD, mental arithmetic), a simple accident scenario on the opposite side of a rural road proved to be a very effective distraction. In a similar way to the rural road situation in the BAS PLUS study, the sudden braking of the leading vehicle occurred at exactly this moment.

The majority of test participants (53 %) reacted so quickly to the optical and acoustic collision warning that the accident could be prevented with BAS PLUS braking support. A further 17 % also avoided the collision, although they only reacted during the automatic partial braking phase. Thirty percent of participants were so distracted that they did not brake in time and a collision occurred. On average, the collision speed was reduced from 45 to 35 km/h with automatic partial braking. The resulting reduction in crash energy of 40 % would have considerably reduced the risk of injury to driver and passengers (Schöneburg et al. 2011). Further test scenarios from concept studies in driving simulators are given in ▶ Chap. 9, "Dynamic Driving Simulators".

5.2 Evaluation of the Controllability of Incorrect Automatic Brake Activation According to ISO 26262

This section describes tests with expert and nonexpert subjects on proving grounds and in the driving simulator for the evaluation of the controllability of incorrect automatic braking (see also Fach et al. 2010). These were carried out in the concept phase of PRE-SAFE® Brake, in order to determine the controllability of incorrect activation and the ASIL (Automotive Safety Integrity Level) according to ISO26262 2011.

The controllability of incorrect interventions *in the system vehicle* was evaluated on the proving ground. While the speed range up to 130 km/h could be studied with normal drivers without endangerment, the controllability up to the system limit of 200 km/h could only be determined with expert drivers. The scenarios for the study were derived from the hazard and risk analysis. Figure 2 shows an example test setup on the proving ground.

Controllability *for following vehicles* was studied in a driving simulator. In order to counteract artifacts such as overcautious time headway, test subjects were guided by colored bars to maintain an appropriate and reproducible time headway to the vehicle in front. The incorrect braking interventions by the vehicle in front occurred with a specified duration at speeds of about 40 and 130 km/h, when the participants were driving within a particular time headway range. The objective measure of controllability was the avoidance of the critical event from the hazard and risk

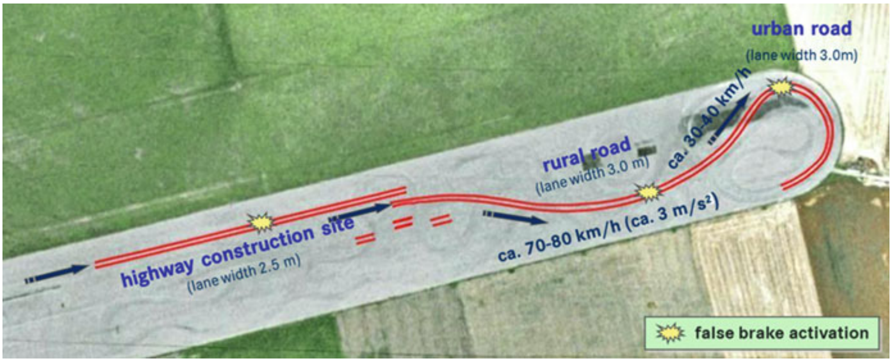


Fig. 2 Setup for testing the controllability of incorrect automatic brake activation on a proving ground (Fach et al. 2010)

analysis (rear-end collision). Studies of controllability of yaw interventions from single-sided braking can be found in Fach et al. (2010) and Simmermacher (2013).

5.3 Evaluation of the Effectiveness of Safety Functions on the Proving Ground

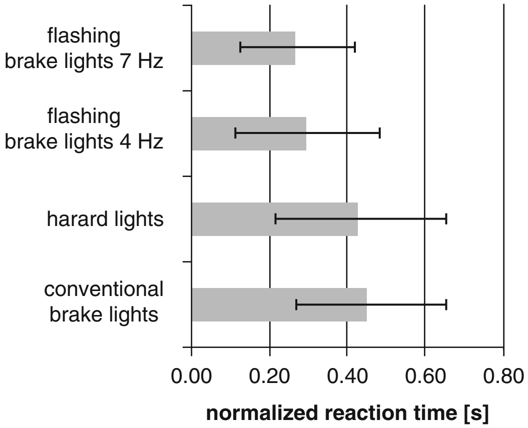
The signal view from the rear during braking can reduce the danger of rear-end collisions by improving the ease of recognition of and reaction to emergency braking by the drivers following. In order to compare the effectiveness of different approaches, tests with nonexpert subjects were carried out on a proving ground, under conditions designed to be as realistic as possible (Unselt and Beier 2003).

Each of the 40 test persons drove at a speed of 80 km/h and a distance of approximately 40 m behind a leading vehicle. This was practiced during a familiarization phase. After several noncritical driving maneuvers, the expert driver in the leading vehicle executed a full emergency braking maneuver. As the primary evaluation criterion, the reaction time between the lighting of the brake lights of the leading vehicle and the activation of the brake pedal by the test subject was determined telemetrically. In order to eliminate the influence of between-subjects variation, the reaction times were normalized according to the individual reaction times determined in five analog braking reaction time tests in a stationary vehicle.

Figure 3 shows that flashing brake lights lead to a significantly faster reaction ($\approx$0.2 s) than conventional brake lights or flashing hazard lights. From the test speed of 80 km/h, the calculated reduction in stopping distance is therefore ca. 4.4 m.

Based on these results, adaptive, flashing brake lights were introduced as standard on Mercedes-Benz passenger cars. The activation level, which varies according to speed, was determined from field tests and takes into account the real braking behavior of vehicle drivers.

Fig. 3 Normalized reaction time (reaction time in test run – mean reaction time in stationary vehicle) to different brake lights of the leading vehicle in an emergency braking maneuver (mean and standard deviation), $n = 40$)



5.4 Evaluation and Optimization of a Safety System for Monitoring Driver Condition in Accompanied Test Drives

Studies have shown that approximately 25 % of all severe road accidents can be attributed to drowsy driving. After the initial evaluation of various approaches to drowsiness detection in driving simulator tests, the ATTENTION ASSIST system was introduced by Mercedes-Benz in 2009 (cf. ▶ Chap. 37, "Driver Condition Detection"). Primarily, the system analyzes changes in the individual steering behavior of the driver, as well as factors such of time of day and trip duration. Therefore, development and optimization of the system could only be carried out with extensive accompanied test drives with a broad test subject collective of more than 550 car drivers.

Before final field validation testing, development versions of the system were evaluated in accompanied night-drive tests under defined conditions (route, time of day, distraction factors). The tests were always carried out with a specially trained test supervisor, equipped with an infrared camera and screen for monitoring the driver, as well as a second set of pedals. In addition to recorded vehicle data, external criteria included physiological measurements and subjective drowsiness ratings based on the Karolinska Sleepiness Scale (Akerstedt and Gillberg 1990).

5.5 Field Validation of RADAR-Based Safety and Comfort Systems

Before introducing new or modified driver assistance systems onto the market, extensive field validation tests with nonexpert drivers are conducted at Mercedes-Benz (Breuer 2009). The tests serve to evaluate and optimize the effectiveness of the systems, demonstrate controllability, and ensure the absence of negative side effects. Table 3 shows example trigger conditions for field validation tests for safety and comfort systems.

Table 3 Safety-related driving situation in field operational tests – examples for trigger conditions and continuous data channels

Potential critical situation	Safety systems: BAS PLUS incl. collision warning, PRE-SAFE® Brake	Comfort system: DISTRONIC PLUS
Use cases	Precritical thresholds exceeded (TTC < x s)	Continuous trip data, e.g., distance-keeping behavior, acceleration/deceleration, situation analysis thresholds exceeded, e.g., time headway < x s, driver override system
	Collision warning activated	
	Adaptive brake support activated	
	Automatic brake intervention activated	
Nonuse cases (false negatives)	“Manual” trigger for missing collision warnings	“Manual” trigger for poor system availability, late reaction, etc.
	Precritical thresholds exceeded (TTC < x s)	
	Indirect: strong braking by driver	
Nonuse cases (false positives)	Evaluation of use cases	“Manual” trigger, continuous transition to system limits, e.g., system decelerates due to vehicle in neighboring lane
	“Manual” trigger, e.g., false collision warnings or automatic brake activations	
System limits and functional deficits	Evaluation of use cases, e.g., warnings which are functionally correct, but “undesirable” from the perspective of the driver	Defined system limits reached, e.g., maximum allowed deceleration
		Warning for driver to take control
		Driver overrides system
		“Manual” trigger
System errors	Fault messages	Fault messages
Misuse	Analysis of the use cases (e.g., provoked warnings/testing the system limits)	Analysis of the use cases
Subjective evaluation	“Manual” trigger questionnaires	“Manual” trigger questionnaires

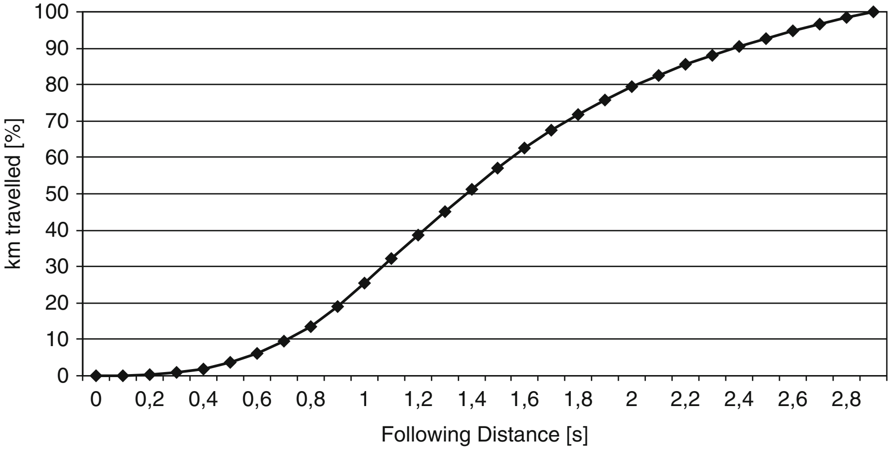


Fig. 4 Time headway using a second-generation ACC system (Breuer and Feldmann 2011)

The following results refer to the analysis of a sample of 936,000 km field data (Germany 79 %, South Africa 10 %, USA 9 %, UK 2 %) from more than 100 drivers (Breuer and Feldmann 2011).

Figure 4 (Breuer and Feldmann 2011) shows the distance-keeping behavior when driving with DISTRONIC PLUS, a second-generation ACC system, based on an analysis of the continuous trip data.

Analysis of a total of 449 collision warnings which were rated as necessary or helpful showed that in more than half the cases, the driver had not yet started braking at the time of the warning. In more than 90 % of situations, the driver started braking at the latest 0.8 s after the start of the warning.

When also considering adaptive brake support (BAS PLUS) and automatic braking (PRE-SAFE® Brake), the frequency distribution by vehicle speed is shown in Fig. 5. While collision warnings and adaptive brake support interventions were also triggered at very high vehicle speeds, automatic brake interventions were only activated at speeds up to about 50 km/h. At typical highway and rural road speeds (70–150 km/h), half of the collision warnings resulted in an activation of adaptive braking support, showing the effectiveness of this combination.

Comparison of the activation frequencies of BAS PLUS and PRE-SAFE® Brake in real driving shows the limited importance of stationary “obstacles.” In addition, adaptive braking support was activated about ten times more frequently than automatic braking. This is a clear indication of the effectiveness of the collision warnings.

Ex post validation studies, based on spare-part orders, have also been used to demonstrate the effectiveness of driver assistance systems for supporting the driver in the avoidance and mitigation of rear-end collisions, while also giving the following traffic better chances of avoiding collisions (Schittenhelm 2013). Analyses of the potential benefits of safety systems and ex post validation on the basis of real accident data are treated in ▶ Chap. 4, “Driver Assistance and Road Safety”.

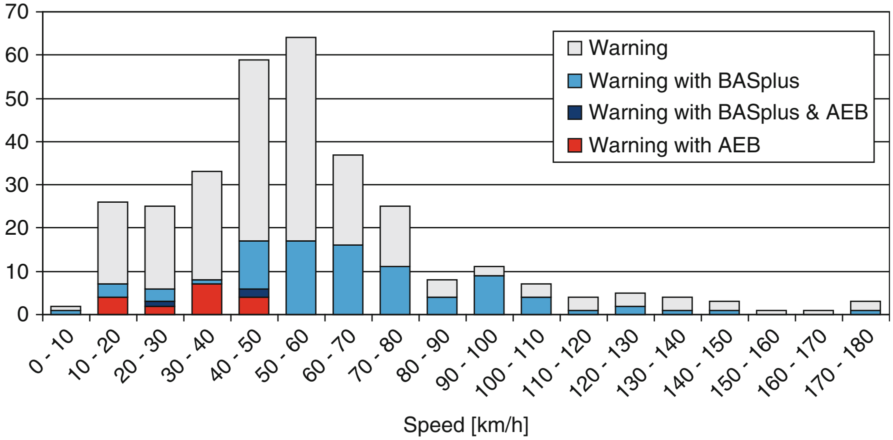


Fig. 5 Activation frequencies of forward collision warning, adaptive brake support, and automatic emergency braking, by vehicle speed (318 situations, Breuer and Feldmann 2011)

References

Akerstedt T, Gillberg M (1990) Subjective and objective sleepiness in the active individual. Int J Neurosci 52:29–37

Bortz J (2005) Lehrbuch der Statistik für Human- und Sozialwissenschaftler (Handbook of statistics for human and social scientists). Springer, Berlin

Breuer J (2007) Fahrerassistenzsysteme: Vom Tempomat bis zum Notbremsassistenten (Driver assistance systems: from cruise control to emergency brake assistance). In: VDA Verband der Deutschen Automobilindustrie (ed) Technischer Kongress 2007. VDA, Frankfurt

Breuer J (2009) Sicherheitsprognosen für neue Assistenzsysteme – Stand und Herausforderungen (Safety prognoses for new driver assistance systems – state of the art and challenges). In: Bruder R, Winner H (eds) 4. Kolloquium Mensch & Fahrzeug. Ergonomia Verlag Stuttgart

Breuer J, Feldmann M (2011) Safety potential of advanced driver assistance systems. In: Pischinger S, Eckstein L (eds) 20. Aachen Colloquium Automobile and Engine Technology 2011. Aachener Kolloquium Fahrzeug- und Motorentechnik GbR. Aachen

Bubb H (2003) Wie viele Probanden braucht man für allgemeine Erkenntnisse aus Fahrversuchen? (How many test subjects are necessary for reliable results in driving tests?) In: Landau K, Winner H. Fahrversuche mit Probanden – Nutzwert und Risiko. Fortschr.-Ber. VDI Reihe 12 Nr. 557. VDI, Düsseldorf

De Waard D (1996) The measurement of drivers' mental workload. PhD thesis, University of Groningen, Haren, University of Groningen, Traffic Research Centre, The Netherlands

Eckstein L (2008) Souveräne Interaktion mit Fahrerassistenzsystemen (Competent interaction with driver assistance systems). In: Technischer Kongress 2008 Verband der Deutschen Automobilindustrie VDA. VDA, Frankfurt

Euro NCAP (2013) Assessment protocol – safety assist. Version 6.0 http://www.euroncap.com/files/Euro-NCAP-Assessment-Protocol—SA—v6.0—0-03d1ee92-316f-4c23-918d-76c7496ba833.pdf. Accessed 30 Oct 2014

Fach M, Baumann F, Breuer J, May A (2010) Bewertung der Beherrschbarkeit von Aktiven Sicherheits- und Fahrerassistenzsystemen an den Funktionsgrenzen. VDI Berichte, Düsseldorf

Hart S, Staveland L (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock P, Meshkati N (eds) Human mental workload. North Holland Press, Amsterdam, p 139–183

ISO 26262 (2011) Road vehicles – functional safety. International Organization for Standardization, Genf

Laurig W, Luttmann A (1988) Planung und Durchführung von Feldstudien (Planning and execution of field studies). In: Rohmert W, Rutenfranz J (Hrsg.): Die Bedeutung von Feldstudien für die Arbeitsphysiologie. Festkolloquium aus Anlass des 75. Geburtstags von Herbert Scholz, Dortmund 10. Juni 1987. Dokumentation Arbeitswissenschaft Bd. 17, Dr. Otto Schmidt, Köln

Neukum A, Krüger H-P (2003) Fahrerreaktionen bei Lenksystemstörungen – Untersuchungsmethodik und Bewertungskriterien (Driver reaction to malfunction of steering systems – test methods and evaluation criteria). VDI-Berichte Nr. 1791, Düsseldorf

Neukum A, Lübbeke T, Krüger H-P, Mayser C, Steinle J (2008) ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen (ACC-Stop&Go: driver behavior at functional system limits). In: Maurer M, Stiller C (eds) 5. Workshop Fahrerassistenzsysteme – FAS 2008. Fmrt, Karlsruhe, p 141–150

RESPONSE Consortium (2006) Code of practice for the design and evaluation of ADAS. RESPONSE 3: a PReVENT Project

Schierge F (2005) Welche Fahrerassistenz wünschen sich die Fahrer? (Which driver assistance systems do drivers desire?) In: VDI-Gesellschaft Fahrzeug- und Verkehrstechnik Der Fahrer im 21. Jahrhundert. VDI-Berichte Nr. 1919. VDI-Verlag, Düsseldorf, p 207–219

Schittenhelm H (2013) Advanced brake assist – real world effectiveness of current implementations. In: ESV Conference. Seoul

Schöneburg R, Baumann K-H, Fehring M (2011) The efficiency of PRE-SAFE®-systems in pre-braked frontal collision situations Paper 11-0207, 22nd International Technical Conference on the Enhanced Safety of Vehicles, 13–16 June, Washington, DC

Simmermacher D (2013) Objektive Beherrschbarkeit von Gierstörungen in Bremsmanövern (Objective controllability of yaw disturbances in brake maneuvers). Dissertation TU Darmstadt

Tattersall S, Petersen U, Breuer J (2012) Ein Messdatenmanagementsystem für die Feldabsicherung von neuen Fahrerassistenzsystemen. (Data management system for field validation testing of new driver assistance systems) In: VDI Wissensforum GmbH (ed) 28. VDI/VW-Gemeinschaftstagung „Fahrerassistenz und Integrierte Sicherheit 2012" VDI Verlag GmbH Düsseldorf

Unselt T, Beier G (2003) Safety Benefits of advanced brake light design. In: Gesellschaft für Arbeitswissenschaft (GfA), International Society for Occupational Ergonomics and Safety (ISOES), Federation of European Ergonomics Societies (FEES): International Ergonomics Conference. Munich, 7th–9th May

Weitzel A, Winner H (2012) Ansatz zur Kontrollierbarkeitsbewertung von Fahrerassistenzsystemen vor dem Hintergrund der ISO 26262 (Approach for the controllability assessment of driver assistance systems with respect to ISO 26262), FAS 2012 – 8. Workshop Fahrerassistenzsysteme. UNI-DAS e.V., Darmstadt

Wierwille W, Tijerina L, Kiger S, Rockwell T, Lauber E, Bittner A (1996) Heavy vehicle driver workload assessment – task 4: review of workload and related research (DOT HS 808 467). National Highway Traffic Safety Administration, Washington, DC

Evaluation Concept EVITA

13

Norbert Fecher, Jens Hoffmann, and Hermann Winner

Contents

Abstract

Within this article, the testing method EVITA (Experimental Vehicle for unexpected Target Approach) for the evaluation of anticollision systems is presented. Using this method, it is possible to generate scenarios in a real test

---

N. Fecher (✉) • H. Winner
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: fecher@fzd.tu-darmstadt.de; winner@fzd.tu-darmstadt.de

J. Hoffmann
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: jens.hoffmann@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_13

drive which are typical for rear-end collisions without endangering the occupants and vehicles. Furthermore, the criterion of speed reduction is explained to evaluate the effectiveness of the system, and finally the latest results are described.

1 The EVITA Dummy Target

There is no universally applicable test method for assessing DAS in critical situations that is completely suitable for use in test drives with real volunteers.

Different versions of collision avoidance systems were developed and evaluated in two joint research projects with Honda R&D Germany and the “Aktiv” [Active] research initiative. For development purposes, we have derived our own evaluation method with a top-down approach.

1.1 Aims

The aim was to develop both a method and a tool for evaluating the performance of anticollision systems in linear traffic flow. One of the requirements was to be able to plot the motion variables of sudden and unexpected braking of the vehicle in front in an otherwise steadily moving line of traffic. The risk to the volunteers arising from the developed test method had to be no greater than in other standard vehicle testing procedures. Another aim in the development of EVITA (Experimental Vehicle for Unexpected Target Approach) was to minimize the influence of the test instrument upon the volunteers, and consequently, great emphasis was placed on keeping the rear view as close as possible to that of a conventional passenger car. As well as facilitating testing with volunteers, keeping the rear view as close as possible to that of a known vehicle means that it could also be used to develop and evaluate sensor concepts for collision warning and avoidance systems.

1.2 Concept

The final concept consists of a combination of a towing vehicle, a trailer, and a following vehicle. When traveling in line with a steady distance being maintained between vehicles, the trailer (known as the dummy target) brakes suddenly, taking the volunteer driving the test vehicle by surprise. Regardless of whether the driver reacts promptly to the maneuver or not, the dummy target is actively pulled out of the collision zone. The dummy target was essentially reengineered in the spring of 2014. EVITA 2.0 is shown in Fig. 1.

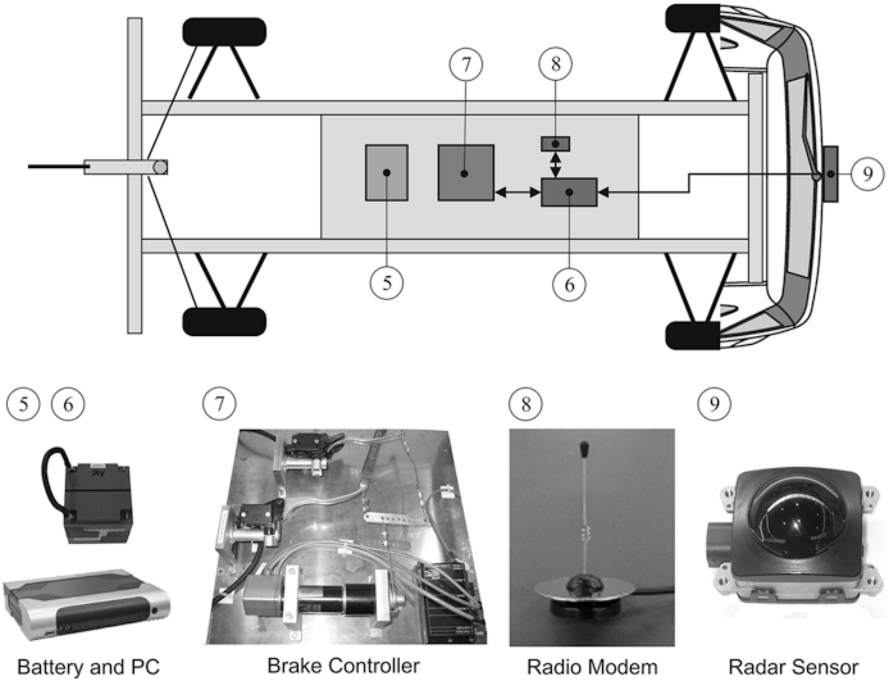
Fig. 1 EVITA (consisting of towing vehicle and dummy target)

1.3 Construction

At the rear of the towing vehicle, there is a rope winch with a friction-lock winch brake and an electric motor. The trailer is only connected to the towing vehicle via the winch rope. The other end of the rope is attached to the Ackerman steering of the trailer's front axle. The trailer's disk brakes are hydraulically operated by an electric motor via the handbrake. At the back of the trailer is the rear end of an Opel Adam. Attached to this rear end is a RADAR sensor. There are computers located in the towing vehicle and in the trailer, and these are interlinked via wireless modems. An aluminum frame from platform technology serves as a basic frame for the dummy target, and this has four individual quad bike wheel suspensions. The large trail of the front axle ensures directional stability. The fanless computer is housed in a moisture-proof casing together with the wireless modem, the energy supply, and the brake control. The brake lights on the rear are functional. The total weight of the dummy target is approx. 200 kg. Figure 2 provides an overview of the dummy target components.

1.4 Test Procedure

On starting out, the trailer is short-coupled behind the towing vehicle. If a RADAR installed on the rear of the trailer and measuring behind it detects a vehicle (target object) at the appropriate test distance, the whole system is activated ready to perform a test. A command from the operator in the towing vehicle releases the rope winch brake and operates the brakes on the trailer. During this process, the towing vehicle continues to travel at constant velocity. Braking of the dummy target causes the winch rope to unwind. While the trailer is slowing down, the processing unit of the distance sensor continuously calculates the time-to-collision (TTC). TTC is a variable formed from distance and relative velocity:



Battery and PC Brake Controller Radio Modem Radar Sensor

Fig. 2 Components of the dummy target

$$TTC = \frac{d}{v_{rel}}; [TTC] = s \tag{1}$$

Here d represents the distance in m to the object in front and v_{rel} represents the relative velocity in m/s. If the TTC is below a specified value, the rope winch brake in the towing vehicle closes and the trailer accelerates toward the towing vehicle that is traveling at constant initial velocity. Acceleration of the trailer up to maximum differential velocity takes approximately 1 s. Once the test is over, the whole combination slows down to a standstill.

1.5 Performance Data

The EVITA performance data are given in Table 1.

2 Measuring Concept in the Test Vehicle

In the selected methodology, the quality of the forward collision prevention measures is measured independently of the EVITA tool. The entire measuring concept for determining the defined evaluation criteria is implemented

Table 1 EVITA performance data

Maximum differential velocity between vehicle approaching from behind and EVITA	50 km/h
Max. braking deceleration of EVITA	9 m/s^2
Shortest TTC before end of a test	0.8 s
Standard test velocity (initial velocity)	50–130 km/h

in the test vehicle, which is fitted with a collision prevention system. Environment sensors classify the EVITA target traveling in front as a relevant target object. Object variables such as distance, relative velocity, and relative acceleration are measured in order to calculate the TTC. A test observer adjusts the settings for controlling the forward collision prevention measures using an operating interface.

The vehicle has a measuring system for the combined capture of CAN and camera data. Three cameras are used. The first camera is directed toward the view in front of the vehicle. In combination with the RADAR data, it provides a reliable interpretation of the situation. The second camera is directed from the combination instrument onto the driver’s face. This helps to monitor where the driver is looking, among other things. The third camera is focused on the vehicle control pedals. It is therefore possible to analyze the driver’s foot movements and to determine action times, for example, transferring from the accelerator pedal to the brake pedal. The repetition rate for each of the three images is 20 ms. The same measuring system records the CAN data to allow the images and signals to be assigned to the correct time. The CAN data include the usual vehicle data such as velocity, lateral and longitudinal acceleration, data about the object in front, and data about the driver’s control operations such as steering wheel angle, footbrake operation, etc.

3 Risks to Test Participants

A system FMEA (failure modes and effects analysis) was performed in order to determine potential system failures, and measures for ensuring safe operation were derived from this. Automated safety check routines are running during each test sequence. If a fault is detected, the system is transferred into a safe and stable state. The level of safety is increased even further by the automated actuation of an emergency brake in the following test vehicle when a TTC of 0.7 s is reached. For the purposes of the tests, the minimum possible TTC set for a collision-avoiding action by EVITA is 0.8 s (see Table 1). If a TTC of less than 0.8 s is reached, it must be assumed that there is a fault with EVITA. If a collision is unavoidable, despite all the precautions, no injuries are expected to the volunteers because of the minimal weight of the dummy target.

4 Evaluation Method

EVITA provides a tool for generating critical accident situations. Below follows a description of the main variables for assessing the quality of collision avoidance systems.

4.1 Effectiveness of a Collision Avoidance System

The reduction in speed of the subject vehicle prior to impact is used as an objective indicator for assessing the effectiveness of a collision avoidance system (especially forward collision prevention measures). This criterion accords with the general aim of collision avoidance systems, which is either to reduce the velocity on impact or to avoid impact altogether. The greater the speed reduction, the more effective the collision avoidance system is deemed to be. In addition to objective effectiveness, the subjective effectiveness perceived by the volunteer driver is also defined. This variable, which is obtained via a questionnaire, is defined as a comparison between different characteristics of forward collision prevention measures, by forming a ranking.

4.2 Test with Volunteers

We know from in-depth studies that a lot of drivers have tried to steer to one side to avoid a rear-end collision (NHTSA Report 2001). The volunteers driving the following test vehicle are therefore induced to take their eyes off the road ahead for more than 2 s by distracting them with a secondary task just before EVITA brakes. The operator sitting in the test vehicle initiates the critical situation while the volunteer driver is looking away. The volunteer driver is then alerted, for example, by the warning elements of the collision avoidance system, when a predefined TTC threshold is reached. Figure 3 shows the idealized velocity curve of the test vehicle plotted over time. We can see the steering maneuver made by the volunteer driver and braking of the dummy target. When the critical threshold is reached, the driver is alerted or some other intervention is initiated. Typically, the volunteer driver then focuses on the situation in front of the subject vehicle and starts to brake.

In the interests of reproducibility, the volunteer driver is shown the reliable distance from the EVITA in front by a traffic-light-type display on the back of EVITA. If the distance is too great, the driver is shown a blue signal and, if it is too short, a red signal. If the distance is between 20 and 25 m, the light shows green. A test is only initiated by the EVITA braking, when the light is green.

4.3 Evaluation Criteria for Forward Collision Warning Measures

An evaluation period is established for assessing effectiveness. This period starts at the point when a warning or a vehicle intervention is initiated and ends at the time of a notional, unbraked impact of the test vehicle against the steadily braking

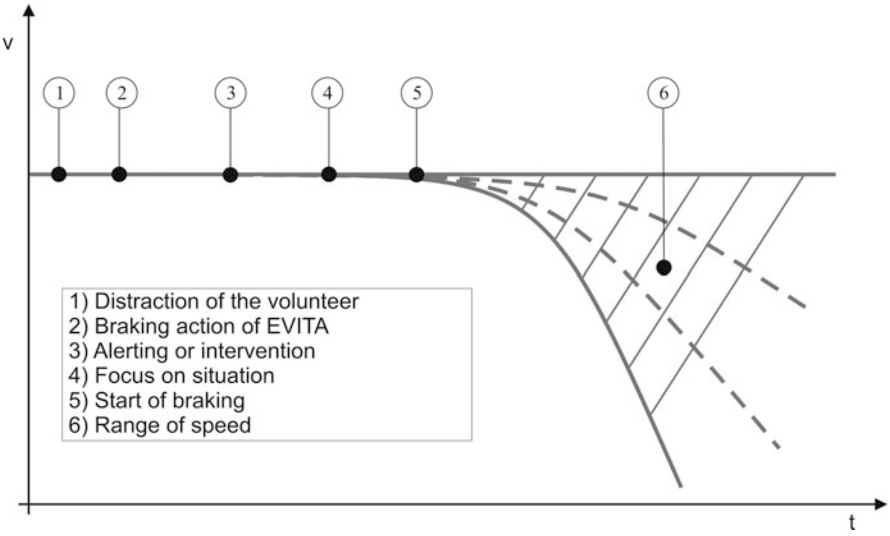


Fig. 3 Idealized test sequence as velocity curve of test vehicle over time

dummy target in front. This collision is "notional" because EVITA automatically prevents an actual collision. The endpoint is determined as a function of the TTC algorithm and the triggering threshold in an unbraked calibration test without a volunteer driver. The evaluation period for a typical warning with the TTC algorithm is 2 s. The warning threshold was defined on the basis of known warning times for forward collision prevention measures. It is then possible to make a cross comparison between warning elements and also to compare them with autonomous braking maneuvers.

The main variable introduced for assessing the quality of forward collisions prevention measures is the reduction in the notional impact velocity, and this is known as effectiveness. For this purpose, an evaluation period is defined from the time of the warning given by the collision prevention system to the time of the notional impact, and at the end of this period, the reduction in differential velocity is determined. In the case of a collision warning system, the driver's response time and the value of the initiated deceleration are the main components of the reduction in differential speed. The overall response time is subdivided into various process steps over the evaluation period.

Literature on the subject provides a lot of information about determining driver behavior in hazard situations, and this is summarized by Bäumler (2008) and Krause et al. (2007). The specification used in this context is based on the definition used by Burckhardt (1985) and/or Zomotor (1987), which is generally applicable to the test conditions. Figure 4 shows the chronological relationship between response times, the evaluation period, the typical velocity curve, and effectiveness.

In the chosen test vehicle, 60 bar corresponds to a deceleration of 10 m/s^2 and therefore to the maximum deceleration at a high friction coefficient between the

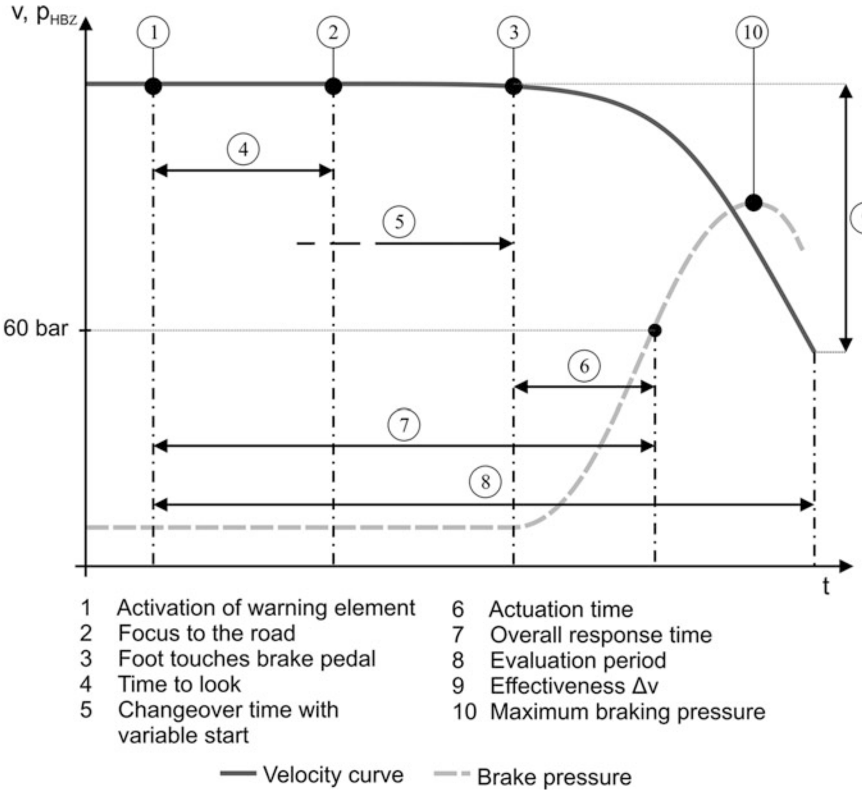


Fig. 4 Definition of the chronological evaluation criteria

road surface and the tires of 1.0. The criteria listed in Table 2 are evaluated during the course of the evaluation period.

4.4 Comparison of Collision Avoidance Systems

A uniform evaluation method forms the basis for comparing different types of forward collision prevention measures. For the purposes of the assessment, test runs are made with an appropriate sub-collective of volunteers looking at different variants. By comparing the average speed reductions for all volunteer drivers over the evaluation period, we obtain the effectiveness of the system variant.

The absolute effectiveness of a collision avoidance system is assessed by using a so-called baseline. To obtain this baseline, some of the volunteer drivers collective are confronted with the critical situation without any intervention by the collision avoidance system, and the differential velocity, for example, is determined.

Only the volunteer's first test can be used as an unbiased basis for assessing the effectiveness of the collision avoidance system. Despite being fully informed in

Table 2 Evaluation criteria during the evaluation period

Objective Effectiveness	Change in speed of subject vehicle
Time to Look	Time taken from time of warning to look at the road
Changeover Time	Time taken from moving foot off the accelerator to first contact with the brake pedal
Actuation Time	Time taken from the foot first contacting the brake pedal to achieving a foot brake pressure of 60 bar
Disturbance	Change in speed of the subject vehicle from start of a false warning where there is no collision risk
Subjective Effectiveness	Tester-assessed descriptor for effectiveness of a warning element in avoiding a collision
Subjective Forgiveness	Tester-assessed descriptor for the "excusability" of a false/unjustified warning

advance about the actual purpose of the tests, after the first test, the volunteer driver is considered to have prior knowledge about the sudden emergency situation and is therefore no longer unbiased. Nowadays, the assessment of driver acceptance is considered to be an important aspect in developing driver assistance systems (Bubb 2003). After the first emergency situation, the subsequent tests are used to obtain additional information, for example, about how false warnings are dealt with or for comparative volunteer rating of different types of collision avoidance system. Test volunteer ratings about the situation they have experienced and their assessment of driver warning elements are obtained by means of questionnaires. Evaluation of these questionnaires provides information about the design of driver warning elements.

4.5 Results

Using the EVITA evaluation tool, various research projects were conducted to investigate the suitability of driver warning elements for use in collision avoidance systems. The results for the following warning elements are presented in this chapter:

1. Tire squeal (*Sound*)
2. Seat vibration with symbol (*Seat Vibration & Symbol*)
3. Braking jerk (*Jerk*)
4. Automated partial braking (*Partial*)

These warning elements are compared in a test with no warning and no intervention (*Baseline*). All warning elements in the test vehicle were activated 2 s before a threatened collision. The auditory icon Sound is provided by a loudspeaker arranged in the center of the dashboard. The volume at the driver's head is 90 dB (A), and the duration is 0.95 s. An electric motor with an imbalance arranged under the center of the driver's seat provides the seat vibration, and a screen installed

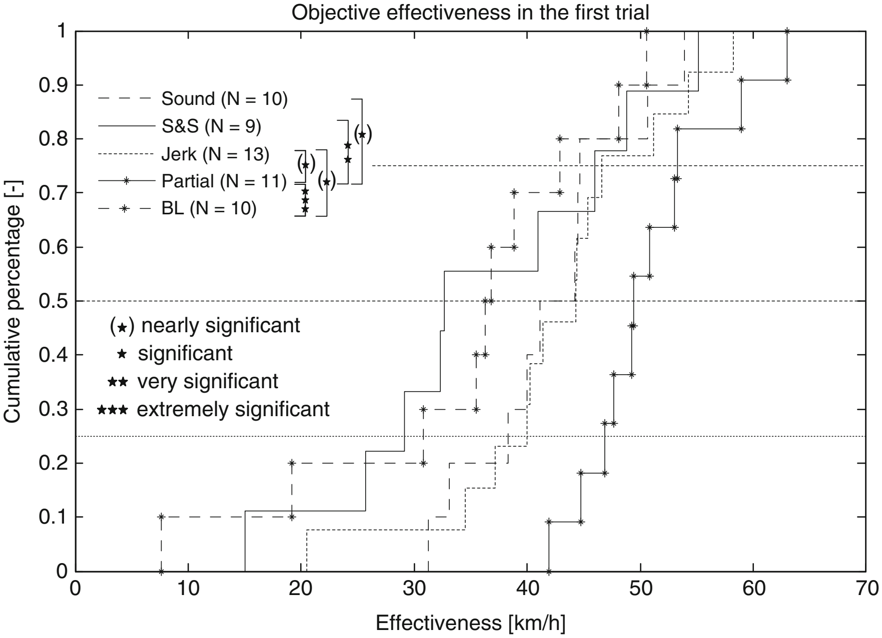


Fig. 5 Effectiveness of warning elements

above the combination instrument displays the flashing red symbol. The symbol size is 75 × 50 mm. The jerk is produced as an acceleration ramp over a period of 0.5 s with a maximum at 5 m/s^2. Automated partial braking is built up at a rate of 6 m/s^2 for a period of up to 1.3 s.

Figure 5 illustrates the effectiveness of the tested warning elements determined in the first test drive. The reduction in speed of the subject vehicle during the evaluation period is determined in the first emergency braking situation.

The cumulative percentage is plotted against effectiveness. The boundaries of the middle 50 % are shown as auxiliary horizontal lines and correspond to the boundaries of a box plot (limit at 25 and 75 %). The letter *N* designates the number of tests. The asterisks describe the significances (this is a measurement for differentiation) between the warning elements. The further the curve is to the right, the more effective the warning element.

We can see that there are differences between the groups *Seat Vibration & Symbol* with *Baseline* as opposed to *Sound* and *Jerk* or as opposed to *Partial*. From a statistical point of view, *Seat Vibration & Symbol* displays no significant difference from a comparative test with no warning (*Baseline*). The curves for *Jerk* and *Sound* are similar, and the hypothesis of equivalence of both distributions cannot be statistically disproved. In terms of effectiveness, *Jerk* displays a nearly significant difference (significance level of 7 % rather than 5 %) against the *Baseline*. *Partial* achieves the highest level of effectiveness with the lowest spread. *Partial* gives an

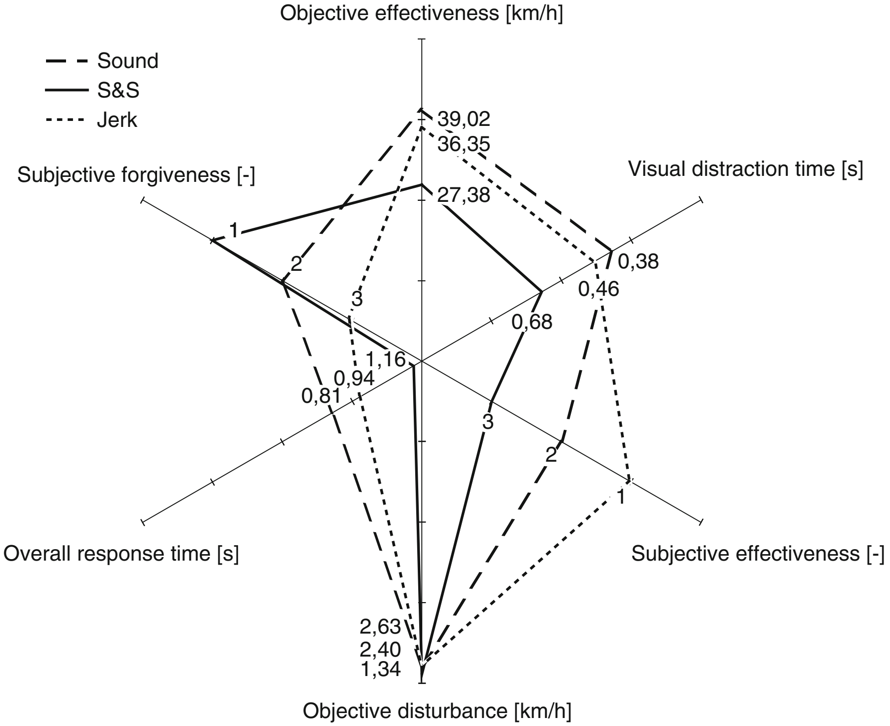


Fig. 6 Features diagram of the warning elements: Sound, Jerk, Seat Vibration & Symbol

effectiveness of 37 km/h, and an even higher value can be achieved with driver override.

A features diagram is useful for providing a comparative picture of different criteria, summarized in Fig. 6, for three of the warning elements from the same test series.

This plots the evaluation criteria of Objective Effectiveness, Visual Distraction Time, Subjective Effectiveness, Objective Disturbance, Overall Response Time, and Subjective Forgiveness as in Table 2. The median is plotted for each warning element. The further a value is from the center (where the axes intersect), the more completely the criterion is fulfilled.

The features diagram can be used to establish weighted assessments in order to produce recommendations. For example, a high weighting can be given to Objective Effectiveness and Subjective Forgiveness. This serves as a recommendation for the warning element *Sound*, because its Objective Effectiveness and Subjective Forgiveness are both higher than for *Jerk*. *Seat Vibration & Symbol* are characterized by lower Effectiveness but higher Forgiveness.

In summary, defining the evaluation criteria is a way of clearly distinguishing between different forward collision prevention measures. Examples of this are the results for the three warning elements: *Sound*, *Jerk*, and *Seat Vibration & Symbol*.

This therefore provides us with an objective assessment, which includes driver perception, to be used in the development of collision avoidance systems.

5 Application in Further Studies

By the end of 2013, extensive tests had been conducted with more than 800 test volunteers. Evaluation of the test structure is very important for translating the findings to the real situation. Evaluation of the tests does not show any obvious peculiarities in the driving behavior of volunteers in normal driving in a moving line of traffic, which could be attributed to the test design. This finding is confirmed by the tester ratings obtained via questionnaires. The aim of avoiding the test design having any negative influence upon the testers has therefore been achieved. Autonomous partial braking is more effective than the baseline by a highly significant margin. Results obtained using the method can be found in Hoffmann (2008), Hoffmann and Winner (2008a, b, c),Winner et al. (2008), and Fecher et al. (2008, 2009).

References

Bäumler H (2008) Reaktionszeiten im Straßenverkehr (Response times in road traffic); VKU (Verkehrsunfall und Fahrzeugtechnik), Vieweg Verlag, Editions 11/2007, 12/2007, 1/2008.

Bubb H (2003) Fahrversuche mit Probanden – Nutzwert und Risiko (Driving tests with volunteers – utility and risk). Darmstädter Kolloquium Mensch & Fahrzeug, Darmstadt

Burckhardt M (1985) Reaktionszeiten bei Notbremsvorgängen (Response times in emergency brake situations), Fahrzeugtechnische Schriftreihe. Verlag TÜV Rheinland, Cologne

Fecher N, Fuchs K, Hoffmann J, Bruder R, Winner H (2008) Analysis of the driver behavior in autonomous emergency hazard braking situations, FISITA 2008, F2008–02–030

Fecher N, Fuchs K, Hoffmann J, Abendroth B, Bruder R, Winner H (2009) Fahrerverhalten bei aktiver Gefahrenbremsung (Driver behavior during active emergency braking). Automobiltechnische Zeitschrift 02/2009, pp. 144ff. (2009)

Hoffmann J (2008) Das Darmstädter Verfahren (EVITA) zum Testen und Bewerten von Frontalkollisionsgegenmaßnahmen (The Darmstadt method (EVITA) for testing and evaluation of anti-collision systems). Fortschritt-Berichte VDI Reihe 12 No. 693, Düsseldorf, ISBN 978-3-18-369312-2

Hoffmann J, Winner H (2008a) EVITA – Die Prüfmethode für Antikollisionssysteme (EVITA – The testing method for anti-collision systems), 5. Workshop Fahrerassistenzsysteme, Walting, April 2008

Hoffmann J, Winner H (2008b) EVITA – Das Untersuchungswerkzeug für Gefahrensituationen (EVITA – The test tool for the investigation of hazard situations), 3. Tagung aktive Sicherheit durch Fahrerassistenz, Garching, April 2008

Hoffmann J, Winner H (2008c) EVITA – The testing method for collision warning and collision avoidance systems, FISITA 2008, F2008–12–019

Krause R, de Vries N, Friebel WC (2007) Mensch und Bremse in Notbremssituationen (Humans and brakes in emergency brake situations). VKU (Verkehrsunfall und Fahrzeugtechnik), June 2007 edition(2007)

NHTSA Report 2001

Winner H, Fecher N, Hoffmann J, Regh F (2008) Bewertung von Frontalkollisionsgegenmaßnahmen – Status Quo (Evaluation of anti-collision systems – Status quo). Integrated Safety, Hanau

Zomotor A (1987) Fahrwerktechnik – Fahrverhalten. Vogel, Würzburg

Testing with Coordinated Automated Vehicles

14

Hans-Peter Schöner and Wolfgang Hurich

Contents

H.-P. Schöner (✉) • W. Hurich
RD/FFS, 059-X820, Daimler AG, Sindelfingen, Germany
e-mail: hans-peter.schoener@daimler.com; wolfgang.hurich@gmail.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_14

Abstract

Many tests of assistance systems can be performed more precisely, much more safely, and more efficiently using coordinated automated vehicles. A testing method that has been developed at Daimler AG to cope with the challenges of testing new assistance systems is presented. The concept for safe operation of automated vehicles on a test track is detailed by describing the technical components and the control strategy of the system. Results from precision and repeatability tests are discussed, and methods of planning tests efficiently and safely are described. The concept of virtual guide rails, which allows a vehicle being tested to act autonomously within a deterministic testing environment, is presented.

In a separate section, the design of automatically driving crash targets is discussed. Two different concepts – the self-driving soft crash target and the over-drivable target carrier – are described in detail. Advantages and limitations of these concepts are discussed.

The chapter concludes with a presentation of several typical test applications of coordinated automated vehicles as single vehicles for better precision and repeatability, as coordinated vehicles to set up complex scenarios that are too difficult to coordinate with human drivers, or as a combination of human-driven vehicles with precisely coordinated crashable targets in challenging traffic situations.

1 Motivation for Testing with Coordinated Vehicles

Driver assistance systems not only support the driver on long journeys by performing routine tasks but also help the driver to react promptly and correctly in critical situations. The very latest assistance systems even react autonomously if the driver does not react in time when an accident becomes unavoidable. For this kind of function, the systems must understand complex traffic situations and differentiate between critical accident situations and uncritical constellations – which is also a challenge for the testing technology with which such systems are validated. In Daimler's research department, a test methodology was developed that allows driver assistance systems to be tested precisely, reproducibly, and safely.

Testing and validating these systems requires a large number of physical tests of the integrated system – despite the increasing numbers of virtual development methods (see ▶ Chap. 8, "Virtual Integration in the Development Process of ADAS"). Quantitative validation requires varying tests over a wide range of parameters. Covering this parameter space completely and as efficiently as possible is a challenge for tests carried out on these systems.

In contrast to testing of vehicle dynamics systems, which react to state variables inside a vehicle, testing driver assistance systems requires the inclusion of additional state variables *outside* the vehicle. For example, the relative position of the vehicle with respect to lane markings is important for testing lane departure warning and prevention systems, and the relative speed and distance of several vehicles to each other is significant for adaptive cruise control systems. If the systems react by supporting the driver's action or if they even act autonomously upon the vehicle's motion, these control systems need to be tested and validated with respect to a large number of different driving situations and environmental constellations to eliminate any risks to passengers and to enable the system to be certified (ISO 2011).

Systematic testing of such systems requires specific driving states of a vehicle to be set precisely on a given test track. Typically, this means controlling a predefined course with a specific speed profile. If several vehicles are involved, all vehicles should move simultaneously in a coordinated manner. Human drivers can control such conditions sufficiently in *one single* vehicle. However, controlling *several* vehicles simultaneously with respect to temporal and spatial specifications means asking too much of human drivers for many constellations. Statistical variation of driving maneuvers might be acceptable or even necessary for some general testing tasks, but for systematic and efficient testing according to precise specifications, for objective comparison of different system variants, and for performing safety-critical maneuvers, it is essential that tests are very precise and can be reproduced exactly.

For specific assistance systems, testing methods have been developed to guarantee comparability (Gulde 2010; Huber and Resch 2008) and/or to improve accuracy and repeatability (Hoffmann and Winner 2008; Ploeg et al. 2008) (see ▶ Chap. 13, "Evaluation Concept EVITA"). There are also some innovative solutions for reducing the risks for drivers performing critical maneuvers (Bock et al. 2007). An analysis of the general applicability of those systems to all relevant accident situations that will have to be controlled by future systems revealed that several tasks could still not be resolved with those systems. The task was still to define and develop a testing system that could manage potentially dangerous maneuvers involving several vehicles in a precise, coordinated, and safe way. Testing with automated coordinated vehicles has proven to do this job as designed since 2009 (Hurich et al. 2009).

Automation and – enabled by its precision – exact coordination of vehicles allows specific improvements to be achieved in the following categories of driving maneuvers (see Fig. 1):

1. Maneuvers that human drivers can barely reproduce, such as swerving maneuvers involving different speeds and distances between the individual vehicles
2. Risky maneuvers that can cause accidents even with small parameter changes, such as close passing maneuvers, which are necessary for fine-tuning and verifying collision avoidance systems on the borderline between intervention and nonintervention
3. Dangerous maneuvers that cannot be performed safely by several human drivers, such as trajectories on a collision course with a high longitudinal difference in speed or even at relatively low speeds with vehicles crossing each other at an intersection.

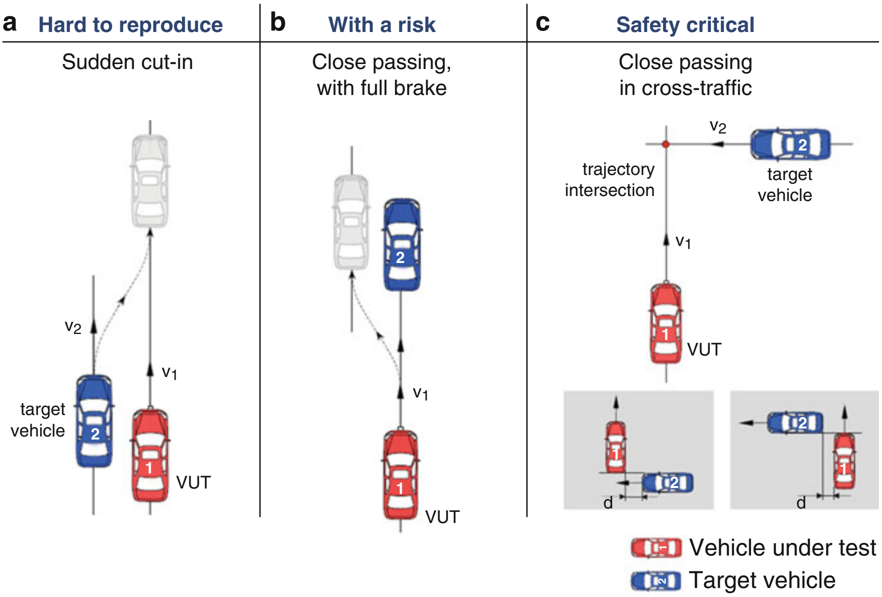


Fig. 1 Maneuver categories with significant improvements through automated driving

2 Requirements for Precision and Reproducibility

From the analysis of future assistance functions to be tested, the accuracy requirements for automated vehicles are deduced. Specifications for lateral and longitudinal accuracy are different. For lateral accuracy, the width of a lane marking (16 cm on rural roads in Germany) is a good reference. Passing an obstacle should therefore be possible at a distance of less than 20 cm. For the precision of a vehicle driving in a lane, this translates into a lateral course tolerance of ±10 cm.

Longitudinal accuracy requirements depend on the speed of the vehicle and are much more difficult to define generally for all tasks. The specifications ultimately required a certain waypoint to be reached within a time tolerance of ±20 ms from nominal timing. This translates into a spatial tolerance for a given time of, for example, 40 cm at a driving speed of 20 m/s (72 km/h).

A *single* vehicle can realize reproducible paths and driving speed profiles simply by maintaining these tolerances. To coordinate *several* vehicles, a common time base is required for all vehicles with a tolerance in the range of 1 ms.

3 Technical Realization

A flexible technical solution for this task was developed in cooperation with Anthony Best Dynamics Ltd., which is the basis for the following description. (Similar solutions from other manufacturers are now also available.)

Fig. 2 Robots control throttle, brake, and steering in the vehicle being tested

3.1 In the Vehicle: Steering Robot, Pedal Robot, Position Measurement, Safety Controller, and Emergency Brake

To replace the driver, actuators are built into the vehicle being tested. The actuators control the throttle, brake pedal, and steering in a similar way to human control actions (see Fig. 2). Such robots have already been in use for some time, mostly for testing vehicles in driving cycles on dynamometers (pedal robots) or for performing challenging steering maneuvers reproducibly (steering robots).

To control the vehicle motion precisely with the goal of reaching a certain location at a given time with a predefined speed and course, it is essential to measure these values exactly. For this task, an inertial navigation system (INS) is installed in the vehicle, which is supported by a differential GPS system. The measurement data is transferred to the vehicle at a rate of 100 Hz. The local differential GPS base station communicates a correction signal every 1 s, which is used to calculate the vehicle's position with a typical accuracy of ±2 cm at 100 Hz. A loss of GPS signal is compensated by the inertial navigation system with an assured precision of better than 10 cm for more than 30 s of GPS outage. This allows the vehicle to pass under bridges without GPS reception without any problems, for example.

Besides the position, differential GPS also supplies a highly accurate time reference signal, which is available equally on all vehicles and is used to synchronize the maneuvers. In particular, all the vehicle maneuvers are started at the same time and/or with a defined time delay based on this reference signal.

The driving maneuvers are controlled by a real-time computer in each vehicle. If a driver is on board the vehicle, he can operate the real-time controller using a connected tablet computer as an interface. This computer stores the test variants and individual control parameters for the respective vehicles and maneuvers. The driver can start and stop single automatic driving sections while sitting in the vehicle, and he can easily regain control of the vehicle at any time.

For driverless operation of the vehicles, an additional safety controller is used, which constantly monitors all the vehicle components to ensure that they are functioning normally as well as its position control and communication with a base station. A spring-loaded emergency brake system and an additional cut-off switch for the vehicle's engine complete the list of safety components that will bring the vehicle to a quick and safe stop in the event of an incident or a system failure.

3.2 In the Base Station: Control Center, Visualization, Coordination, Safety, and Security

From the base station, the automated vehicle tests are started and supervised by a human operator (see Fig. 3). The base station communicates by WLAN with all vehicles within a range of approx. 1 km. A safety controller in the control station permanently exchanges a watchdog signal with the safety controllers in the vehicles. The principal components of the base station are a PC to control the single vehicle controllers remotely, a second PC with separate base station software (which is used to load the maneuvers and control the timing), and a control console for manual remote control (with a steering wheel and pedals for rough positioning of the vehicle).

From the base station, five vehicles can be controlled simultaneously on an enclosed test area. This allows highly complex scenarios to be controlled, which should cover all traffic situations to be tested according to current expectations.

Fig. 3 Remote control of automated vehicles from control station

In the base station, the vehicles and their systems are monitored and test maneuvers are planned, the necessary information is sent to the vehicles, and the vehicle maneuvers are started. With coordinated maneuvers, the control station ensures that all vehicles use the same set of coordinated single maneuvers and assigns the coordinated starting time. During a test maneuver, the trajectories – which include spatial *and* temporal track data – of all vehicles are constantly monitored automatically and visualized for the operator. The integrity and communication of all systems are verified, and if necessary, the test can be stopped at any time by software control from the base station controller or by a hardware emergency stop button. The complete test procedure (including planned stopping trajectories after an emergency stop) is known to each vehicle at the start of a maneuver. Consequently, the relatively unstable wireless communication is not within the control loop and does not mitigate the safety of the test.

A single vehicle can be controlled remotely by hand from the base station, where a video channel shows the vehicle's view of the road ahead. The vehicle can be controlled from the console with a steering wheel and pedals for the throttle and brake. The operator can steer the vehicle at a low speed, even with a significant time lag due to the communication link. This function allows the vehicle to be brought to a suitable starting position for an automatic maneuver, which is especially useful if a maneuver has been interrupted for any reason.

Another task of the operator at the base station is to ensure safety on the test track. A safety concept that considers test track accessibility, the reliability of all technical components, and safety measures to avoid errors while planning or operating the vehicle trajectories is essential for the safe and secure operation of such a system. The requirement for robust communication coverage is the limiting factor for the size of the test area for coordinated automated vehicles. Using dedicated LTE communication channels provides the option of covering distances larger than 1 km.

3.3 Other Systems: Data and Video Transmission, Data Synchronization, and Aerial Photographs

Accurate GPS time signals allow measurement and video data to be documented for all vehicles with synchronous timestamps so that the signals can be evaluated off-line once testing has been concluded. Precisely timed maneuvers also enable cameras or other triggered objects, like traffic lights or pedestrian dummies, to be triggered to operate with the correct timing. Even aerial documentation using camera drones is easy to carry out: the drone flies to the predefined position with an optimal view of the point of interest and takes pictures at precisely the right time. This provides a very good way of documenting close passing maneuvers in particular.

4 Planning Maneuvers

4.1 Planning Individual Trajectories

There are two different options for planning test maneuvers:

1) The test vehicles can be driven by human drivers, even with the steering and pedal robots in place. In learning mode, the control system can record the trajectory driven by the human driver and save it in the maneuver catalog for further repetition. Some parameters, like scaling of speed, lateral shift of the trajectory, and starting time, can be modified when the maneuver is retrieved for a new experiment.
2) For coordinated maneuvers involving several vehicles, planning with graphical planning tools in the maneuver editor is normally more efficient. The trajectory can be compiled from predefined segments (straight tracks, curves, lane changes, sine-curves, etc.) and placed in a certain position on the mapped test track. The whole test can be simulated in advance to check the limitations of vehicle dynamics for the trajectory. Naturally, this necessitates mathematical models describing the properties of the vehicle with sufficient accuracy.

Planned trajectories of all single vehicles can be simulated separately or in combination and visualized in the base station.

4.2 Planning and Examining Coordinated Trajectories

The correct coordination of several vehicles should also be verified by simulation, especially in order to check for minimal vehicle distances. Depending on the maneuver, the distance between two vehicles in an interesting phase of the maneuver can be modified by varying the starting delay between the vehicles. Critical maneuvers involving several vehicles are first simulated for each individual vehicle and then verified by multiple test drives of the single vehicles. The test drive simulations are validated with respect to simulation accuracy and repeatability. The trajectories are then checked for unwanted collisions in combined simulations. This enables even quite close passing maneuvers to be performed safely and reproducibly.

In the case of coordinated vehicles, it must be ensured that the test can be stopped at any time without creating safety-critical situations. For example, a vehicle that brakes due to an emergency stop must not block the trajectory of another vehicle. This can be tested by the simulation tools (Hurich et al. 2009; Pick et al. 2010; Schretter et al. 2009). Each vehicle can be programmed to avoid such situations by predetermining a specific procedure when an emergency stop is triggered: braking immediately, braking with a delay, or even evasive acceleration, steering, and final braking within a defined time and magnitude.

4.3 Precision and Repeatability

The precision and reproducibility of the control system was verified with a position measurement system independently of the GPS measurement. The precision of differential GPS, which is around ±2 cm, could be reproduced consistently for the closed loop control of the trajectory, which involved driving in close proximity to obstacles with a fixed position. Braking tests onto a target with high braking deceleration (e.g., 7.5 m/s^2) show a reproducibility of the stopping point within ±3 cm. For this level of precision, however, the vehicle controller must be tuned to the specific vehicle type, and the trajectory must include a sufficiently long settling time for dynamic deviation. During acceleration and highly dynamic steering maneuvers, temporary deviations in the range of decimeters could be observed. However, these deviations are reproducible when performing the same driving maneuver multiple times and can therefore be accounted for.

Long-term stability of the position control has also been checked by driving to reference points on the test track. Reproducibility of the control over several hours could be clearly demonstrated by repeatedly driving along a pattern in snow over several hours (Schöner et al. 2011a). In total, the required lateral accuracy of ±10 cm can be reached easily. The repeatability of trajectories is much better compared to human drivers.

The longitudinal accuracy depends on the dynamics of the trajectory and on the engine power of the vehicle. During rapid changes in the controlled speed value, a temporary deviation from the reference value is inevitable. The specified precision for reaching a certain waypoint within a time slot of ±20 ms is achievable under the condition of a sufficiently long settling time within the trajectory. The repeatability of trajectories is generally very good, and deviations after steps in reference values are highly reproducible. However, a change in automatic gear shift timing, which generally depends on temperature, may be considered for safety-critical situations and avoided by choosing a suitable speed profile or shift limitations. The control system monitors and shows deviations constantly and can trigger a test shutdown if spatial or timing deviations become too large. Aspects of reproducibility must be considered when planning maneuvers with high precision requirements. The control software allows “critical sections” of the trajectory to be defined, in which particularly small tolerances for position and timing error can be defined.

4.4 Virtual Guide Rails

Coordinated automated vehicles allow software and hardware of autonomously acting driver assistance systems to be tested safely during the development phase. For this purpose, it is important that there is a certain amount of freedom for the system to control the vehicle in a given scenario. A trajectory that is precisely controlled by robots would be counterproductive because a steering or braking action of the robots could be interpreted as corrective driver intervention and thus lead to the assistance function being aborted.

“Virtual guide rails” allow a corridor around the nominal trajectory to be specified for the robot-controlled vehicle being tested. Steering and/or braking intervention of the assistance systems is possible within this corridor without interference by the robot controller. Only if the vehicle is about to leave this corridor because of too little or too much intervention by the assistance systems does the robot system take control and bring the vehicle back on course or to a stop.

5 Self-Driving Targets

Automated vehicles, i.e., vehicles equipped with driving robots as described above, are well suited for collision-free traffic scenarios. However, if a crash mitigation system is to be tested in situations that are not completely controllable, crashable objects should be used as targets. Most crash targets in common use are stationary objects, or they can either drive only in a straight line or are coupled to a drawing vehicle. Collision avoidance tests for intersection scenarios and situations involving turning vehicles in particular are scarcely possible with such targets. Only a completely self-driving target vehicle can solve this problem.

To find a solution that can be seamlessly integrated into the rest of the test environment, a collision target should have the following properties:

- Fully automated, similar to automated real vehicles
- No additional installations in the test environment
- Suitable for crashes from all directions
- With a three-dimensional structure
- The visual appearance of a real vehicle from all sides
- A RADAR signature equivalent to the signature of a real vehicle from any direction

5.1 Soft Crash Target

A first target concept that fulfills these requirements is shown in Fig. 4 (Schöner et al. 2011b). This crashable vehicle consists of a narrow chassis with electric drive and a controller that is compatible with the robot vehicle control as described above, which enables it to drive precisely along predefined trajectories. Its driving performance is sufficient for city driving scenarios as accelerations of around 4 m/s^2 and decelerations up to 8 m/s^2 are possible.

The crash target is built from deformable segments (air cushions) that are kept in shape by pressurized rubber hoses of 10 cm in diameter, which define the edges of the segments. The cushions as a whole are not pressurized, and openings at the lower side allow air to escape during a crash (see Fig. 5). This cushion construction creates a damping effect during crashes, and the differential speed of collision is reduced smoothly with a more or less constant force, providing a maximum crumple zone. The shape of the cushion also easily accommodates the shape of the colliding vehicle. This allows the crash forces to be distributed over a large area

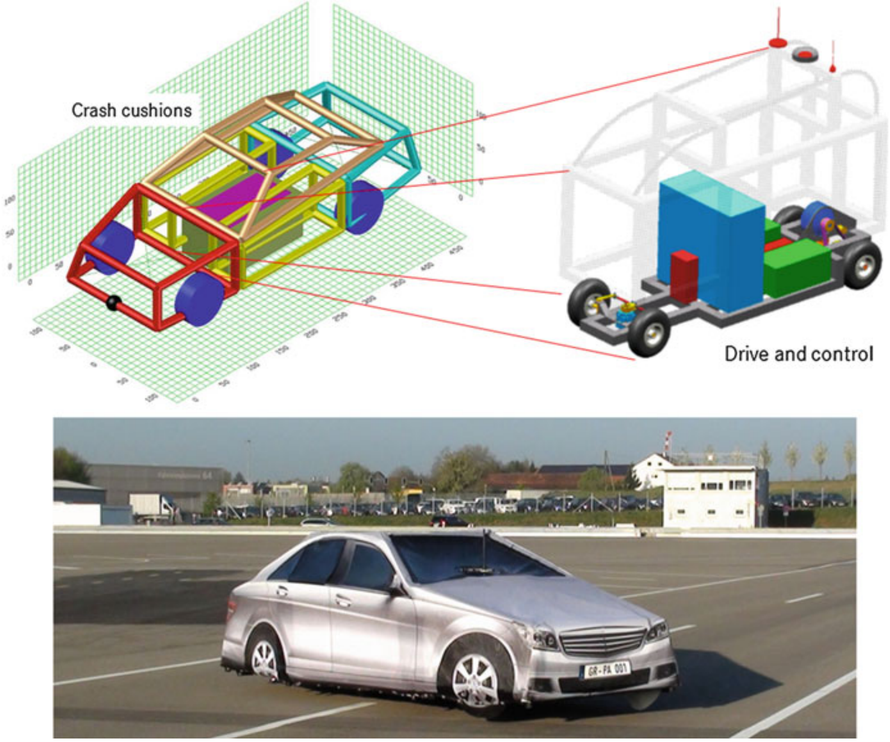


Fig. 4 Concept of the self-driving soft crash target

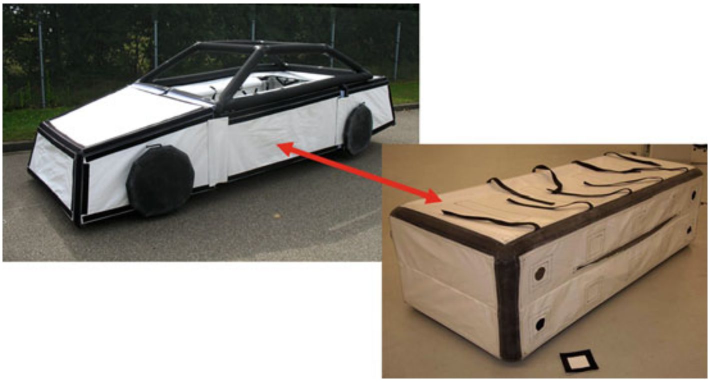


Fig. 5 Crash cushion of the soft crash target with absorber characteristics

and over the whole duration of the crash and thus minimized. Measurements show that this leads to substantially smaller peak crash forces compared to a conventional pressurized “balloon car”with less chance of damage to the vehicle being tested. The pressurized rubber hoses ensure that the target regains its shape independently after a crash.

The self-driving soft crash target is suited to testing vehicles with a driving collision partner, especially during turns at intersections. The target is configured for similar RADAR reflectivity to that of a real car by metal foils, and its shape closely resembles that of a real car. This allows it to be used for RADAR-based, laser-scanner-based, and camera-based assistance systems. The free choice of trajectories and the vehicle-like appearance from all directions make it suitable for many different traffic scenarios. Differential speeds during a crash of up to 50 km/h in a longitudinal direction and up to 30 km/h in a lateral direction are possible without any risk of damaging the vehicle being tested or harming the target itself. The control electronics and the electrical system are set up and tested for high acceleration during such crashes.

5.2 Over-Drivable Target Carrier

If tests of the system vehicle are required with even higher differential speeds or if the risk of damage has to be even lower, the concept of a self-driving over-drivable target carrier can be applied (see Fig. 6). Based on a suggestion from Daimler AG, this concept was developed by DSD in Austria (Steffan et al. 2012). A similar

Fig. 6 Over-drivable, self-driving target carrier with rear-end target

concept was developed independently and more or less at the same time by Dynamic Research, Inc., in California (Zellner 2011; ABD 2013).

The target carrier can be driven over if a collision takes place, and the target itself, which is fixed on top of the target carrier, is pushed away by the colliding vehicle. High differential speeds are only relevant for longitudinal crashes. For such maneuvers, it is sufficient to use only a front or rear target instead of a complete vehicle mockup on the target carrier. For example, the front or rear portion of the soft crash target can serve this purpose. With this concept, differential speeds of more than 100 km/h during a crash have been performed without damaging the vehicle or crash target. The target carrier can also be used for carrying pedestrian or bicycle dummies as targets (see ▶ Chap. 11, "Test Methods for Consumer Protection and Legislation for ADAS").

Due to the low height of the target carrier vehicle, ground clearance and maximum driving power of the target carrier are limited. For this reason, the target carrier is less flexible for applications than the soft crash target. The ability to drive over the target carrier works only if the vehicle crashes into the target. The target carrier may not drive from the side into the wheels of the test vehicle because it would hit the wheels very hard. Therefore, the target carrier should only be used for lateral crashes in which the vehicle and target are precisely coordinated. If the target carrier is used in conjunction with human-driven test vehicles for lateral crashes, it is essential to allow for the timing error of a human driver to ensure the correct collision configuration.

6 Examples of Automated Driving Maneuvers

6.1 Automatic Maneuvers of Single Vehicles

Automatic maneuvers have advantages even for tests with a *single* vehicle:

Driving along curves and in circles: Some assistance systems are activated when a specified lateral acceleration is reached. An automatically driven maneuver with a predefined speed and track curvature can provide this lateral acceleration very precisely and, for example, with a continuously increasing value. Such maneuvers can be performed much more efficiently than with a human driver.

Misuse and durability: For standardized testing of airbag systems, several situations with very high lateral and longitudinal accelerations are required; only under specific conditions may an airbag be deployed. For these tests, vehicles must be driven across a ramp at up to 70 km/h, perform a full brake crossing a curb, or ram a simulated boar. Durability testing is another example with extreme stress on the vehicle and on the test driver. Such maneuvers can now be performed using automated vehicles, which avoids subjecting the test driver's health to unnecessary stress.

Fig. 7 Close passing within intersection at 70 km/h

6.2 Coordinated Maneuvers with Several Driverless Vehicles

In maneuvers with several participating vehicles, the benefits of coordinated automated vehicles become fully apparent. Much of the testing of collision avoidance or collision mitigation systems involves verifying performance in situations that closely resemble the specified use case. In many of them, the assistance system must show that it does not act unintentionally so that incorrect emergency brake situations can be avoided reliably. For such situations, close passing is required, and often with evasion maneuvers at the last moment.

With automated vehicles, such tests can be performed precisely, reproducibly, and safely for both driver and vehicle. Parameters like initial speed, smallest distance, and course angle can be controlled easily. A new software version of the system being tested or a new sensor variant can be tested in exactly the same maneuver as before, thus enabling efficient and reproducible results of the tests.

The greatest challenge for coordinating two vehicles is close crossing at an intersection. (A cross-traffic assistance system has been available in Mercedes vehicles equipped with “intelligent drive” since 2013). Crossing at intersections puts the highest requirements on spatial and timing precision of the test. (By comparison, spatial precision is sufficient for safe operation in situations with longitudinal traffic like passing vehicles or oncoming traffic). Any fault in the test procedure also presents a significant risk of considerable damage. For these reasons, close passing maneuvers in cross-traffic situations may not be performed with human drivers on board. With coordinated automated vehicles, cross-traffic situations with vehicles driving at up to 70 km/h and minimum distances of less than 1 m could be performed repeatedly and safely (see Fig. 7).

Fig. 8 Cut-in maneuver with coordinated vehicle and target

6.3 Maneuver with Driver, with Triggered and/or Synchronized Targets

A test variant with a coordinated vehicle involves driving only part of the maneuver automatically, especially a portion after a specific trigger point, which needs very precise execution. This is essential for tests in which crashable targets perform a critical maneuver, but the test vehicle is driven by a human driver to test the actions and reactions of different drivers. For this variant, no steering or pedal robot is required in the vehicle being tested; only exact position measurement and a communication link to the other vehicle or to the base station are required. Depending on the relative position or other trigger conditions of the vehicle being tested, a lane change maneuver of the crash target can be initiated, for example. Before the trigger point, the target drives along its trajectory on the adjacent lane (see Fig. 8).

This method allows a coordinated self-driving crash target to be used with very little planning outlay for precise repeatable maneuvers with human drivers.

7 Future Developments

In the future, testing of highly and fully automated vehicles must be performed on automotive test tracks. Coordinated automated vehicles and the procedures and equipment described above can contribute significantly to such tests. Such vehicles must prove that they can cope with any conceivable situation without incident. Systematically testing many variants of critical situations can only be performed efficiently with automated vehicles and will be required for thorough testing. It can be foreseen that the complexity of driving situations will further increase and that coordinating more than two vehicles under test conditions will be necessary. Robust, safe, and efficient testing technology with automated maneuvers will be one of the challenges for the development of future vehicles.

References

ABD (2013) ADAS soft targets. http://www.abd.uk.com/en/adas_soft_targets/abd_dri_guided_soft_target_vehicle. Accessed 1 Mar 2015

Bock T, Maurer M, Färber G (2007) Vehicle in the Loop (VIL) – a new simulator set-up for testing advanced driving assistance systems. In: Driving simulation conference, Iowa City

Gulde D (2010) So testet AMS, Teil 9: Assistenzsysteme (Testing of driver assistance systems at AMS). Auto-Motor-Sport

Hoffmann J, Winner H (2008) EVITA – das Untersuchungswerkzeug für Gefahrensituationen (EVITA – testing tool for dangerous situations). In: 3rd Conference on active safety, München

Huber B, Resch S (2008) Methods for testing of driver assistance systems. SAE Paper 2008-28-0020

Hurich W, Luther J, Schöner HP (2009) Koordiniertes Automatisiertes Fahren zum Entwicklen, Prüfen und Absichern von Assistenzsystemen (Coordinated automated driving for development and test of driver assistance systems). In: 10th AAET symposium, Braunschweig

ISO 26262 (2011) (DIN-ISO 61508): Road vehicles – functional safety

Pick AJ, Hubbard MJ, Neads SJ (2010) Near-miss collisions using coordinated robot-controlled vehicles. In: Proceedings of the 10th international symposium on advanced vehicle control, Loughborough, pp 634–639

Ploeg J, Hendriks FM, Schouten NJ (2008) Towards nondestructive testing of pre-crash systems in a HIL setup. In: IEEE intelligent vehicles symposium IV, 4–6 June 2008, Eindhoven

Schöner HP, Hurich W, Luther J, Herrtwich RG (2011a) Koordiniertes Automatisiertes Fahren für die Erprobung von Assistenzsystemen (Coordinated automated driving for testing of driver assistance systems). Automobiltechnische Zeitschrift ATZ

Schöner HP, Hurich W, Haaf D (2011b) Selbstfahrendes Soft Crash Target zur Erprobung von Assistenzsystemen (Self-driving soft crash target for testing of driver assistance systems). In: 12th AAET symposium, Braunschweig

Schretter N, Sinz W, Schöner HP (2009) Planung und Realisierung von automatisierten Fahrmanövern zur Erprobung von aktiven Sicherheitssystemen (Planning and realization of automated driving maneuvers for test of driver assistance systems). In: 3rd symposium virtual vehicle, Graz

Steffan H, Moser A, Ebner J, Sinz W (2012) UFO – ein neues System zur Evaluierung von Assistenzsystemen (UFO – a new system for the evaluation of driver assistance systems). In: 13th AAET symposium, Braunschweig

Zellner JW (2011) Guided soft target. http://www.dynres.com/prod_guidedtarget.html. Accessed 1 Mar 2015

Part IV

Sensors for DAS

Vehicle Dynamics Sensors for DAS

15

Matthias Mörbe

Contents

Abstract

Driver assistance systems require fundamental information from sensors. In order to meet the high standard of safety and availability of these systems, the product quality is essential. Therefore, the selection process plays an important role during the purchasing process in the automotive industry. The surrounding and specific application defines the sensor design and manufacturing process of the sensor.

For advanced system functions, the requirements of sensor data content and precision are increasing continuously; however, there are still some values which have not changed since the first ABS was introduced into the market. Even though the function remains relatively unchanged, the technology itself is changing. This opens up the possibility for the introduction of new features.

M. Mörbe (✉)
Bosch Engineering GmbH (BEG-SC/PJ-T), Heilbronn, Germany
e-mail: matthias.moerbe@de.bosch.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_15

This chapter provides a general overview of the key figures related to sensors for safety systems with details to the design and with special considerations to the fitment in the vehicle.

1 Introduction

In many areas, a sensor component for a driver assistance system is selected regardless of its function. Rather, the conditions are based on VDA or ISO standards introduced worldwide in the automotive industry and on the rules that system suppliers and vehicle manufacturers have established for themselves.

These standards are regarded as the fundamental basis for the level of quality achieved today. In addition to its importance for the availability of the systems and the vehicle in the narrow sense, in many cases, sensor quality is also vitally important to the safety of the system as a whole. The cost and effectiveness of monitoring the sensor signals depend on this quality.

Sensors in the motor vehicle are not an end in themselves; they supply essential information for driver assistance systems. As the cost of these systems is a decisive factor for their acceptance on the market, both the cost and number of sensors must be reduced to the bare minimum.

A sensor is selected for a system based on the following two key aspects:

- General selection criteria that apply to any sensor
- Technical data for the function in question

The information in this chapter is intended to explain what to take into consideration during this selection process. I leave an in-depth treatise on the subject to specialist literature and internal company documentation. The sensor data contained herein has been taken from current documents for vehicle manufacturers.

My special thanks go to all colleagues who assisted me with this presentation.

2 General Selection Criteria

For the selection process, it is recommended to systematically order the different sensor requirements in a matrix, using the same layout and method for each supplier. This considerably facilitates the comparison of different offers. A model of this selection matrix, with a technical level and a commercial level, is shown in Fig. 1. The content can be added to as necessary.

During the first selection stage, it is a good idea to refrain from weighting the individual factors. This ensures that each criterion is treated with the same care and attention. If two offers are very similar in the final selection, however, weighting can provide greater transparency. In addition to factors that are simple to measure, there are of course a great number of the so-called “soft” factors. These include

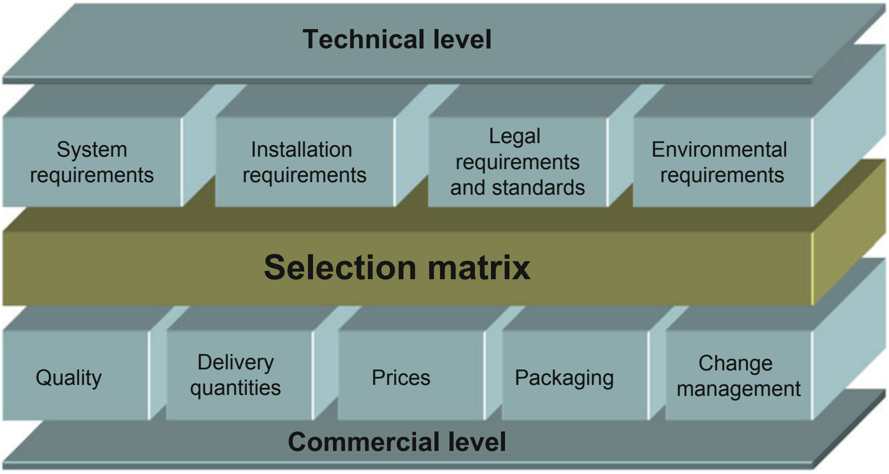


Fig. 1 Selection matrix, sensor components for driver assistance systems

reliability as reported verbally, trust in compliance with confidentiality agreements, fast reaction times for quality issues, and – if necessary – willingness to enter into a long-term cooperative relationship.

2.1 Requirements on the Technical Level

The requirements for a sensor in the motor vehicle are divided into four main categories:

- System requirements
- Installation requirements/geometry
- Environmental requirements
- Legal requirements and standards

The requirements are documented in specifications by vehicle or system manufacturers and are subject to a documented change service. This constant influx of changes is the result of technological advances, and compliance with these changes must be verified again before each component selection. Many years of observation of this process have clearly shown that noncompliance with changes is a significant cause of subsequent complaints. However, the truth is that it is not just compliance with a change, but the analysis and assessment of this change – including in relation to its interaction with functions and further requirements in other areas or systems – that counts. A highly suitable method for systematically identifying this interaction is the DRBFM (Design Review Based on Failure Mode) method developed by Toyota.

System Requirements System requirements are divided into physical quantities resulting from the conversion of the measured variable, the electrical interfaces, and the functional description in the context of the system. As a rule, clearly measurable parameters can be determined for signal conversion. Each parameter is additionally presented with tolerances, resolutions, and accuracy, in context with other requirements. I must emphasize the importance of clear measurability, because this has a major influence on the cost of tests during production that are part of the delivery conditions. The definition of variables that can only be described during further processing of the signal in the system must remain the exception to the rule.

For many applications, standards have already been defined for the electrical interfaces. These are intended to ensure that a sensor from one supplier could also be provided by another supplier. However, reduction to these standards permits this simple exchange only in a few isolated instances. This is due to the additional conditions based on system function, which cannot simply be reduced to electrical or mechanical quantities. For safety systems, the development process also needs to be assessed. Important fundamental assessment criteria for the FMEA (failure mode effect analysis) and an FTA (fault tree analysis) are derived from the methodology, the depth of simulations performed, and manufacturing process developments.

The distinction between static and dynamic interactions also adds to the complexity of this subject. This complexity is so great that sensor characteristics not described in the specification may have unknowingly been taken into consideration in a system's basic development. If the sensor changes, e.g., in the technology used, as the system evolves, complications with dramatic effects may arise only very late in the development process.

As a result, specifications are becoming ever more complicated, and the need for expert knowledge is growing. Expert knowledge must be used to observe the meaning of parameters and their tolerances in relation to system function. Simulation provides assistance here, but this also has certain limits. So far, dynamic processes in the vehicle and its electrical system could only be simulated in the system as a whole to a limited extent. As long as sensor and electrical system models are not sufficiently refined, checking compliance with system requirements using actual hardware and software remains unavoidable.

Installation Requirements As the level of equipment of vehicle dynamics control (VDC) systems in the widest sense grows, more and more requirements for sensor installation are being included in the vehicle's design specifications. Given the number and diversity of vehicle shapes and sizes, the number and diversity of installation conditions for suitable sensors are increasing accordingly. A uniform trend is discernible where size is concerned: the smaller and more well adapted a sensor's design, the better for the automotive design engineer.

The limits are set based on handling during vehicle production and service. This is particularly the case for fastening using bolts or screws and the accessibility of electrical plug-in connections.

However, the installation location is the source of the most significant interactions between sensor and vehicle. The greatest error when assessing the suitability of a sensor can arise if the installation location is regarded as a static, stable parameter. The vibrations that occur vary in their frequency spectrum, amplitude, and resonant rise, depending on the dynamics of driving conditions. But the handling of the vehicle itself is also reflected in this profile. An incomplete list without prioritization by importance clarifies what needs to be borne in mind.

Disturbing factors at the sensor installation location:

- Change in the air gap of the wheel speed sensor due to axle load and bearing play with the steering wheel at full lock
- Jolts due to slamming doors
- Vibrations due to handbrake activation
- Jolts due to seat adjustments
- Pulses due to water passing through at high speeds
- Driving over lane markings on race tracks
- Pulses due to stone chips on dirt roads
- Splash water in the engine compartment when driving through water
- Jolts and knocks due to tools during vehicle assembly
- Uneven road surfaces during different driving maneuvers
- Road surfaces and tire properties
- Knocks/jolts from unsecured objects in the footwell
- Right-/left-hand drive versions
- Additionally installed high powered audio systems
- Cell phones positioned in locations not intended for this purpose
- Temporary external magnetic fields

Additional items can be added to this list for each sensor, and each individual situation reflects the experience of the sensor manufacturer, which it has taken into consideration in its design. Synthetic tests can only be performed to a limited extent, because the excitation energy cannot be generated and introduced in all cases. Moreover, changes due to the aging of the vehicle occur, the course of which cannot always be filtered out for the individual disturbing factor.

In addition to the abovementioned wealth of underlying conditions at the installation location, modifications to the vehicle during the course of its development are also a factor. As a rule, only a masked prototype of the vehicle based on a predecessor model is available at the time decisions are made regarding the system and hence the sensor. However, the body and axle characteristics of the series-produced version are important for reliable sensor operation. A list of influencing factors will show what should be noted when assessing the installation location.

Factors for assessing the installation location:

- Panel thicknesses
- Beading, indentations, tension zones around punched/bent parts
- Weight of further fastening and mounted parts, e.g., seats

- Apertures
- Carpets and insulating materials
- Wheelbases and different configurations
- Two-door/four-door/station wagon versions
- Automatic/manual transmission
- Right-/left-hand drive versions
- Selected wheel bearings
- Types of suspension
- Spring/damper rate of chassis
- Engine versions
- Engine vibrations in motorcycles

Special attention must also be paid to sensor applications in which the vehicle manufacturer is employing identical parts for several platforms. In this case, responsibility for suitability at installation locations that have not been assessed must be borne by the company using the sensor.

The multidimensional factors at the installation location mean that the use of sensors in motor vehicles entails considerable expenditure. These costs must be taken into consideration in the selection matrix.

Legal Requirements and Standards These requirements on the technical level are divided into demands on function, stemming from the system, and requirements for the materials used.

In the future, systems will be classified according to an ISO 26262 safety standard. This will also give rise to requirements for sensors in process chains, affecting both the technology and the development process. This standard is specifically adapted for motor vehicles in the form of ISO 26262.

Safety requirements may also change due to the multiple use of sensor signals by different systems. If a steering angle sensor or yaw rate sensor, for example, is used not only by the ESP system but also by a superimposed steering or rear axle steering system, the safety requirements for signals from these sensors will become more exacting. The result may be that all signal processing in the sensor must be redundant. If these steering systems are not planned as 100 % equipment, a simplified, less expensive version without redundancy may be required for the vehicle series.

Under the umbrella term “environmentally friendly design,” materials that constitute a potential risk in automotive engineering are increasingly restricted or prohibited outright. Using material data sheets, component manufacturers and their suppliers must attest to what materials and quantities the components contain. Auxiliary materials employed during production are also subject to this rule, insofar as the substances they contain will also be present in the delivered product.

Here, it is irrelevant whether these now prohibited materials are distributed or diluted to such extent that they are below the detection limit in the finished product. Since the standards are amended annually, before bringing products onto the market, it is necessary to check whether this is still permissible under the new laws. It is the duty of the supplier to demonstrate freedom from pollutants. Special rules exist, e.g., for spare parts for older vehicles.

In order to check whether the requirements arising from legal regulations and standards are satisfied, both wide-ranging knowledge of the entire vehicle system and in-depth knowledge of the production processes of the components and assemblies used are required.

The material data sheet must be a fixed part of every sensor component quotation. The data are ascertained using sophisticated procedures, and this also determines supplier selection.

Environmental Requirements This term is understood to cover all climatic and dynamic requirements in relation to operation in the vehicle.

Standard ISO 16750 and the specifications of vehicle manufacturers describe the loads to which the sensor is exposed at the installation location. The aim is to ensure error-free operation over the systems' entire operating time and service life. Checking these specifications and implementing them in an electronic circuit, selected components and an electromechanical construction subject the design engineer to the most stringent demands. Satisfying each and every parameter under all conditions is not economically viable and also does not make technical sense. Rather, the design engineer must regard each parameter in its real context of vehicle operation. This observation must cover all the interactions of the installation location with its ambient conditions.

This example provides further clarification: a maximum temperature applies for a wheel speed sensor. This value is generated by extreme brake temperatures and continues rising even in the residual heating phase, when the vehicle is stationary. A temperature shock may occur if the vehicle subsequently drives through dirty water, which may also contain deicing salt. If this scenario was not reconstructed in a laboratory test, a number of underlying conditions have not been fully covered. The sensor head is installed in a subframe and therefore protected from heat radiating from the brake disk. The large mass of the subframe permits only a slow change in temperature due to its thermal capacity. This is true in both directions. Whether or not the sensor head is wetted with dirty water as the vehicle drives through depends on how the sensor is positioned on the axle. Finally, the question remains as to the frequency of this maneuver: how many vehicles will be run under these conditions, and how frequently does such extreme braking occur? This example purposefully omits in-depth numerical detail, because it is intended simply to provide an idea of the complexity of these relationships.

Special cases are missing covers, damage to cables and connections, bolts not tightened to the specified torque, spare parts with different specifications, and use in motorsport.

2.2 Commercial Level

In supply agreements, compliance with all functional requirements on a technical level is assumed, on the basis of technical documentation and a set of drawings.

Deviations must be documented in special lists. These deviations are usually agreed for a series, a quantity, or a period of time.

The principal issues at this level are:

- Quality
- Delivery quantities
- International supplies
- Subsequent deliveries
- Packaging
- Change management

The design and selected technologies determine the subject matter of the contracts agreed on this level to a considerable extent. With the customer's purchasing department, the design engineer must identify the particular underlying conditions and set out agreements.

3 Technical Sensor Characteristics for Driver Assistance Systems

3.1 Sensors and Installation Locations

Main functions of signals in the systems: the designation of the main axes in the direction of travel as x, the transverse direction as y, and the vertical axis as z is internationally standardized; see Fig. 2. This ensures that sensor labels and parameter definitions do not result in different assessments.

Wheel Speed For all vehicle dynamics systems, wheel motion is the variable that determines wheel velocity, acceleration, and direction of rotation. This is used to calculate the coefficient of friction or wheel slip and also the vehicle speed. The difference in speed between the front and rear wheels in motorcycles is a control variable of traction control. The dynamics of wheel motion is the most important variable for controlling vehicle deceleration and driving stability on all surfaces.

Steering Angle For controlling vehicle stability, the steering angle as input information of the driver's intention is the measured variable on which all measured values relating to vehicle dynamics are based and checked for plausibility. The steering angle is not measured on the wheel.

Yaw Rate Signal Rotational movements on all three planes are measured in order to determine the dynamics of the vehicle body. Movement around the z-axis is measured for ESP systems, the rolling motion around the x-axis for rollover detection, and the pitching motion around the y-axis for chassis control.

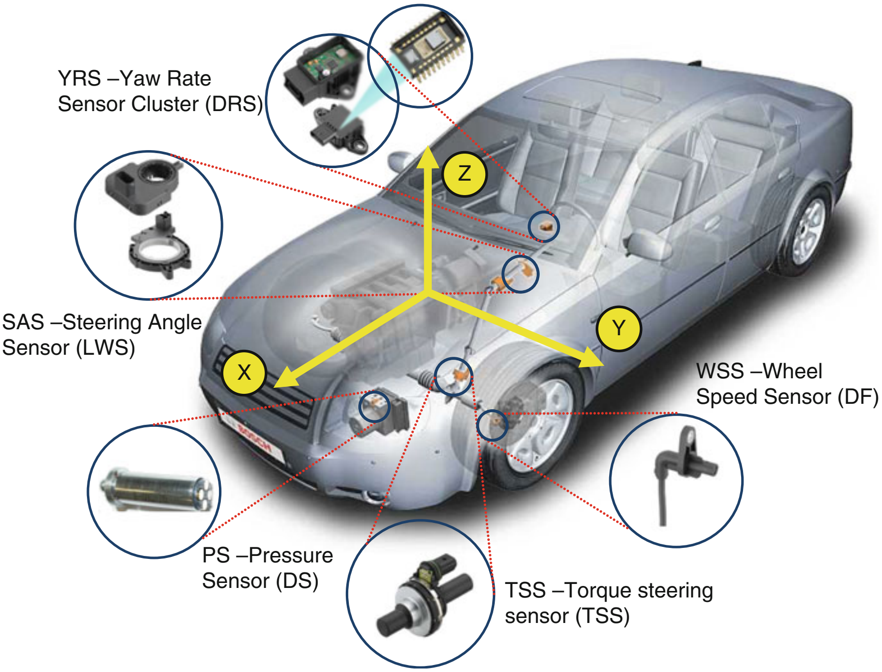


Fig. 2 Overview diagram of vehicle with sensors, with illustration of dimensional axes

Acceleration Sensors The x-sensor is used to record acceleration and deceleration. This also enables the determination of static downward and rearward forces. The acceleration sensor on the y-axis measures radial acceleration with the wheel at full lock and is employed statically to measure road surface inclinations. Acceleration sensors on the z-axis are used to detect vehicle body movement in chassis control systems.

Brake Pressure Sensors The pressure in the master brake cylinder is measured in order to detect the braking force intention of the driver. In high-end VDC systems, the individual brake circuits – or even the pressure of each wheel brake cylinder – are measured. For distance control systems, the pressure must be measured after the accumulator or pump in all cases.

Brake Pedal Travel Sensors In today's braking systems, the driver's braking intention is coupled directly to the hydraulic pressure buildup. The recuperative braking systems of hybrid and electric vehicles additionally require the driver's braking intention to be detected by a pedal travel sensor, in order to control generative braking in interaction with hydraulic braking.

Steering Torque Sensor For power-assisted steering systems, a torque sensor is incorporated in the steering column, to provide haptic feedback of the steering motion (Mörbe and v. Hörsten 2003).

3.2 Wheel Speed Sensor

Functional and Structural Illustration First introduced for ABS in 1978, most wheel speed sensors were inductive in nature. Due to demands to enable wheel velocity to be measured almost down to zero, the passive sensor had to be replaced by an active model (Mörbe and Zwiener 2003). From 1995 onward, these sensors, with measuring elements using the Hall effect or AMR principle, have almost completely superseded the passive versions (Walter and Arlt 2007; Mörbe 2006). In commercial vehicles, however, inductive sensors are still used today, if the axle designs have not been adapted. The basic function and structure of these sensors are illustrated in Fig. 3.

The wheel speed sensor technical data are based on the following parts:

- Sensor head
- Electric cable including sleeves, fastening elements, and a plug

The sensor head is additionally divided into the following zones:

- Sensor zone
- Cable zone

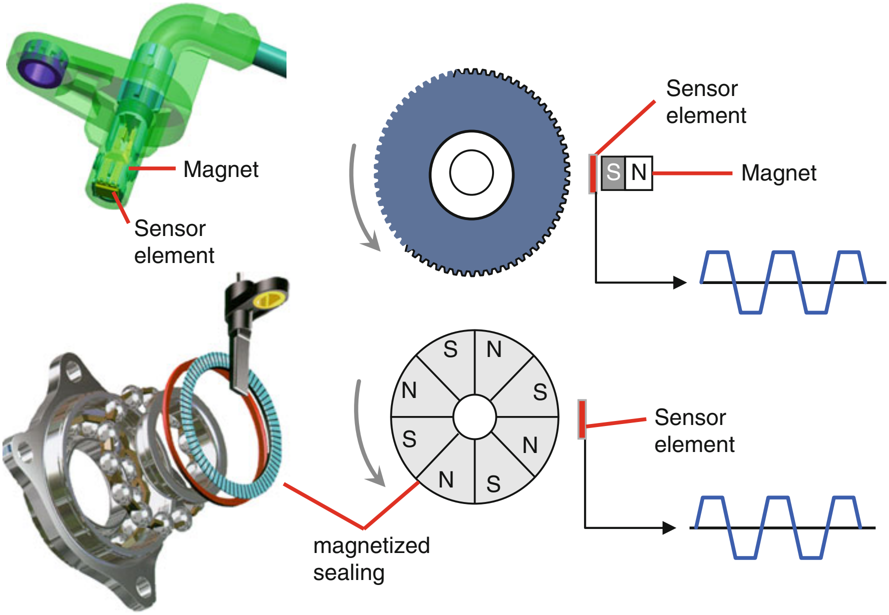


Fig. 3 Wheel speed sensors

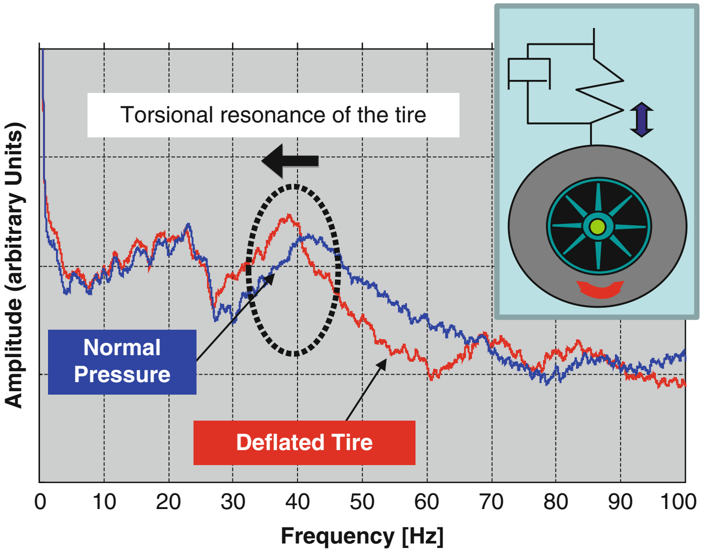


Fig. 4 Functional illustration of indirect tire pressure monitoring using the resonance method

The precise location of the individual zones is established in the quotation drawing.

Different axle constructions require differently designed sensors. The decisive factor is the position of the sensor elements in relation to the pulse wheel or encoder. Active sensors also permit the wheel's direction of rotation to be measured. It is therefore also possible to define a left- and a right-hand installation location.

Tire pressure monitoring can be determined directly by measuring the pressure in the rim or indirectly based on the rigidity of the tire as a spring/mass system. With indirect measurement, the shift in the tire's resonant frequency is calculated by means of a special algorithm. The basic principle is shown in Fig. 4. However, to achieve this, the wheel speed signals have to be emitted as square-wave signals with especially low flank jitter. This jitter is the result of ancillary mechanical effects in the wheel suspension and of thermal noise in the analog part of the sensor element, plus analog/digital conversion in the evaluation circuit.

Wheel Speed Sensor Technical Data **Storage Time**

Criterion	Value
From production date	10 years
Storage temperature	−40 °C to +50 °C

Minimum Life Expectancy

Criterion	Value
Under consideration of temperature limits	15 years
Operating time	12,000 h

Ambient Temperature

Criterion	Value
For sensor zone	−40 °C to +150 °C
For cable zone	−40 °C to +115 °C

The supply voltage must be in the range 4.5–20 V.

Output Signal

All wheel speed sensors work with two switched current levels in a cable with two wires. The lower current level consists of the sensor element's own current consumption and a controlled correction variable. The upper current level is shown as an additive variable based on an additional switched, temperature-compensated current source (Welsch 2006).

Criterion	Value
Signal frequency	1 to 2,500 Hz
Lower signal level I_L	5.9 to 8.4 mA
Upper signal level I_H	11.8 to 16.8 mA
Signal ratio I_H/I_L	≥ 1.9
Signal rise/fall with EMC capacitor and defined test circuit	8 to 26 mA/μs
Duty cycle	$0.3 \leq t/T \leq 0.7$

Tests

The tests listed below are the characteristic of installation on the wheel on the exterior of the vehicle; they are individual tests and are always conducted on all new parts.

Test Conditions

Unless otherwise specified, the following applies to all tests listed below:

Test condition	Value
Test conditions acc. to	IEC 68–1
Ambient temperature	23 °C ± 5 °C
Relative humidity	50 % ± 15 %
Voltage supply Uv (DC)	12 V ± 0.1 V
Input capacity of control unit (incl. cable)	≤10 nF

The characteristics must apply on completion of each test.

Insulation Resistance Measurement

The wheel speed sensor is immersed in a 5 % NaCl solution. The test voltage is applied for the test duration between an electrode in the solution and the short-circuited connector pin. The connector zone is located outside the brine.

Test condition	Value
Test voltage	400 V DC
Test duration	2 s
Test criterion in new condition (R_{Isol})	≥100 MΩ
Test criterion over service life (R_{Isol})	≥5 MΩ

Broadband Random Vibration Test

Test condition	Value
Test conditions acc. to	IEC 68–2-34
Determination of main axes	Defined by customer
Test holder	Defined by customer

The cable is attached at a distance of 50 to 120 mm from the wheel speed sensor head, with the first attachment point (sleeve, panel) on the part of the test holder that is also oscillating.

3.3 Steering Angle Sensors

A very popular design for a steering angle sensor is illustrated in Fig. 5.

The CVH (circular vertical Hall) or GMR (giant magnetoresistive) measurement principle is employed as a noncontact method that guarantees absolute measurement.

The absolute angle is measured using two target wheels, which have a transmission ratio to the steering column hub that differs by two teeth. The target wheels bear magnets that bring about a change in resistance that is proportional to the angle in the GMR elements situated opposite. The analog voltages are digitized, and the out-of-phase voltage profile allows the position of the new sensor to be precisely allocated within three clockwise or anticlockwise revolutions, using the Nonius principle, for example. Counting starts from the center position, i.e., driving straight ahead. For reasons of system safety, a CAN interface is used as the system interface. The calculated angular velocity of the steering wheel can also be transmitted via this interface. A correction factor can be calculated by means of a mathematical operation, shown in Fig. 6 as the yellow and blue field.

The different installation positions require individually designed mechanics. Identical part strategies can also be defined within a manufacturer's platforms, however.

Steering Angle Sensor Technical Data The CAN interface specification is comparable with all other applications of this kind in the motor vehicle and is therefore not explained separately.

The conversion of mechanical variables and their tolerances is of particular importance for system functions. This is especially the case since the plausibility of the steering angle sensor, as an input variable of the driver's intention, must be

Fig. 5 Possible installation position of steering angle sensor

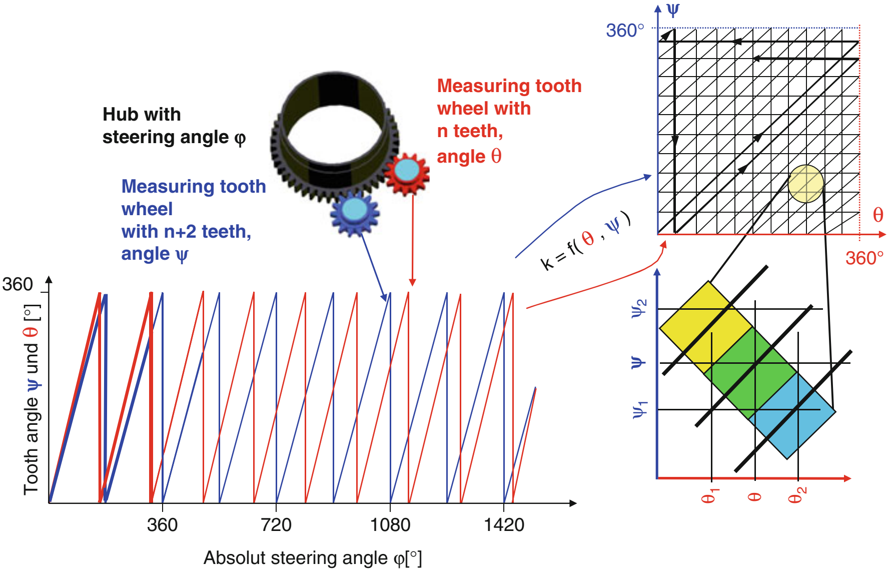


Fig. 6 Schematic diagram of the Nonius principle in the steering angle sensor

checked in the systems together with other signals. If it is also used to control a rear axle steering system, more stringent requirements regarding hysteresis and linearity must be considered. Furthermore, these applications require redundant signal processing.

Functional Characteristic Parameters

The stated values are only valid if the sensor is mounted on the steering column as per the drawing.

Nominal Measuring Range

Function	Value
Angular range	−780° to +779.9°
Angular velocity of steering wheel	0 to 1016°/s

Sensitivity and Resolution

Function	Value
Angle: equivalent to 1 bit (over measuring range)	0.1°
Velocity: equivalent to 1 bit (over measuring range)	4°/s

Nonlinearity

Function	Value
Angle (over measuring range)	−2.0° to +2.0°

Hysteresis

Function	Value
Angle (over measuring range)	0° to 4°

Zero Calibration

Offset calibration of zero takes place via the CAN interface, while the steering wheel and vehicle wheel are moved in one direction. The initialization procedure is contained in the service manual of the system application.

Zero Drift

Function	Value
Maximum zero tolerance between mechanical and measuring sensor interface	−5° . . . +5°

Repetitive Accuracy of Zero

Function	Value
Switch-on repetitive accuracy	−0.5° to +0.5°

Angular Velocity of Steering Wheel

Function	Value
Maximum velocity (<5 s)	−2,500°/s to +2,500°/s

Signal Delay

Function	Value
Delay time between ignition on and a valid output signal without steering movement	≤200 ms

Torque

The torque that has to be added is extremely important for all steering system components. If high torques are produced, the steering geometry is not sufficient to ensure the steering response when driving straight ahead.

Repetitive Accuracy of Zero

Function	Value
Torque (average over measuring range)	≤6 N cm
Temperature	+23 °C
Rotational speed	50°/s ± 10°/s

3.4 Yaw Rate and Acceleration Sensors

Yaw Rate and Acceleration Sensor Technical Data **Measurement Principle**
The purpose of these sensors is to measure the rotation of the vehicle around its axes, as well as lateral, longitudinal, and vertical acceleration. This enables the dynamic status in the room to be clearly determined.

In many applications, the sensor element for the yaw rate is surface micromachined (Golderer et al. 1998; Willig and Mörbe 2003; Schier et al. 2005; Axten and Schier 2007) and connected to a control system for the drive and the evaluation circuit. It works on the principle of the gyroscopic effect. An electrostatic comb-drive causes the oscillating vibration of a seismic mass. Rotation of the vehicle, e.g., around the z-axis (vertical axis), gives rise to a Coriolis force on an acceleration sensor, and the capacitive change of this sensor can be measured. Synchronous demodulation of the measured Coriolis force, which uses the velocity of the seismic mass, generates a signal that is proportional to the yaw rate. Likewise, the sensor elements for acceleration also consist of surface micromachined measuring elements.

In the sensor modules that have been in use for some years now, the minimum charge quantities are boosted in analog form and digitized for further processing (Hagleitner and Kierstein 2005). However, this signal processing with a hardwired logic is extremely costly and time consuming, including during development, due to the changes required. Therefore, additional, smaller microcontrollers are integrated, so that these changes can be achieved using software. In Fig. 7, we see the structure of a multiaxial inertial sensor with various sensor modules. An additional, central microcontroller allows complex calculations to be carried out in the sensor

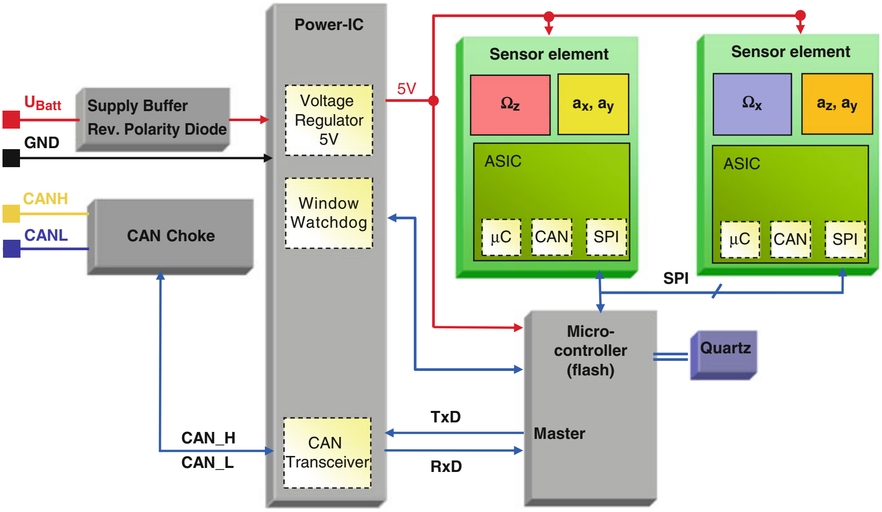


Fig. 7 Block diagram of a multiaxial inertial sensor for motorcycles with CAN interface (Willig and Lemedja 2012)

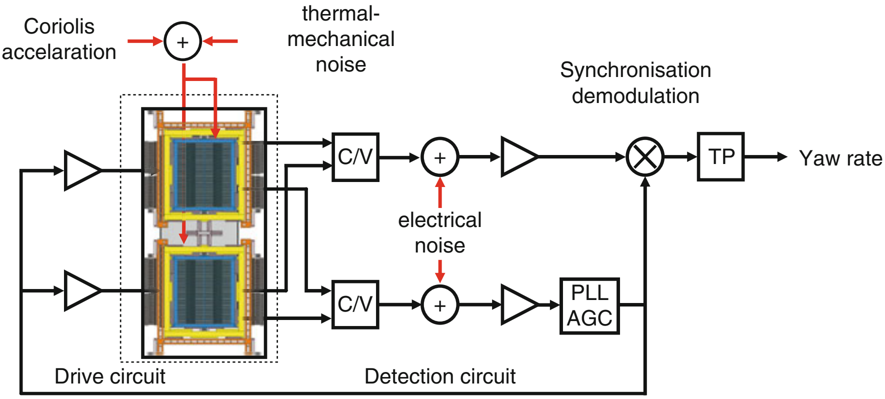


Fig. 8 Micromachined yaw rate measuring element, block diagram of signal evaluation with disturbance variables

already, using the acquired signals. The disturbing factors of the micromechanics shown in Fig. 8 limit sensitivity and achievable resolution, in particular. Sensor modules with 6 DOF (degrees of freedom) are supplied in packages of 2 × 2 mm housings for multimedia applications and form the basis for further miniaturization in the motor vehicle (Baus et al. 2014).

Communication between the sensor and control unit is achieved via a CAN interface that conforms to the customer specification. This enables the multiple use of different systems, and signal transmission itself is less susceptible to interference.

Installation Location

The installation location of the sensor cluster should be selected such as to ensure that only the dynamic movements of the vehicle occur at this spot. The center tunnel or the area around the A-pillar crossmember is particularly suitable. For motorcycles, the mounting location must be largely decoupled from engine vibrations (Welsch 2006). Installation on the floor of the vehicle under the seats must be examined with particular care. Due to the measuring principle of the sensor, which is based on acceleration, secondary, interfering accelerations with high amplitudes and critical frequency ranges, which are not caused by the vehicle movement, must be limited at this installation point. Rocker panel and crossmember connection points have proven effective. Fastening to thin trim panels is to be avoided.

Where service life is concerned, a spectral acceleration test is suitable for making sure that vibrations possibly occurring in the vehicle do not interfere with the sensor. This test is many times more difficult than what would normally be expected in a vehicle. It is intended to simulate the long-term effect (vehicle service life) on the sensor in a short time. It is not possible to define a universal acceleration test for checking function. The accelerations actually occurring in a vehicle are not constant and vary in terms of time, frequency, temperature, and amplitude.

Functional Data

Yaw Rate

Function	Minimum	Typical	Maximum	Unit
Nominal measuring range	−163		+163	°/s
Limit of measuring range	−1,000		+1,000	°/s
Nominal sensitivity		200		LSB/°/s
Sensitivity error (at $\vartheta_{operation}$ over t_{life})	−4	±2.5	+4	%
Nonlinearity	−1	±0.5	+1	°/s
Differential nonlinearity (in increments of 5°/s)	−4		+4	%
Offset, absolute (over t_{life}, measured at ϑ_{op})	−3	±1.5	+3	°/s
Offset drift, operation to operation (over t_{life}, measured at $\vartheta_{operation}$)	−2.0		+2.0	°/s
Resolution, absolute (quantization)			0.1	°/s
Time until availability		0.75	1	s
Cross-sensitivity	−5	±2.0	+5	%
Filter corner frequency (–3 dB)		15		Hz
Output noise		0.05	0.2	$°/s_{\text{rms}}$
Acceleration sensitivity	−0.25		+0.25	°/s/g

Acceleration Signal (Longitudinal and Lateral Acceleration)

Function	Minimum	Typical	Maximum	Unit
Nominal measuring range	−4.2		+4.2	g
Limit of measuring range	−10		+10	g
Nominal sensitivity		490.5		LSB/g
Sensitivity error (at $\vartheta_{operation}$ over t_{life})	−4	±2.5	+4	%
Nonlinearity	−0.072	±0.036	+0.072	g
Offset (new sensor, measured at ϑ_{room})	−0.030		+0.030	g
Offset (over t_{life}, measured at $\vartheta_{operation}$)	−0.1	±0.05	+0.1	g
Offset drift, operation to operation (over t_{life}, measured at $\vartheta_{operation}$)	–0.07		+0.07	g
Rate of change of offset	−0.03		+0.03	g/min
Resolution, absolute (quantization)			0.01	g
Time until availability		0.150	0.250	s
Cross-sensitivity	−5	±2.5	+5	%
Filter corner frequency (–3 dB)		15		Hz
Output noise		0.004	0.01	g_{rms}

3.5 Brake Pressure Sensors

For all vehicle dynamics systems that act by means of the hydraulic braking system, the pressure built up in the system must be measured. A pressure estimation model will suffice for a simple ABS system.

For a simple ESP system, it is sufficient to measure the pressure in the master brake cylinder. For ESP systems with more demanding functional requirements, pressure sensors with up to three independent channels (master cylinder, two brake circuits) are employed. The decisive feature of all pressure sensors in the brake circuits is reliable tightness during operation. This tightness must be ensured by means of burst protection with multiple overload safety. It is essential to avoid even small leaks that are compensated by the brake fluid reservoir. Due to its corrosive effect on the environment, escaping brake fluid can lead to serious faults, e.g., in an installed electronic system. Therefore, the best course of action is to seal the measuring element using welded joints. Measuring cells must be provided with a drift monitor in cases where the absolute value of pressure is assigned a safety function. If an automatic distance control system is controlled via the braking system, the system pressure is a measure of the deceleration that needs to be achieved. To compensate for the temperature dependency of hydraulic pumps, temperature measurement is also integrated. Here, it is important to note that this temperature measurement only represents the value at the installation location of the pressure sensor, not in the system as a whole. The pressure sensor is not necessarily suitable as a means of measuring the brake pedal travel. The brake pedal has already traveled

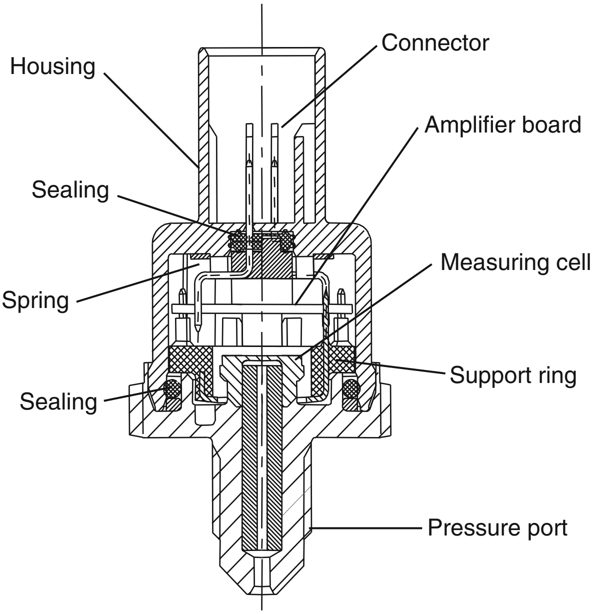


Fig. 9 Typical arrangement of a pressure sensor in the brake control system (screwed version)

some distance before pressure actually builds. This travel is necessary in order to open holes to the brake fluid reservoir. For using the signal to activate the brake lights and deactivate cruise control, accuracy at zero is not always sufficient, or the function is expected to already be activated before brake pressure builds up.

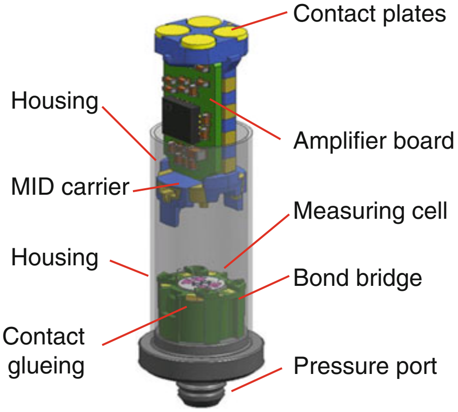
In all known pressure sensor designs, the hydraulic connection is achieved either using a screw connection (see Fig. 9) or a press-fit connection (see Fig. 10). A specific requirement is for a minimal air volume in the sensor, so that special venting can be dispensed with. The brake fluid is degassed when first poured in and is therefore capable of absorbing this small volume of air. The amplifier ICs are specially designed for this particular application and have adjustment features for adapting the sensitivity, offset, and temperature profile. This adjustment takes place during production by means of pressure (air) and temperature. For sealing against external influences, the decision whether to situate the sensor in the hydraulic power unit or outside in the engine compartment is taken during the design phase.

Pressure Sensor Technical Data **Electrical Parameters**
Unless a different temperature is specified, an ambient temperature from −40 to +120 °C is assumed. Functional data and functional parameters

Function	Minimum	Typical	Maximum	Unit
Supply voltage (normal operation)	4.75	5.0	5.25	V
Switch-on delay (output signal not specified during this time)			10.0	ms
Nondestructive supply voltage range	−5.25		16.0	V
Current consumption (normal operation)	9.0		20.0	mA

(*continued*)

Fig. 10 Typical arrangement of a pressure sensor in MID (molded interconnect device) technology for a brake control system with press-fit technology (Hanser 2011)



Function	Minimum	Typical	Maximum	Unit
Undervoltage detection (supply voltage, output signal switched to alarm)	3.7		4.2	V
Overvoltage detection (supply voltage, output signal switched to alarm)	6.0		7.5	V
Open circuit detection, cable break of signal, ground or supply cable (based on supply voltage)	96 %		100%	V
Short circuit detection, signal/supply cable (based on supply voltage)	96 %		100%	V
Short circuit detection, signal/ground cable (based on supply voltage)	0 %		4%	V
Short circuit detection, supply/ground cable (sensor connected to control unit, based on supply voltage)		34 %		V
Short circuit detection, supply/ground cable (sensor connected to load resistance, based on supply voltage)		100 %		V

Functional Data

Functional Parameters

Function	Minimum	Typical	Maximum	Unit
Pressure range, nominal	0		250	bar
Maximum pressure			350	bar
Pressure on destruction	500			bar
Maximum vacuum	−1.0			bar
Increase in volume			0.05	cm^3
Resonance frequency of diaphragms		200		kHz
Lower switch-off frequency	0	0	0	Hz

(*continued*)

Function	Minimum	Typical	Maximum	Unit
Upper switch-off frequency (–3 dB); determined by fixed filter coefficients	150			Hz
Nominal frequency		100		Hz
Phase error (at nominal frequency)			35	°

3.6 Brake Pedal Travel Sensors

Magnetic methods have established themselves for measuring brake pedal travel (Willig and Mörbe 2003). In this application, noncontact detection of change and high signal reliability are vital. The changes may be rotational or linear movements, depending on the design of the brake actuation unit; see Fig. 11. The moving magnet is specifically adapted in size and shape to the travel that needs to be measured. The Hall sensor positioned opposite measures the change in direction of the magnetic flux in both the x- and y-directions. The linear travel or change in angle can be calculated by means of an arctangent function. The design and number of magnets determine the measuring range and sensitivity of the sensor. The signals are emitted redundantly, with cross characteristics and monitoring bands for detecting short circuit to ground and to the supply voltage. The plausibility of the output signals is verified using the aggregate signal of the output signals.

Functional Data
Brake Pedal Travel

Criterion	Value	Unit
Power supply	5.0	V
Current consumption	<30	mA
Output signal	1 kHz (10–90 %)	PWM
Measuring range dependent on magnet	Typ. 45 linear	mm
	Typ. 30 rotational	°
Resolution	10	bit
Accuracy		
Operating temperature range	–40 to +120	°C

Functional Parameters
Brake Pedal Travel

Criterion	Value	Unit
Power supply	5.0	V
Current consumption	<30	mA
Output signal	1 kHz (10–90 %)	PWM
Measuring range dependent on magnet	Typ. 45 linear	mm
	Typ. 30 rotational	°
Resolution	10	bit

(*continued*)

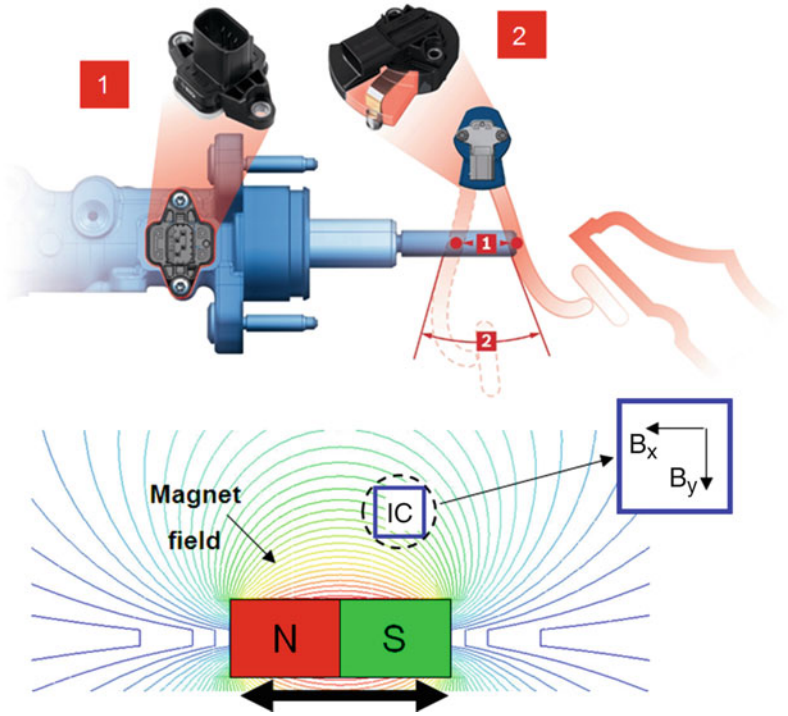


Fig. 11 Functional illustration of brake pedal travel sensor as a linear travel sensor (*1*) or an angle sensor (*2*)

Criterion	Value	Unit
Accuracy	+/−0.4	%
Accuracy based on 45 mm travel	Typ. +/−0.18	mm
Operating temperature range	−40 to +120	°C

References

Axten E, Schier J (2007) Inertial sensor performance for diverse integration strategies in automotive safety. In: Valldorf J, Gessner W (eds) Advanced microsystems for automotive applications. Springer, Berlin, pp 251–263

Baus M et al (2014) Skalierbarer MM7 Inertial-Sensor-Cluster für aktive und passive Sicherheits- und Fahrerassistenzfunktionen im Kraftfahrzeug (Scalable MM7 Inetrial-sensor-cluster for active and passive safety and driver assistance functions in automobile). In: Tille T et al (eds) Sensoren im Automobil V, vol 132, Haus der Technik Fachbuch. Expert Verlag, Renningen, pp 53–67

Golderer W et al (1998) Yaw rate sensor in silicon micromachining technology for automotive applications. In: Ricken DE, Gessner W (eds) Advanced microsystems for automotive applications. Springer, Berlin, pp 69–77

Hagleitner C, Kierstein KU (2005) Circuit and system integration. In: Baltes H, Brand O, Fedder GK, Hierold C, Korvink J, Tabata O (eds) Advanced micro and nanosystems, vol 2, CMOS-MEMS. Wiley-VCH, Weinheim, pp 513–577

Mörbe M (2006) Standardisierung von Sensorschnittstellen – Chance oder Risiko (Standardisation of sensor interfaces – chances and risks). In: sensor4car conference “Sensorsystemtechnik und Sensortechnologie”, Böblingen, 25–26 Oct 2006

Mörbe M (2012) Advanced systems for motorcycles based on inertial sensors. In: Stan C (ed) Development trends of motorcycles IV, vol 127, Haus der Technik Fachbuch. Expert Verlag, Renningen, pp 122–144, 2013

Mörbe M, von Hörsten C (2003) Force and torque sensors. In: Marek J, Trah HP, Suzuki Y, Yokomori I (eds) Sensors for automotive technology, vol 4. Wiley-VCH, Weinheim, pp 450–462

Mörbe M, Zwiener G (2003) Wheel-speed sensors. In: Marek J, Trah HP, Suzuki Y, Yokomori I (eds) Sensors for automotive technology, vol 4. Wiley-VCH, Weinheim, pp 403–416

Published by Carl Hanser, Munich: Kleines Teil, große Wirkung (Small part, large effect). In: Form+Werkzeug 5/2011 p 24, 25

Schier J, Willig R, Miekley K (2005) Mikromechanische Sensoren für fahrdynamische Regelsysteme (Micro mechanical sensors for drive dynamic control systems). ATZ 107:11

Walter K (2011) Erfassung des Fahrerbremswunsches in Hybrid- und Elektrofahrzeugen (Detection of driver breaking intention in hybrid and electrical cars). In: Specialist VDI conference “Innovative Bremstechnik”, Stuttgart, 12.–13.12.2011

Walter K, Arlt A (2007) Drahtlose Sensorik – Anforderungen in sicherheitskritischen Fahrdynamiksystemen (Wireless sensors – Requirements for safety critical drive dynamic systems). In: sensor4car conference “Sensorsystemtechnik und Sensortechnologie”, Fellbach, 24.–25.10.2007

Welsch W (2006) Intelligente Schnittstellen für Raddrehzahlsensoren (Intelligent interfaces for wheel speed sensors). In: sensor4car conference “Sensorsystemtechnik und Sensortechnologie”, Böblingen, 25.–26.10.2006

Willig R, Lemejda M (2012) A new inertial sensor unit for dynamic stabilizing systems of powered two wheelers 9th international motorcycle conference, Cologne, 1–2 Oct 2012

Willig R, Mörbe M (2003) New generation of inertial sensor cluster for ESP-and future vehicle stabilizing systems in automotive applications. SAE Permissions, Warrendale

Ultrasonic Sensors for a K44DAS

16

Martin Noll and Peter Rapps

Contents

Abstract

Typical characteristics of ultrasonic sensors for parking assistance systems in automobiles are described. The chapter includes the principles of sound conversion into electrical energy and vice versa, material and production of piezoceramic transducer elements, acoustic design, as well as structure, setup,

---

M. Noll (✉)
CC-DA/EAU3, Robert Bosch GmbH, Leonberg, Germany
e-mail: martin.noll@de.bosch.com

P. Rapps
Karlsruhe, Germany
e-mail: peter.rapps@o2mail.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_16

and installation of complete sensor elements into a vehicle. Simulation, measurement, and evaluation techniques for obstacle detection and distance calculation in the close range around the car are another major focus of this chapter. Performance characteristics and robustness factors together with a summary about current and future application in various configurations for advanced parking systems on the way to fully automated parking represent a roundup at the end.

1 Introduction

Ultrasonic sensors are used in widely different application areas. Examples to be mentioned here are material testing, medical diagnostics, underwater sonar technology, and in industrial proximity switches. The physical principles and numerous examples of applications are described in many places in the literature (Bergmann 1954; Lehfeldt 1973; Kutruff 1988; Waanders 1991). Unlike this, a use in the automobile has been found only comparatively recently with the parking assistance systems based on ultrasonic technology that were introduced at the beginning of the 1990s, though this application has become widespread since then.

This chapter therefore covers the specific requirements and the design of the ultrasonic sensor components in greater detail for use in parking assistance systems. A large area is taken up at the beginning by the piezoceramic ultrasonic transducers that above all because of the robustness of their environmental properties are particularly suitable for applications in the automobile and have therefore gained wide acceptance here.

2 Principles of Ultrasonic Conversion

2.1 Piezoelectric Effect

The piezoelectric effect discovered in 1880 by Jacques and Pierre Curie describes the linear electromechanical interaction between the mechanical and the electrical states in a crystal. Mechanical deformation of the crystal creates an electrical charge proportional to this deformation that can be tapped as an electrical voltage (direct piezoelectric effect). Vice versa, mechanical deformation can be created by applying an electrical voltage to the crystal (reciprocal or inverse piezoelectric effect). This makes piezoelectric materials principally suitable for generating mechanical oscillations and causing deformation by applying electrical fields in the reverse as sensors for detecting mechanical oscillations and deformation. Since the direct and the inverse piezoelectric effect always occur together, piezo transducers can be used both for transmitting and for receiving sound.

2.2 Piezoelectric Ceramics

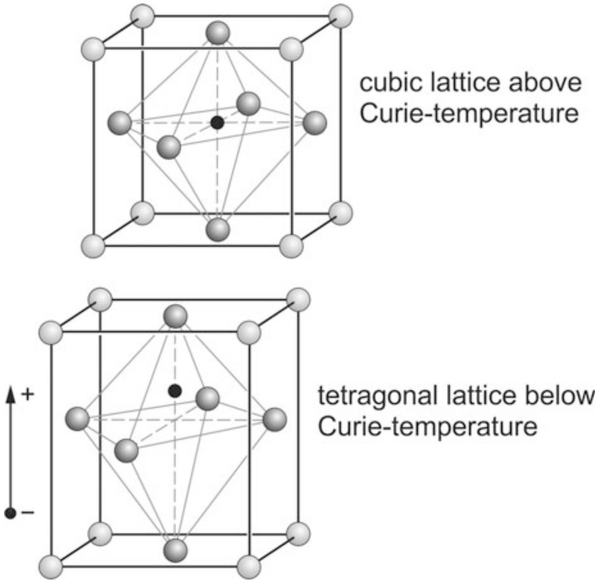
2.2.1 Materials

In practice today, ferroelectrical ceramics are among the most widespread of materials with the piezoelectric effect. There are also organic materials that exhibit the piezoelectric effect. Though because of the low robustness, their use has not played a major role in the vehicle up to now. The more important piezoelectric ceramic materials of today are based on the oxidic mixed crystal system of lead zirconate and lead titanate that is called lead zirconate titanate (PZT). The specific properties of these ceramics, like the high dielectric constant, depend on the molar ratio of lead zirconate to lead titanate as well as on the substitution and doping with additional elements. Many possibilities for modification are feasible in this way to give materials with widely different properties.

Dissymmetry in the distribution of positive and negative electrical charges sets in within the lattice of a cell below the so-called Curie temperature. This results in a permanent electrical dipole moment of the single cell (Fig. 1). The ferroelectricity is given from the development of domains with uniform electrical polarization that are aligned by the polarity, i.e., by a strong electrical constant field applied for a short time. The polarity is associated with a change in the length of the ceramic.

The work in the field of lead-free piezoelectric ceramics has been intensive as part of the efforts not to use lead in the vehicle wherever possible. There is however no alternative to using today's ceramics that can be expected in the short term from this research.

Fig. 1 Crystal lattice



2.2.2 Production

The starting point in the production of piezoelectric ceramics of the type PZT is the oxides of the metals lead, titan, and zirconium. These are calcined after admixing. The constituents of the mixed crystal systems are given from the chemical bonds established between the materials during this thermal process. The material that forms during the calcining is grinded and conditioned together with additives to give the so-called ceramic in the green state. The ceramic material is still soft in this state and can be easily shaped as desired.

The electrodes are produced by suitable metallization of the ceramic surface once the required shape has been realized. Common methods here are the electrochemical deposition of metals, vapor deposition, sputtering, and thick-film technology. In thick-film technology, a paste of metal particles together with organic and inorganic binders is sprayed or printed onto the surface and then fired. It shall be observed here that the inorganic constituents of the binders migrate to a certain degree into the ceramic material to thereby modify the piezoelectric properties. The shape for the electrodes is realized either by screen printing or by subsequent lasering in the case of spraying.

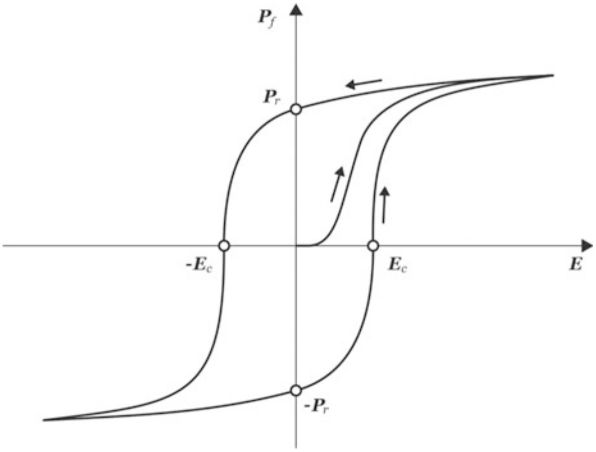
During polarization, an electrical DC voltage is applied to the electrodes. The upper limit for the magnitude of the voltage is given by the breakdown voltage in the ceramic and the lower limit by the operating voltage later, whereby the polarization voltage always has to be larger than this.

2.2.3 Hysteresis

Analogous to ferromagnetism, ferroelectrical materials exhibit a relationship between E, the applied electrical field, and P_f, the polarization, in form of a hysteresis as shown schematically in Fig. 2.

During the polarization the new curve is run and after cutting out the electrical field, the remanent polarization, P_r, then remains.

Fig. 2 Ferroelectrical hysteresis loop



2.2.4 Piezoelectric Constants

The physical constants like elasticity and permittivity are tensors because of the anisotropic nature of the polarized piezoceramics, whereby the direction of the polarity is arranged as a rule in the z-axis and the index 3. The x-axis and the y-axis with the indices 1 and 2, respectively, designate the axes vertical to this.

The dielectric constant $\varepsilon_{33}/\varepsilon_0$ lies typically in the range between 1,500 and 3,000. A further major parameter is the coupling factor that describes the ratio of converted energy to the energy expended. In the case of ultrasonic transducers with thin piezoceramic disks, this is the planar coupling factor k_p describing the coupling between the electrical field in the z-direction and the mechanical effects in the x- and/or y-directions that is of particular significance.

The coupling factor with the ultrasonic transducer is dependent both on the ceramic material and on the design of the ultrasonic transducer. The information about the coupling factor given in ceramic datasheets refers to a standard disk and the value for this lies typically in the range between 0.6 and 0.7.

2.2.5 Depolarization

When using piezoceramics, particular attention shall be paid to the special aging effects that can lead to changes in the material properties in the course of the service life (in the vehicle up to 20 years). The main aging effect here is a gradual depolarization of the material. This starts directly after being polarized and continues in a logarithmic relationship with time. Preaging by storing at an elevated temperature is a method that can be applied for stabilization of the material.

Other depolarizing mechanisms are of a thermal, electrical, or mechanical nature. Degradation of the piezoelectric property is accelerated by heating. Above the Curie temperature that for the materials commonly used in the automotive field lies above 300 °C, the piezo effect disappears completely and polarization is only possible after first allowing to cool down and then polarizing again. In practice, the rule of thumb applies that the maximum service temperature shall be half the Curie temperature in degree Celsius.

Depolarization can also take place by applying a constant electrical field in the direction opposite to the polarization. The same effect is given by a mechanical force applied to counteract the mechanical deformation that has occurred during the polarization. It must be ensured by the mechanical design of the ultrasonic transducer and from the electrical circuitry that these depolarizing effects remain negligibly small over the service life.

3 Ultrasonic Transducers

Ultrasonic transducers that emit sound in air or shall receive sound again in air are – in contrast to applications in liquids and solids – dependent on relatively large amplitudes to decouple sufficient energy in the air and to couple this in again from the air. The mechanical deformation of the piezoceramic itself does not suffice for this which is why mechanical amplification of the effect is needed. This is realized

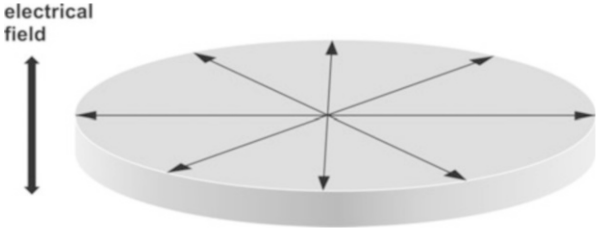


Fig. 3 Planar oscillations of a piezoceramic disk

with customary vehicle ultrasonic distance sensors in that a piezoceramic platelet is adhesive bonded over the whole area onto a metallic diaphragm. When an AC voltage is now applied between the electrodes, then both the diameter and the thickness change (Fig. 3). Since the platelet is attached to the diaphragm by adhesive bonding, these changes are transferred to the diaphragm as bending oscillations that generate considerably larger oscillation amplitudes when operated at the resonance frequency.

Conversely, a sound wave causes bending oscillations by the diaphragm and by this also causes a change in the diameter of the piezoceramic platelets. An electrical DC voltage develops between the electrodes that is then amplified and processed further electrically. Ultrasonic transducers are usually used both for transmitting and receiving ultrasonic sound.

The oscillating diaphragm has to be held firmly in place at the edge. This is realized in practice by adhesive bonding the piezoceramic disk to the base of an aluminum pot. The base acts as a diaphragm and the stable side walls of the pot hold the diaphragm firmly in place on the outside. The oscillations thereby concentrate mainly on the diaphragm, though the side walls are included in the oscillations to a certain extent as well. This is of significance to the extent that by this, the tensioning of the pot does have an influence on diaphragm movement.

The diaphragm is normally excited to the basic oscillation (see Fig. 4). Higher modes are in principle beneficial as well though these do however lead to larger side lobes forming, and this can have a detrimental effect on the spatial detection properties.

To make transducers more robust and easier to control, the inside of the diaphragm has to be dampened acoustically in a specific manner, e.g., by including a silicone foam, the material, and cell structure of which have been matched to the working frequency.

Even though the efficiency of the transducer is reduced, the advantages of this are however:

- Harmful sound emitted into the sensor interior is absorbed immediately.
- The robustness toward external coatings on the diaphragm (e.g., from dirt or moisture) and frequency-changing temperature/aging influences is increased.

For ultrasonic-based parking assistance systems, a working frequency in the range around 40–50 kHz has be found to be the best compromise between the

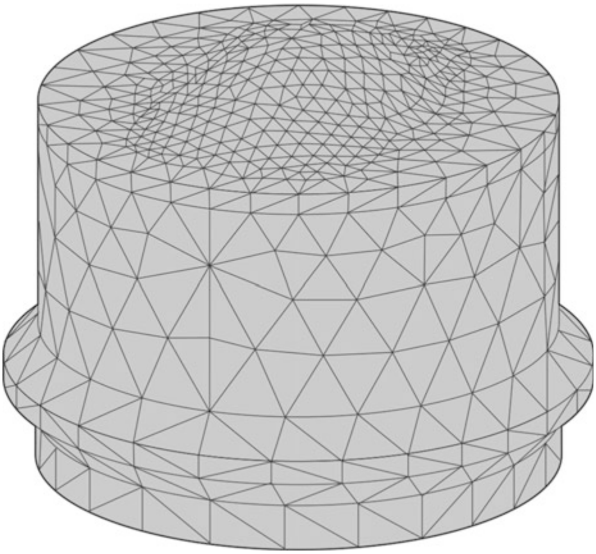


Fig. 4 FEM simulation of the oscillating diaphragm (base of the pot excited into resonance)

competing requirements for good system performance (sensitivity, range, etc.) on the one hand and greater robustness toward noise from extraneous sound sources on the other hand. Higher frequencies lead to lower echo amplitudes because of higher dampening of the airborne sound, whereas for the lower frequencies the proportion of the sources of interfering sound in the vehicle environment is always increasing.

3.1 Equivalent Circuit

A piezoceramic ultrasonic transducer can be shown close to its resonance frequency by an electrical equivalent circuit (Fig. 5) consisting of a resonant circuit in series with a parallel capacitance C_0.

The electrical parameters C_s and L_s thereby correspond to the mechanical parameters of the springiness of the diaphragm and its oscillating mass. R_s is an expression for the losses due to friction, ferroelectrical hysteresis, and the emission of sound. The serial resonance frequency is given by

$$f_s = 1/\left(2 \cdot \pi \cdot \sqrt{L_s \cdot C_s}\right)$$

C_0 is the disk capacitance of the piezoceramic. The value for C_0 in the adhesive bonded state of the ceramic is considerably lower than before bonding when the mechanical deformation can develop without being impeded. The transducer is tuned either in parallel or in series to increase the bandwidth of the system. Figure 5 shows parallel tuning: The electrical parallel circuit must be tuned to the same resonance frequency as the mechanical series circuit.

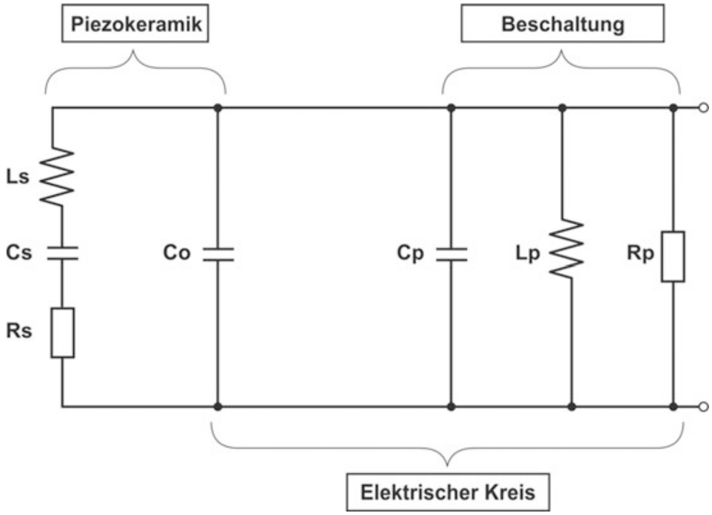


Fig. 5 Equivalent circuit of an ultrasonic transducer with parallel tuning

Since C_0, the capacitance of the piezoceramic, has to exhibit a marked positive temperature dependency, it is purposeful to compensate this effect by a parallel capacitance with a negative temperature dependency. In this way the resonance frequency of the electrical circuit can be kept stable with regard to the temperature.

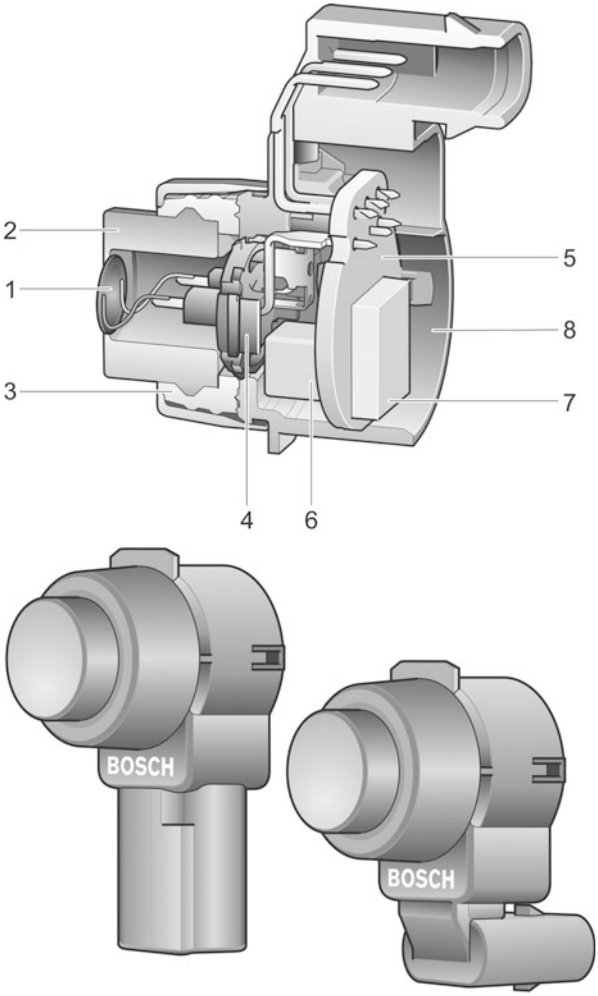
When an ultrasonic transducer is used in the transmit and receive mode, then it is necessary for measuring the distance from nearby objects that the diaphragm oscillations after the active transmission signal settle within as short a time as possible so as to allow the system to be receptive again as soon as possible. Quick stabilization and decay behavior – especially in parking assistance applications – are therefore a major quality and function characteristics for the ultrasonic sensors that are used here.

Conversely in the transmission mode, it is advantageous when the mechanical oscillations of the diaphragm resound as quickly as possible after applying the electrical AC voltage. Shorter ultrasonic pulses are possible in this way. Practical values for the effective transmission duration are typically around 300 μs, whereas for the subsequent decay these can last a further 700 μs.

4 Ultrasonic Sensors for the Vehicle

The basic function of an ultrasonic parking assistance system, the major characteristics of the associated components, and the interaction of these are described in examples given in Robert Bosch GmbH (2004). The properties of the sensors are gone into in greater detail again in the following since these constitute the core of each system and have a fundamental influence on the function and quality of the whole system.

Fig. 6 Sectional view of the sensor module (*1* piezoceramic, *2* diaphragm pot, *3* decoupling ring, *4* contact, *5* circuit board, *6* transformer, *7* ASIC, *8* housing with plug-in connection)



4.1 Sensor Assemblies

The main components of the sensor are the acoustic transducer element (analogous to a combination of speaker and microphone), the electronics, and the housing with plug-in connection. An example of such a build is shown in Fig. 6.

4.1.1 Acoustic Transducer Element

The acoustic part of the ultrasonic sensor is essentially made up of the pot-shaped aluminum body, on the base on the inside of which the piezoceramic disk is adhesive bonded. During installation in the vehicle, this so-called diaphragm pot is fitted to be more or less flush with the outer skin of the bumper and as a rule is in the same color as the installation environment. Decisive in both the design and the build of the “front end” is the complete decoupling of the diaphragm oscillations from the sensor housing and the holder mounted in the vehicle. This is the reason why the diaphragm pot is embedded in a decoupling ring made of a soft silicone material, the acoustic properties of which remain almost unchanged over the whole service temperature range and at low temperatures in particular. The design of the diaphragm pot has also been optimized such that any oscillations at the edge in the region of the clamping on the outside will only exhibit amplitudes that are as small as possible. Also, the electrical contacting of the piezoceramic has been designed by using thin litz wire or strands of wire such that there can be no acoustic coupling on the circuit board in this way.

4.1.2 Electronics

All ultrasonic distance sensors used in the vehicle contain electronic components, the scope of which can vary greatly depending on the system design (partitioning sensor and analyzer ECU). An approximate classification into the following three types can be made:

- Sensors with a purely analog interface
- Sensors with a purely digital interface
- Sensor with a data interface analogous with time

Sensors with a purely analog interface are triggered during transmission with an AC voltage and return the raw or (pre)amplified analog echo signal to the higher-level control unit. The scope of the electronics here is limited to a few passive and discretely active components. Unlike for sensors with a purely digital interface, a distance is computed directly in the sensor from the runtime of the ultrasonic pulse and then signaled to the control unit as a datum.

The most common sensors are those with a data interface analogous with time that are typically triggered by a pulse for transmitting, the length of which gives the transmission duration. On the same – bidirectional – signal line the electronics returns a switching pulse to the control unit at that point in time when an echo is received. The distance information is the difference in time between the two switching edges from the transmission echo pulse. The sensitivity of the echo detection can thereby include a programmable time or distance dependency to fulfill the various boundary conditions for installation in the vehicle bumper (height, angle, lateral installed distance, protruding detachable parts like trailer coupling, number plate holder, etc.) in a manner as universal as possible. A block diagram with the main functions of generating the transmission signal, conditioning the echo signal, and sequencing control for such a sensor is shown in Fig. 7.

4.1.3 Housing

Besides protecting the transducer and the electronics from environmental influences, the job of the sensor housing is to make the plug-in connection with the harness and snapping into place in the sensor holder possible. Because of the

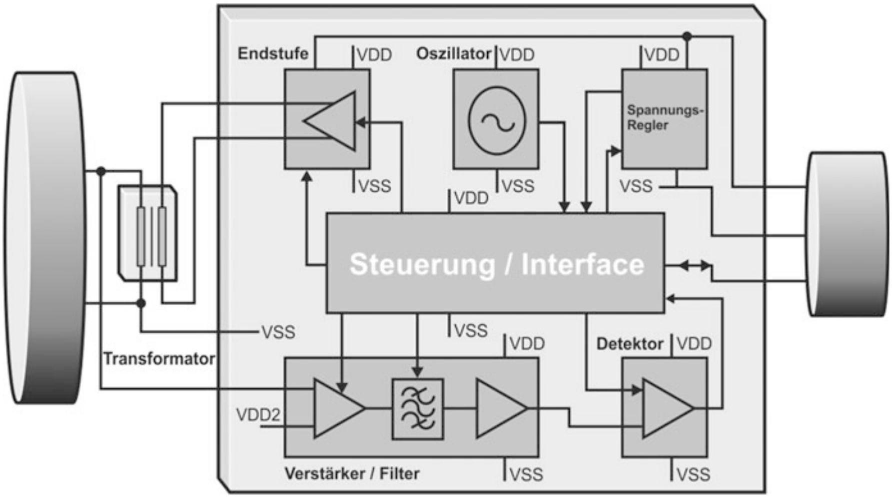


Fig. 7 Block diagram of the ultrasonic sensor with the data interface analogous with time

installation in the splash water area of the vehicle, the housing is generally filled out with a potting material that seals the electronics from water while preventing at the same time that undefined cavities can detrimentally influence the acoustic behavior. The potting material shall be selected such that no damage to components or to solder joints in the electronics can take place due to temperature changes.

5 Antenna and Beam Design

The directionality characteristics or the antenna diagram of an ultrasonic parking assistance sensor constitutes one of the crucial characteristics for the quality of the resulting object detection and the function based on this for acquisition of the surroundings. This should be spatially homogeneous, i.e., shall not exhibit any appreciable interference effects or side lobes so as to keep the dependency of sensor performance on the angle as small as possible. For complete coverage of the vehicle width with as small a number of sensors as possible, the horizontal sound distribution part from this should exhibit a large effective opening angle (approx 120–140° for detection of a reference object in the close range up to about 50 cm). At the same time however, the vertical opening angle should be designed to be as small as possible so that reflections from the road surface – especially in the case of a rough road, e.g., gravel/paved surface – will not cause any signals that might lead to pseudo-obstacles being displayed. Actually for the installation of the sensors in the bumper, an effective vertical opening angle has asserted itself that with about 60–70° is only about half as large as the horizontal angle.

5.1 Simulation

Short development times and widely different boundary conditions for installation of the sensors in the bumper call for efficient and exact predictions already in the very early stages of the project for the acoustic sensor performance to be expected as a function of the installed position, the installation angle, and above all the particular installation environment. Mature simulation methods, models, and tools are the ideal auxiliaries here for making reliable statements possible during the design phase without having to depend on a costly and time-consuming production of prototype parts and the trials based on these.

For the emission of sound, the boundary element method (BEM) has been shown in recent years to be the method best suited for this. Here and contrary to the finite element method (FEM) for three-dimensional problem assignments, only the sound-emitting surface is discretized, not however the surrounding volume as well. The computing efforts required are reduced considerably here because of the significantly lower number of data points.

The emission properties of an ultrasonic transducer installed in the vehicle bumper are subject to several influencing variables. On the one hand these depend to a considerable extent on the vibration velocity distribution on the diaphragm excited by the piezoceramic. These can be determined both experimentally and by simulation (see Fig. 4). The climatic boundary conditions shall also be observed and above all the temperature because of the influence from the speed of sound on the wavelength depending on the excited frequency. Finally, the geometry is also included in the immediate vicinity of the transducer. It is this geometry in particular that depending on the bumper design, installation criteria, holder, and securing concept can have a very significant effect on the resulting sound emissions and in the same way on the spatial reception properties as well. As an example Fig. 8 shows the difference between a planar sensor installation (left) and a slightly deeper installation (right) that results in a considerable constriction of the sonic lobe and hence would also cause a very inhomogeneous detection performance. With the help of simulation, such undesirable interference effects can be largely minimized – even when these cannot be prevented entirely – in that the influencing geometry is optimized specifically with regard to a distribution of the sound emission that is as homogeneous as possible.

On the basis of the computed emission, it is also possible while taking the attenuation of the airborne sound and the reflections of sound at defined obstacles into account to determine the field of view depending on the obstacle geometry. Models can be found in the appropriate literature both for the attenuation of the noise as a function of the medium (here: air), temperature, and humidity (water vapor content of the air) and for the reflections at objects.

Since the simulation, including modeling, is subject to certain limits and conditions, comparisons on a regular basis of the simulation results with real measurements are indispensible for validation and developing the methods further.

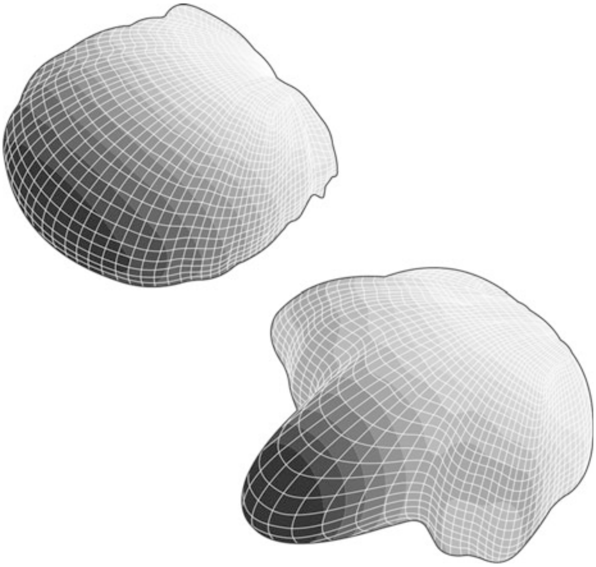


Fig. 8 3D simulation of the sound pressure distribution for a planar sensor installation (homogeneous course for the angle) and deeper installation, caused by the inclination of the bumper area from the diaphragm surface (severe constriction because of interference/side lobes forming in the protruding installation environment)

6 Distance Measurements

The distance measurements using ultrasonic technology according to the pulse/runtime principle are designed to be very straightforward from the technical viewpoint because of the comparably low speed of sound. On the basis of the electronic time measurement between the start of a transmission pulse and the reception of the returned echo signals, the distance from the reflecting obstacle can be calculated directly from the speed of sound on which this is based.

The absolute accuracy of the measurement is thereby influenced by different factors. On the one hand these are the physical dependencies of the speed of sound on the properties of the air as the propagation medium. Here it is above all the temperature of the air that is the deciding influencing variable compared to the other parameters (le.g., the density) that can almost be neglected. Besides this, the echo strength plays a role because it is especially for small signals where a delay in the runtime of the detected signal has to be accepted. This is due to the creation of the echo signal over time because of the limited bandwidth of the piezoceramic ultrasonic transducer. Because of the comparably low requirements on determining the distance with an accuracy to within centimeters for the parking assistance applications, all of these dependencies can be easily tolerated.

Critical are the geometrical measurement accuracies with respect to the limitations of the vehicle that are determined mainly by the position, expansion, geometry, and orientation of the obstacles to be detected relative to the sensor. Errors in the measurements due to this can easily amount to up 20 cm and more. Crucial for a

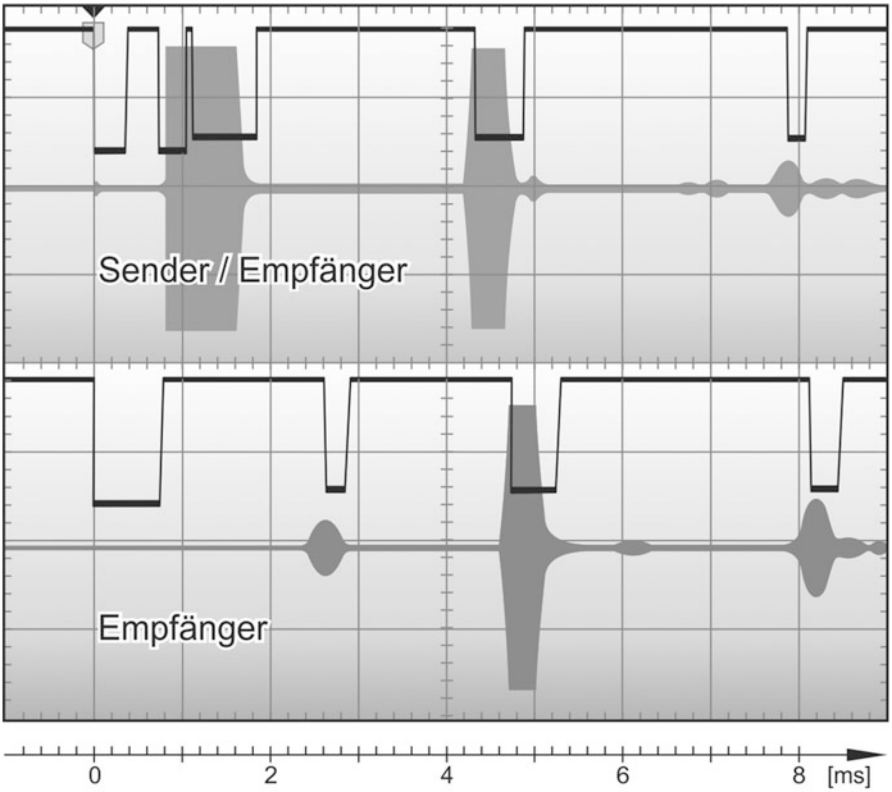


Fig. 9 Example for the signal diagram of two adjacent sensors (*top*: transmitter/receiver, *bottom*: receiver)

reduction of these errors in the measurements attributable to the geometry are on the one hand the use of several sensors over the whole width of the vehicle (typically 4 or 6) and on the other hand to apply the so-called trilateration principle (see Sect. 6.1). Here for each transmission pulse of a sensor, both the received echo signal itself (direct echo, DE) and the echo signal received from the respective left and/or right adjacent sensor (cross echo, KE) are taken for computing the obstacle distance. In this way the position of the next obstacle can be determined approximately within the sensor plane, and hence the actual distance from the vehicle can be calculated as a projection on the bumper.

Figure 9 shows an example of a signal diagram comprising the transmission/reception signal of an active-operated transmitter (top half of the diagram) and the reception signal of an adjacent passive-operated receiver (lower half of the diagram).

Recorded for both sensors are both the digital signals on the bidirectional line between sensor and control unit and the associated sensor internal 50-kHz ultrasonic signals.

A whole system consisting of up to 12 sensors in the front and rear bumpers has to be designed for a carefully matched sequencing control that on the one hand allows high-repetition rates for each individual sensor (necessary for realizing short overall measuring times) while on the other hand will prevent mutual faults from incorrectly allocated signals from different sensors with certainty.

6.1 Trilateration and Object Localization

The very large horizontal opening angle of the sensors and the comparatively moderate installed distance (typically between 40 cm and 70 cm) will result in large overlapping of the effective detection areas. Besides the purely distance measurements, this makes it possible to determine the position of the objects to be detected relative to the vehicle. The basic principle thereby applied of trilateration (see Fig. 10) computes the point of intersection of the two circle arcs with the radii DE1 and DE2 (≅ runtime of the direct echoes) starting from S1 to S2 at the distance d. The height D of the triangle thereby formed forms the

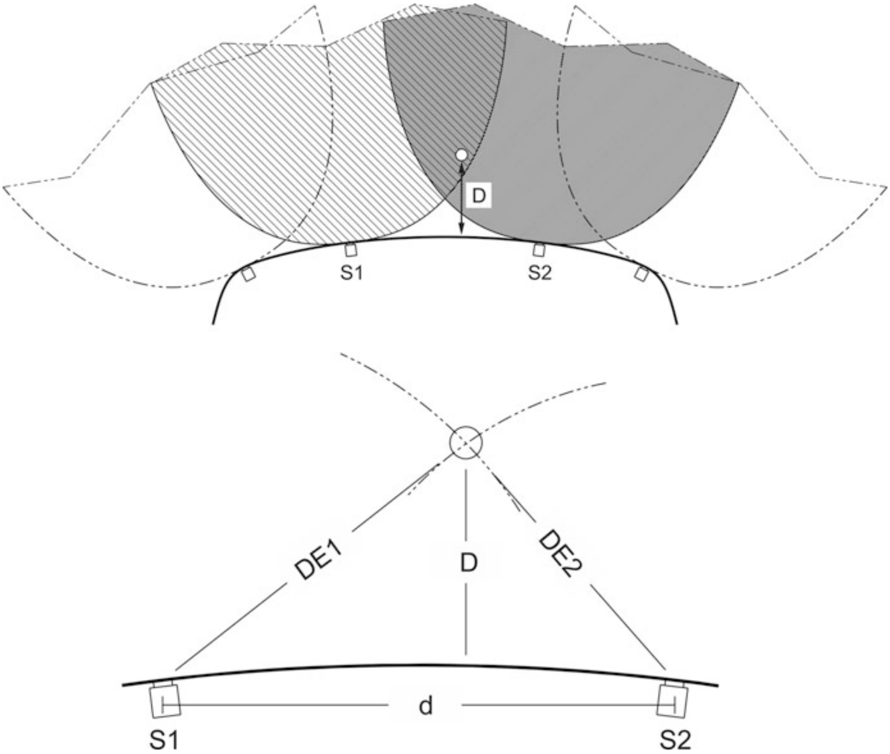


Fig. 10 Illustration of the trilateration principle for calculating object position and object distance D relative to the baseline of the two sensors S1 and S2 at the distance d

projection on the baseline and thus corresponds to the distance to be displayed of the detected object from the bumper. The true curvature of the bumper in the installation plane of the sensors can in the first instance be neglected here or by using a correction factor – e.g., in the region of the vehicle corners – be included in the calculations. D is calculated from the Pythagoras theorem according to the equation

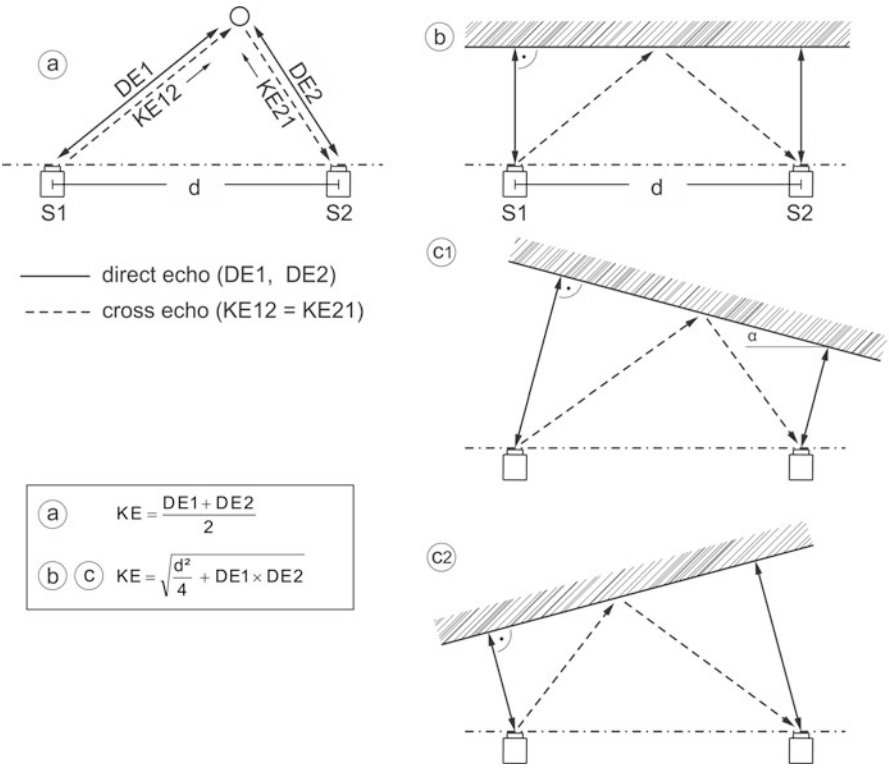
$$D = \sqrt{DE1^2 - \frac{\left(d^2 + DE1^2 - DE2^2\right)^2}{4d^2}}$$

As can easily be seen from Fig. 10, these calculations for the position and distance to prevent the risk of collisions by inexact distance displays are of major significance, especially for the small obstacle distances. The relative differences between the measured (DE1, DE2) and the true distances (D) can be particularly large in the safety range that is of the greatest interest and with the most benefit for the driver.

The special challenges when applying this principle in the field of parking assistance are that the assumption made before of a "point-shaped" obstacle (round post, traffic sign, etc.) only applies in practice in very few cases. Instead, the objects to be detected can be of any size, shape, or spatial orientation. Examples here given in Fig. 11 are of three simple "scenes" showing (a) a cylindrical object, (b) a parallel wall, and (c) an inclined wall.

Mastering this diversity of real scenes is ensured by algorithmic processing all the echo signals available that in the first instance differentiate local "point-shaped" obstacles from expanded "wall-shaped" obstacles, where the specially modified equations for computing the shortest distance are then used. Including the cross echoes (KE) between sensor X and sensor Y (KE12/KE21 in Fig. 11) in the calculations is of particular significance here. How the length of the cross-echo route differs from the sum of the two direct echo routes depending on the object type is relatively pronounced and can therefore be taken as a criterion for the classification "point/wall." The theoretical relationship between cross echo and direct echoes is included in Fig. 11 as well for both cases. The "wall equation" thereby applies both for straight and for inclined orientation toward the base distance d. Once the classification "wall" has been made on the basis of the measured echo runtimes, the inclined position "α" can then also be determined in a straightforward manner from the differences in the runtime between DE1 and DE2.

With four or six sensors in the front or rear bumper, it is possible in this way despite the very poor spatial resolution of a single sensor to create maps for the objects/surroundings for the installed plane of the sensors where reproduction of position, expansion, and orientation of one or several objects in front of or behind the vehicle are good. In combination with advanced tracking algorithms where the vehicle's own movement as sensed by the wheel sensors and the steering angle is continuously compared with the current measurement data from the sensors, more accurate and reliable parking assistance systems can be realized in this way provided there is adequate computing capacity.



$$\text{(a)}\quad KE = \frac{DE1 + DE2}{2}$$

$$\text{(b) (c)}\quad KE = \sqrt{\frac{d^2}{4} + DE1 \times DE2}$$

Fig. 11 Illustration of the shortest-echo routes for the direct echo (DE) and cross echo (KE) using the example of two neighboring sensors and different object types/orientations

As “advanced maneuvering functions” become more widespread with the automatic steering and/or braking interventions as well as the spatially resolved HMI displays (“bird’s-eye view”), this method of object localization is the basis for a significant increase in customer benefit not only in terms of accuracy reliability and robustness. It also makes expansion of the display possible and also monitoring the sides of vehicle for collision in the sense of side protection. Objects that have once been “located” by the front or rear sensors on the right or the left and have already been entered in the map of the surroundings can still be tracked relative to the vehicle and be displayed after leaving the range of detection when the steering angle and path traveled are included in the analysis of the object position.

7 Holder and Securing Concepts

There are many different requirements to be fulfilled by the design for the sensor and mounting this in the bumper. To be mentioned first is the design that shall support the possibility to integrate the sensors to be inconspicuous as possible so as

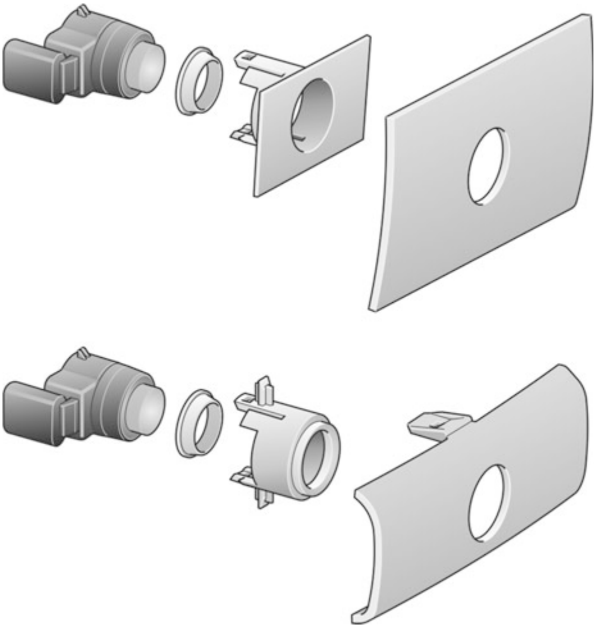


Fig. 12 Examples typical for mounting for the sensor module

to be hardly seen from the outside. This is accompanied by the requirement that it must be possible to paint the visible part of the sensor (oscillating diaphragm) in all bumper colors available without any impairment of the function.

Stability to oscillations, heat, weather, and humidity of the fully mounted sensor module, as well as the reliability in decoupling the diaphragm from its installation environment, furthermore play a major role for correct functioning under real operating conditions over the lifetime of the vehicle.

Two ways for mounting in the bumper using a suitable holder and method of securing are shown in Fig. 12. In the example at the top, the holder is adhesive bonded or welded on the inside over the whole area with the bumper (ultrasonic welding constitutes the method for securing that is widely used in series production). In the lower example, the holder can be snapped into position directly with the bumper using the appropriately attached tabs. The sensor module is then pushed in from the back with the premounted decoupling ring and locked in place in the holder.

The separate design necessary for the decoupling ring for acoustic reasons is needed so that the protruding sensor diaphragm can be painted on all sides in the matching bumper color before installing the module.

8 Performance and Reliability

The wide usage of ultrasonic sensors in the field of parking assistance systems is based on a series of characteristics where this technology is superior to other rival measuring techniques (e.g., RADAR, infrared, capacitive or inductive measuring

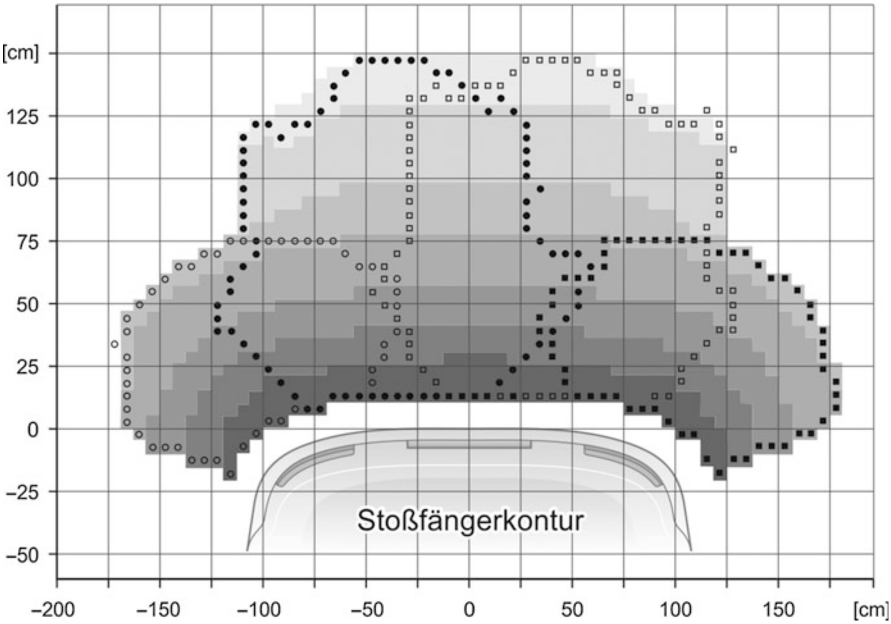


Fig. 13 Example of FoV measurements for a four-channel system

methods), e.g., cost-effective production and higher stability to weathering. Also the detection quality in many areas is independent of the type of obstacles to be detected. Materials of relevance like metal, plastics, wood, masonry, and glass do not absorb sound at the surface and thus provide reflections of almost the same strength for the same geometry. Limitations are only given in the case of partly noise-absorbing materials, though these hardly play a role in practice (e.g., foam). A peculiarity is given in the detection of persons where depending on the clothing, a slightly reduced measuring range has to be expected.

The performance of a sensor or of a sensor system can best be shown by the field of view (FoV) or dimensional check of the detection field and assessed by comparison. Typically to this end, a tube of 7.5 cm in diameter that has been defined as the standard test body in the so-called MALSO standard for the design of passenger car parking assistance systems is used as a reference object (ISO 17386 2004). An example of such FoV measurements is shown in Fig. 13 for a four-channel system where the detection ranges were made visible for the direct echoes of the four sensors by a separate border in each case.

The measuring range in this example is set for the two center sensors to 150 cm and for the outboard sensors to 80 cm and by this corresponds to a typical application case.

The dependable availability of the sensors can in practice be limited by two factors. On the one hand strong extraneous acoustic emitters in the region of ultrasonic working frequency in the immediate vicinity of the vehicle can lower the signal-to-noise ratio such that measurements are no longer possible. Of relevance in practice here are above all compressed air noises (e.g., air brakes in trucks) and metallic grating noises, e.g., from railed vehicles. On the other hand, any layers of dirt, snow, or ice on the sensor diaphragms can form a sound bridge with the bumper that can prolong the decay behavior of transmission excitation in an undefined manner. In both cases the system responds as rule by indicating a fault to the driver or a pseudo-obstacle at a distance that is less than potentially real obstacles. Critical situations where the driver is either not informed of collision risks or is informed of these too late can hence be prevented in this way.

9 Summary and Outlook

Since their initial serial use in 1992, piezoceramic ultrasonic sensors for parking assistance systems have been continually developed further with regard to their mechanical, acoustic, and electronic properties. Modern series production sensors are of compact and robust construction, can be integrated very inconspicuously in the painted or unpainted vehicle bumper, are matched in particular to the optimum angle characteristics of their acoustic transmission/reception behavior, and can be adapted individually to the customer’s request by electronic means and the different installation conditions in the vehicle.

Possible further developments in the future regard to, e.g., enhanced and expanded functionality, better self-diagnostics capability, reduction of the minimum measuring distance, as well as optimized filter mechanisms for increasing the robustness and signal-to-noise ratio.

Parallel to developing the ultrasonic sensor technology, new vehicle applications and multiple usage for the expanded scope of functions of the “normal” parking assistance have recently attracted the interest of the vehicle manufacturers. The basis for this is the very good cost-performance ratio of the ultrasonic sensors produced in large quantities. In the first instance here, it is the exactness in checking the dimensions of a gap for parking in linear parking spaces, on the basis of which the decision can be made whether there is adequate space available for parking the vehicle. Used to this end there are sensors at the front fitted in the corners at the sides of the bumper. These detect parking vehicles together with limits at the sides and the corner positions, dimension checks of the depth of the gap for parking are made for possible barriers, and information is provided about the distance from the curb. On the basis of this, the first series applications for the automatically controlled parking maneuver in linear spaces for parking were launched on the market. A similarly large market penetration is expected in the coming years as was the case as of about 1998 for the standard parking aid systems. Equally conceivable is an expansion of the function for supporting parking in traverse spaces.

A completely new application promises to be in the field of the side view assist (SVA) that also with the help of ultrasonic sensors covers the “blind spot” in a range of up to about 3 m immediately next to and at the sides behind the vehicle. Both other vehicles in the flow of traffic and those slowly overtaking can be detected by the driver's vehicle up to speeds of about 140 km/h and are displayed to the driver when initiating a maneuver to overtake. The necessary suppression of oncoming traffic and/or stationary obstacles, e.g., crash barriers, can thereby be realized by appropriately installing sensors in the corners of the front and rear bumpers. To this end the echo signals of both sensors are analyzed for occurrence with time, and the plausibility of the results obtained then checked.

Another interesting application is given from the further development of the ultrasonic transducer from a purely distance sensor to become an additional angle sensor. Based on the difference in the runtime of a reflected wave front between two directly adjacent transducer elements (dual sensor, consisting of transmitter/receiver and pure receiver), the direction of the reflecting object can be deduced directly (Ide et al. 2004). This information can be linked with the steering wheel angle so as to indicate a risk of collision to the driver during a parking maneuver – especially in the region of the vehicle corner – as a function of the steering wheel.

In conclusion it shall also be mentioned that besides reliable and robust ultrasonic sensors, the benefits and acceptance of all the functions described above presuppose both equally large efforts in the algorithmic signal processing and a mature display strategy. All three factors mutually matched to an optimum form the basis for the continual increase in market penetration and success in launching innovative additional functions of ultrasonic-based driver assist systems.

References

Bergmann L (1954) Der Ultraschall und seine Anwendung in Wissenschaft und Technik (Ultrasound and its application in science and technics). Hirzel-Verlag, Stuttgart

Ide H, Yamauchi H, Nakagawa Y, Yamauchi K, Mori T, Nakazono M (2004) Development of steering-guided park assist system. In: 11th world congress on ITS, Nagoya

ISO 17386 (2004) Manoeuvring aids for low speed operation (MALSO). International Standard Organization, Switzerland

Kutruff H (1988) Physik und Technik des Ultraschalls (Physics and technics of ultrasound). Hirzel-Verlag, Stuttgart

Lehfeldt W (1973) Ultraschall kurz und bündig (Ultrasound clear and brief). Vogel-Verlag, Würzburg

Robert Bosch GmbH (2004) Sicherheits- und Komfortsysteme (Safety and comfort systems). Vieweg-Verlag, Wiesbaden

Waanders JW (1991) Piezoelectric ceramics, properties and applications. Philips Components Marketing Communications, Eindhoven

Automotive RADAR

17

Hermann Winner

Contents

H. Winner (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_17

Abstract

RADAR sensors are used in many driver assistance systems. We could ask whether RADAR for automobiles is similar to RADAR used in aircraft or military applications. The answer would be yes and no: yes, because the basic physical principles are valid for all domains, and no, because the requirements are very different. Whereas sometimes the requirements are less ambitious, enabling new concepts to be implemented, in other aspects, the requirements are higher due to the more complex traffic environment. The fundamentals of RADAR technology laid out in this chapter give an understanding of how RADAR works in typical automotive applications and why principle limitations define the performance. At the end of the chapter, the current technology of automotive RADAR is demonstrated by examples from industry, including their specifications.

1 Introduction

RADAR (*ra*dio *d*etection *a*nd *r*anging) has its origins in the military technology of the Second World War and also remained bound to military applications for a long time. Its first use in the transport sector as a speed monitoring system led to rather negative experiences for many motorists. However, applications felt to be useful for drivers were also considered at an early stage as demonstrated by a magazine article (Fonck 1955) published in 1955. In the 1970s, an extensive research project was conducted with the aim of developing series-capable RADAR sensors for collision protection. Although this project sponsored by the German Federal Ministry of Education and Research advanced the development of RADAR, the time was not yet ripe for series production. The technical requirements that enabled the use of RADAR for driver assistance were not in place until some 20 years later. The first time a vehicle with RADAR was available was in 1998. The key function, however, was not collision warning but rather adaptive cruise control ACC (see ▶ Chap. 45, “Adaptive Cruise Control”) even though collision warning was incorporated as a functional part of this system. Further RADAR-based ACC systems followed at short intervals.

RADAR technology received another boost around 5 years later due to the development of the automatic emergency brake (see ▶ Chaps. 46, “Fundamentals of Collision Protection Systems,” and ▶ 47, “Development Process of Forward Collision Prevention Systems”) and lane change assistance (see ▶ Chap. 49, “Lane Change Assistance”).

There are currently four bands available for use in road traffic (24.0–24.25 GHz, 76–77 GHz, and 77–81 GHz in addition to a UWB band suitable for close range (see

Sect. 4.2) of 21.65–26.65 GHz). They are all used except for 77–81 GHz band. The 76.5 GHz range, which has been explicitly regulated for automotive RADAR and is available worldwide, dominates at present. In the meanwhile, the 24 GHZ range has also claimed a larger share of the market, especially with midrange sensors.

As in comparable cases of innovation, development of the first generation of RADAR for cars was the result of blood, sweat, and tears. However, remarkable things were achieved during development, in spite of many a disappointment in the battle for low costs. The costs were in the three-figure rather than the five-figure range. Nevertheless, huge efforts were made to reduce costs in the follow-up generations. The first trends towards technology convergence can already be seen. Nevertheless, there are still huge differences between the individual solutions, making it necessary to go into this in breadth in this chapter. Nor will it be possible to dispense entirely with calculations as RADAR cannot be understood without touching on the principles of telecommunication engineering. An attempt will nevertheless be made to present the theoretical considerations in a way that anyone with minimal previous knowledge can understand. So readers with good previous training in telecommunications may therefore be surprised not to find any condensed technical language and formulae. Reference is made to standard works (Skolnik 2008; Ludloff 2008) for the principles and definitions of RADAR used here; these standard works contain much more detailed observations about RADAR in general. As RADAR was previously at home in military and civil aviation and shipping, the topic of "automotive RADAR" has hardly been addressed until now. Consequently, this chapter specifically provides an overview of automotive RADAR technology, which boasts very different requirements and solutions compared to the areas of use referred to above (smaller distances, lower Doppler frequencies, high multitarget capability, small size, significantly lower costs).

2 Propagation and Reflection

When RADAR beams leave the sensor, this does not happen as a spherical wave with the same intensity in all directions but rather in a bunched manner. The antenna is responsible for this (see also Sect. 3). The so-called directive antenna gain G_D describes the ratio between the intensity $P(\phi, \vartheta)_\mathrm{max}$ in the solid angle of the strongest emittance and the value $P_\mathrm{total}/4\pi$ of a homogeneous isotropic radiator of identical total output $P_\mathrm{total} = \iint P(\phi, \vartheta)\mathrm{d}\phi\mathrm{d}\vartheta$. In this case, ϕ of the azimuth angles are in the horizontal plane and ϑ of the elevation angles are in the vertical plane. The greater the antenna gain is, the more tightly the rays are bunched. The effective antenna gain G also considers the antenna losses which are mainly line losses. The equivalent isotropically radiated power (EIRP) arising from the product of the total transmitting power and the antenna gain is the decisive variable for two criteria: firstly, for the radio license which depends on the output in the solid angle range of

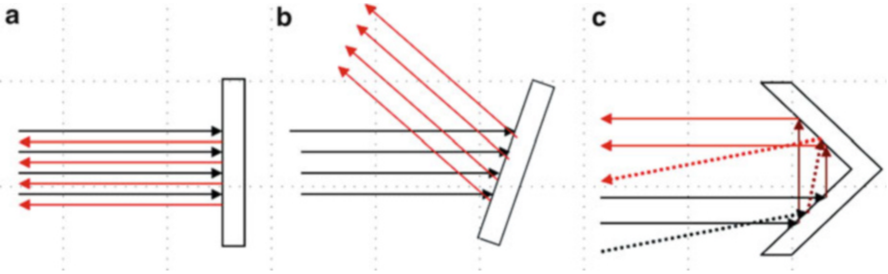


Fig. 1 Examples of directional reflection. (**a**) 90° reflection on a plate, (**b**) $\neq 90°$ reflection on a plate, and (**c**) 90° double mirror

the maximum (given in dBm (EIRP) where dBm relates to the base 1 mW) and, secondly, for the maximum achievable range.

Even more factors, however, have to be considered for the latter. The reflectivity of the RADAR target is quite obviously one of them. This is given as the so-called RADAR cross section (RCS) σ. In product with the square of the wave length, i.e., $\sigma\lambda^2$, the proportion reflected in a solid angle is described by the homogeneously distributed power entering the target. The unit of σ is an area. This area corresponds exactly to the central cross-sectional area πa^2 of a spherical reflector with the radius a. The relevant targets for automotive use in the medium and more distant range have values of $\sigma = 1 \dots 10{,}000\ \text{m}^2$. Significantly smaller values ($\sigma = 0.01 \dots 0.1\ \text{m}^2$) must be assumed if a pedestrian has to be detected at close range. The spread depends on the one hand on the type of target but is even more dependent on the geometry and the orientation. With greater distances, a metal plate oriented perpendicular to the direction of transmission and receiving has a back-scatter cross section of

$$\sigma_{\text{plate}} = 4\pi \frac{A^2}{\lambda^2} \tag{1}$$

Where $A = 1\ \text{m}^2$ and 76.5 GHz ($\lambda \approx 4\,\text{mm}$), the resultant RCS is $\sigma \approx 0.8 \times 10^6\ \text{m}^2$. Thus a box wagon with a flat rear of 4 m^2 may result in a strong backscatter with an RCS of $12.5 \times 10^6\ \text{m}^2$ (in the long range) but may collapse completely if rotated by one degree at a distance of approximately 60 m (cf. Fig. 1a and b). The remaining backscatter then originates only from the edges or the axle components. An ideal retroreflector is formed by three right-angled triangular surfaces that are perpendicular to each other, known as a corner (cube) reflector. With a perfectly oriented corner reflector, all incident radiation with a wave length that is significantly smaller than the dimensions is reflected back in the direction from which the radiation was emitted, as is illustrated in Fig. 1c for the two-dimensional case. For a three-dimensional corner reflector which consists of three equal-sided, right-angled triangles lying perpendicular to each other with an edge length a and the diagonal dimension $L = \sqrt{2}a$ in accordance with Fig. 2, an RCS of

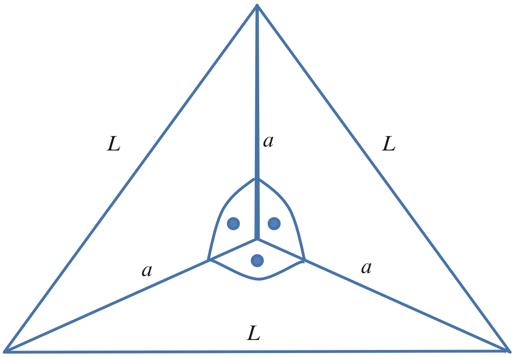


Fig. 2 Geometry of a corner cube reflector

$$\sigma_{\mathrm{CR}} = \pi \frac{L^4}{3\lambda^2} \Leftrightarrow L = \sqrt[4]{3\sigma_{\mathrm{CR}} \frac{\lambda^2}{\pi}} \tag{2}$$

is calculated according to Wolff (2014).

With a geometry such as this, it is possible even with small dimensions $(L = 35 \text{ cm})$ to simulate a very intense reflection of $\sigma_{\mathrm{CR}} \approx 1000 \text{ m}^2$ corresponding to a highly reflective truck. The following are deemed to be typical RADAR cross sections: 100 m^2 $(L = 20 \text{ cm})$ for a car, 10 m^2 $(L = 11 \text{ cm})$ for a motorcycle, and 1 m^2 $(L = 6.2 \text{ cm})$ for a person. In the ISO standards for ACC (TC208/WG14 2002) and FSRA (TC204/WG14 2008), a RADAR cross section of $10 \pm 3 \, \text{m}^2$ is specified for the detection field measurements where it is pointed out that these cover 95 % of vehicles. Small RADAR cross sections occur mainly for vehicles with flat or concave surfaces that reflect RADAR radiation away. High values are mainly attributable to corner reflectors. Thus the supporting posts of crash barriers with their u-profile display very high RADAR cross sections with the result that a very high number of these targets show up in the object list. As prismatic reflectors, the steps up to truck driver's cabs are also so highly reflective that they still carry sufficient signal power to the receiver even outside the main RADAR beam. On the one hand, the high dynamics of the RADAR cross section over four to five orders of magnitude means that classifying the object via the RADAR cross section will necessarily remain unsuccessful. On the other hand, the dynamics of the RADAR cross section increase the dynamic requirement on the receive path which should therefore not be below 70 dB and even then is not safe from clipping.

Besides the RADAR cross section, the radial distance r (range) affects the signal strength at the receiver. As already observed, the output in a solid angle element remains constant, at least if absorption losses are not taken into account. The area of this angular segment enlarges with the square of the distance; the same applies to the reflected beam, with the result that we can assume an r^{-4} drop-off for targets outside the close range. Only in a few cases is the absorption k in dB/km so high that it must also be taken into account. At 76.5 GHz, the atmospheric attenuation is below 1 dB/km and therefore only 0.3 dB for the return path to a target 150 m away.

However, an attenuation maximum of around 15 dB/km exists at 60 GHz (cf. ▶ Chap. 18, “Automotive LIDAR”). Although this attenuation leads to a slight decrease in the receiving power, it has the advantage that there is significantly less fear of overreaching than at 76.5 GHz and the RADAR beams would therefore “stray” less. As the bands around 60 GHz are used for military purposes in many parts of the world, this option was not available. Heavy rain, particularly with big raindrops that achieve the magnitude of the wave length, results in seriously strong attenuation which leads to a significant reduction in the range, while the visual range often exceeds that remaining for the driver. In addition to the attenuation effect, heavy rain results in an increased interference level (clutter). It mostly acts like an increased noise level and in this way decreases the signal-to-noise ratio (SNR) and, in turn, the achievable range. However, apparent targets can also be generated if, for example, the spray from a truck traveling alongside reflects the RADAR beams. Another interference effect of a “watery environment” occurs due to covering the guard (radome) of the beam exit area. Because of the high dielectric constant, water has a high refractive effect on mm waves with the result that uneven water coverage leads to undesirable “lens effects” that may severely distort the determination of the azimuth angle.

The last influence on the reception quality mentioned here is multipath propagation. This relates on the one hand to vertical multipath propagation via the reflection on the road surface. Regardless of polarization and wet road surfaces, reflection is almost total due to the ever decreasing grazing angle at longer distances (Schneider 1998). Thus the RADAR beams take routes of different lengths and therefore arrive at the receiver with different phases: If these phases differ by uneven whole number multiples of 180°, the interference is referred to as destructive; with multiples of 360°, it is referred to as constructive interference. The destructive interference occurs at specific gaps depending on the height of the RADAR and the reflection’s focal point above the road surface, with the result that the radar’s detection performance is noticeably affected. This is mostly not a problem because deflection and rebounding of the target vehicle or the driver’s vehicle, unevenness of the road, and elongation of the objects resulting in multiple reflections eliminate the interference hole and at finite relative velocity additionally destroy the interference gap condition associated with it. Vertical multipath reception is therefore expressed as a signal power “shaker” with the factor V_{mp}^2, $0 \leq V_{mp} \leq 2$, which means that we must basically expect a detection loss or drop-out rate during detection that can be described stochastically. With horizontal multipath propagation, there is reflection onto vertical surfaces, areas that are approximately parallel to the direction of travel. In addition to walls, it is mainly crash barriers that facilitate multipath propagation. In this case, the signal loss during negative interference is less disruptive than distortion of the azimuthal directional information. Scanner antennas (see Sect. 3.2) with narrow RADAR lobes react less sensitively to this than twin- or multibeam antennas; however, methods exist (Heidenreich 2012) for array antennas which determine the reflected path as a separate (ghost) target by assuming two targets and are thus able to reduce the effect on the main path.

If we observe the receiving amplitude over a longer distance range, it is possible to ascertain a harmonic periodicity due to a transformation in the reciprocal distance domain (i.e., $1/r$) whose "frequency" is indicative of the product of sensor height and the height of the target reflection point (Diewald et al. 2011; Diewald 2013) such that it is possible to distinguish between a bridge that can be driven under and an obstacle situated on the carriageway. The latter obstacle, however, may generate similar patterns via lateral reflectors (e.g., via crash barriers), because here, too, the product of the normal distances to the reflector plane may be of a similar size.

If we summarize the influencing factors described in this section, we can deduce the maximum range for a detection. The power P_R of the signal received is calculated as

$$P_\mathrm{R} = 10^{-2kr/1000} \cdot \sigma\lambda^2 \cdot G_\mathrm{t} \cdot G_\mathrm{r} \cdot V_\mathrm{mp}^2 \cdot P_\mathrm{total}/(4\pi)^3 r^4 \tag{3}$$

If the same antenna is used for transmitting and receiving, then the antenna gain for transmitting is equal to that for receiving, i.e., $G_\mathrm{t} = G_\mathrm{r}$ and $G_\mathrm{t} \cdot G_\mathrm{r} = G^2$. The signal received must be sufficiently higher than the electrical noise so that detection can take place. Depending on any other signal evaluation for flare suppression, the threshold is above the electrical noise (output P_N) by a factor $SNR_\mathrm{threshold}$ of approximately 6–10 dB.

The achievable range r_max is determined by equating the receiving power of Eq. 3 with the detection threshold $P_\mathrm{N} \cdot SNR_\mathrm{threshold}$. Disregarding the attenuation, i.e., $k = 0$, it can be calculated analytically:

$$r_\mathrm{max} = \sqrt[4]{\frac{\sigma\lambda^2 \cdot G_\mathrm{t} \cdot G_\mathrm{r} \cdot V_\mathrm{mp}^2 \cdot P_\mathrm{total}}{(4\pi)^3 \cdot P_\mathrm{N} \cdot SNR_\mathrm{threshold}}} \tag{4}$$

The range must be determined numerically in the case of finite attenuation. It is nevertheless easy to estimate the effect of the attenuation: if it is possible to assume an attenuation-free range of 200 m, then this decreases at 21 dB/km to 140 m ($(200/140)^4 \approx 6$ dB times (1 km/ (2 · 140 m)) ≈ 3.5), at 60 dB/km to 100 m and at 240 dB/km to 50 m. Therefore, basically all factors that determine the radar's theoretical range are known. In practical use, however, further limits are set by the signal processing as is described in the following section.

3 Measurement of Distance and Speed

To understand how RADAR works, we have to digress into the mathematics of telecommunications. From the author's point of view, the mathematical relationships derived in the following sections have been kept to a minimum and presented as far as possible in layperson's terms. Reference is made to standard works on

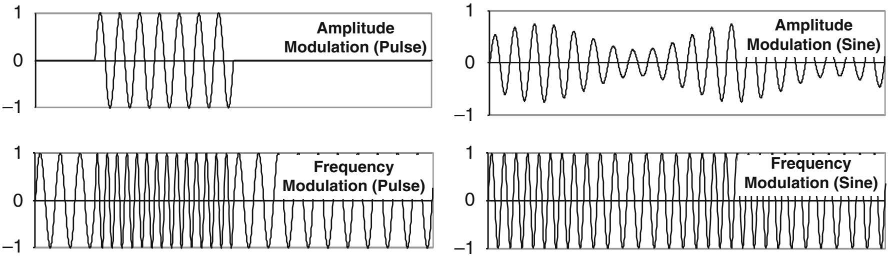


Fig. 3 Idealized modulation examples; *left*: modulated pulse, *right*: modulated sinusoidal signal; *top*: amplitude modulation, *bottom*: frequency modulation

RADAR, such as those by Skolnik (2008) or Ludloff (2008) for a more in-depth consideration of RADAR technology.

3.1 Basic Principle of Modulation and Demodulation

The emission and reception of electromagnetic waves is the only requirement necessary for RADAR to work. However, this creates nothing more than a carrier for the information. The information itself which is needed for measuring the distance has to be modulated to this carrier on the transmitter side and demodulated again on the receiver side. Simply put, the emitted wave train must be given an identifier for recognition and a time reference for measuring the time of flight. This task is referred to as modulation. Recognition and the determination of time patterns require demodulation.

In a general form, the radiation sent can be described as a harmonic wave function:

$$u_{\mathrm{t}}(t) = A_{\mathrm{t}} \cdot \cos\left(2\pi f_0 t + \varphi_0\right) \tag{5}$$

It is therefore possible to carry out modulations with the three variables: amplitude A, frequency f_0, and phase φ. For the purposes of illustration, the amplitude modulation (mainly as pulse modulation) and frequency modulation used for RADAR applications in cars are shown in Fig. 3 in an idealized form.

3.2 Doppler Effect

The Austrian Christian Doppler predicted in 1842 that an electromagnetic wave will undergo a frequency shift if the observer and transmitter move relative to each other. The same also happens if the RADAR beam is reflected by an object moved relative to the RADAR. Thus a RADAR beam to an arbitrary distance r and back again to the receiver travels a real figure z of $z = 2r/\lambda$ wave lengths in total.

Therefore a phase lag of $\varphi = -2\pi z$ arises. If r now changes with $\dot{r}$, then the phase also experiences a change of $\dot{\varphi} = -2\pi\dot{z} = -4\pi\dot{r}/\lambda$. Thus Eq. 5 for the received signal $u_{\mathrm{r}}(t)$ can be rewritten as follows:

$$u_{\mathrm{r}}(t) = A_{\mathrm{r}} \cdot \cos\left(2\pi(f_0 - 2\dot{r}/\lambda)t + \varphi_{\mathrm{r}}\right) \tag{6}$$

The Doppler effect is expressed as the frequency change f_{Doppler} which is proportional to the relative velocity and to the reciprocal value of the wave length $\lambda = f_0/c$ (speed of light c), where the frequency shift is positive when approaching ($\dot{r} < 0$) and negative when departing.

$$f_{\mathrm{Doppler}} = -2\dot{r}/\lambda = -2\dot{r}f_0/c \tag{7}$$

Note: In addition to the phase shift due to the time of flight, phase rotation also takes place during reflection. With ideal total reflection, such as can be assumed for metals, this amounts to π, as is the case during inversion. However, this detail is virtually unimportant as the absolute phase is not used in any evaluation. Only the differences are used.

With a carrier frequency of 76.5 GHz, due to the relative velocity $\dot{r}$ in SI units (i.e., in m/s), we obtain a Doppler shift of $f_{\mathrm{Doppler}} = -510\ \mathrm{Hz} \cdot \dot{r}$ or, at the other frequency of 24 GHz customary for driver assistance applications, approximately a third of this, that is to say, $f_{\mathrm{Doppler}} = -161\ \mathrm{Hz} \cdot \dot{r}$. The values must be divided by 3.6 for the calculation with km/h. With an assumed relative velocity of −70 m/s (−252 km/h) on approach, the maximum Doppler frequencies amount to 35.7 kHz with the result that for a measurement according to the Nyquist theorem, a sampling rate of at least 71.4 kHz is required for unambiguous determination.

Basically, the relative velocity information can already be determined with a continuous wave of constant frequency. However, the carrier frequency is too high for directly measuring the shift in the carrier band which even at maximum relative velocity is only just a millionth of the carrier frequency. In reality, by mixing as described in the following section, it is possible to measure at much lower frequencies.

3.3 Mixing of Signals

The process of signal multiplication is described as mixing in high-frequency technology. The product of two harmonic signals $u_1(t)$ and $u_2(t)$, described similarly to Eq. 5 with the cosine function, of the frequencies f_1 and f_2 and phases φ_1 and φ_2 can also be described by the addition theory of harmonic functions according to Eq. 8 as the sum of two harmonic functions, each with the difference or the sum of the original arguments:

$$\cos x \cdot \cos y = \frac{1}{2}\{\cos(x - y) + \cos(x + y)\} \tag{8}$$

Thus the product of the transmitted signal (Eq. 5) and the received signal (Eq. 6) becomes the frequency conversion product $u_{\mathrm{t,r}}(t)$:

$$u_{\mathrm{t,r}}(t) = \frac{1}{2}A_{\mathrm{t}}A_{\mathrm{r}}\left\{\cos\left(2\pi\left(\frac{2\dot{r}}{\lambda}\right)t + \varphi_0 - \varphi_{\mathrm{r}}\right) + \cos\left(2\pi\left(2f_0 - \frac{2\dot{r}}{\lambda}\right)\mathrm{t} + \varphi_0 + \varphi_{\mathrm{r}}\right)\right\} \tag{9}$$

As the sum signal (the second term) has very high frequency, this fraction is simply eliminated by the electronics (cables, amplifier) which are not designed for this frequency. As a result, the low-frequency difference signal is left over

$$u_{\overline{\mathrm{It,r}}}(t) = \frac{1}{2}\mathrm{A_r A_t}\cos\left(2\pi\left(\frac{2\dot{r}}{\lambda}\right)t + \varphi_0 - \varphi_{\mathrm{r}}\right) \tag{10}$$

The information regarding the frequency shift is found in the cosine argument. However, it is not the argument that is measured but rather the cosine function which has no clear inverse function. This means that the mathematical sign in particular is not accessible since a cosine function with a positive frequency is identical to that with negative frequency. Here, it helps to mix with a signal shifted by 90° in relation to the transmission signal, that is to say, a multiplication; instead of multiplying with the original cosine function, the sine function associated with the transmission signal is now used.

$$\sin x \cdot \cos y = \frac{1}{2}\{\sin(x-y) + \sin(x+y)\} \tag{11}$$

Thus, after suppression of the sum signal, a mixed signal described on the basis of the sine is available:

$$u_{\overline{\mathrm{Qt,r}}}(t) = \frac{1}{2}A_{\mathrm{r}}A_{\mathrm{t}}\sin\left(2\pi\left(\frac{2\dot{r}}{\lambda}\right)t + \varphi_0 - \varphi_{\mathrm{r}}\right) \tag{12}$$

Although the sine function is an uneven function, this is just as inadequate as the cosine mixed signal (Eq. 10) for distinguishing whether a negative or positive Doppler shift is the cause of the difference frequency.

However, if both signals are generated, it is possible to find the uniqueness by comparison with each other: with a positive Doppler frequency, corresponding to an approach, the directly derived signal (index I: in-phase = real part) compared to that from the second signal shifted by 90° phase (index Q: quadrature = imaginary part) also exhibits the 90° phase, but in the case of a negative Doppler frequency, this is shifted by $-90°$:

$$2\pi\left(\frac{2\dot{r}}{\lambda}\right)t + \varphi_{\mathrm{r}} = \arctan\left(\frac{u_{\overline{\mathrm{Qt,r}}}(t)}{u_{\overline{\mathrm{It,r}}}}(t)\right) \tag{13}$$

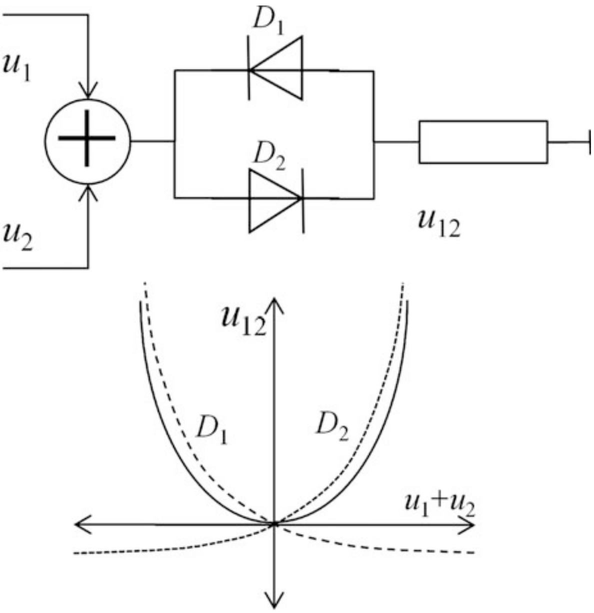


Fig. 4 Schematic diagram of a two-diode sum mixer

Even with a vanishing Doppler frequency, it is possible with Eq. 13 to determine a differential phase. This presupposes, however, that the signals $u_{\overline{\mathrm{Qt,\,r}}}(t), u_{\overline{\mathrm{It,\,r}}}(t)$ do not contain any other direct components.

As described up to this point, mixing is a simple and straightforward mathematical action. Digital multiplication is eliminated for technical implementation because affordable analog/digital converters are too slow for the RADAR frequencies used in cars. Multiplication using analog multipliers is also only possible to a limited extent at these frequencies (see below). Fast nonlinear components such as Schottky diodes (metal/semiconductor transition), however, permit the so-called sum mixing. For this, the two signals to be mixed are first additively superimposed as shown in Fig. 4. The voltage sum $u_1 + u_2$ leads to a current that can be measured via the resistance as voltage drop u_{12}.

The characteristic of the two diodes can be developed individually and also in sum as a Taylor series. With the dual-diode arrangement presented here, the uneven terms vanish in the ideal case so that the following terms remain:

$$u_{12} = A_2(u_1 + u_2)^2 + A_4(u_1 + u_2)^4 + \ldots; \quad A_n = \frac{\partial^n}{n!\partial u^n} D(u) \tag{14}$$

$$\begin{aligned} u_{12} = {} & A_2\left(u_1^2 + 2u_1 \cdot u_2 + u_2^2\right) \\ & + A_4\left(u_1^2 + 4u_1^3 \cdot u_2 + 6u_1^2 \cdot u_2^2 + 4u_1 \cdot u_2^3 + u_2^4\right)^4 + \ldots \end{aligned} \tag{15}$$

The desired product $u_1 \cdot u_2$ is found in the multiplied quadratic fraction. Virtually all other frequency conversion products lead to high-frequency signals (just as the uneven ones do should there be no symmetry). Only the product terms with the same exponent (e.g., $u_1^2 \cdot u_2^2$) deliver contributions to a lower frequency signal and as harmonics may lead to distortions, in particular false detections (false-positive errors). Therefore, the even fractions of the Taylor expansion must be designed as small as possible with powers higher than two.

Active mixers with the so-called Gilbert cell already come very close to the ideal multiplier. With sufficiently fast field-effect transistors, the two input signals can be multiplied with each other because the oscillator voltage is used as the control voltage for the amplification of the other received signal. Silicon technology is no longer adequate for the frequency range 76–77 GHz. Gallium-arsenide (GaAs) must be used instead or, in recent times, the more reasonably priced silicon-germanium (SiGe) technology. Compared to passive mixers, the conversion losses arising during mixing are lower, resulting in a higher signal-to-noise ratio.

3.4 Pulse Modulation

3.4.1 Requirements for Pulse Duration and Bandwidth

Pulse modulation is the easiest to visualize (cf. Fig. 3 top left). In this case, a short wave train of pulse length τ_P is formed. Technically, this is implemented by means of a fast electronic switch which is supplied by a continuously operated oscillator. Such an ideal pulse requires a bandwidth reciprocal to the pulse length even if the oscillation within the pulse corresponds exactly to Eq. 5. In reality, the signal arises from multiplying a flat wave according to Eq. 5 and a window function which is described for an ideal pulse, switching rapidly on and off, of

$$F_{\mathrm{Rect}}(t) = 1 \text{ für } |t - t_0| < \tau_{\mathrm{P}/2},\ 0 \quad \text{sonst} \tag{16}$$

as a rectangular window about the pulse center t_0. This leads in the frequency range to convolution of the discrete frequency line f_0 with the Fourier transform of the window function known as the sinc function $\mathrm{sinc}(\pi f \cdot \tau_P) = \sin(\pi f \cdot \tau_P)/\pi f \cdot \tau_P$. As the sinc function falls away weakly (amplitude envelope with f^{-1}), a major portion of the pulse power falls in frequency bands that are intended for other applications.

Although the ratio of in-band to out-of-band power can be improved by lengthening the pulse, nevertheless the measure does not reduce the energy per pulse scattered in the other bands unless the pulse rise or fall is lowered. On the other hand, the very steepness at the beginning and end makes it possible to differentiate the time of flight. The entire output between rise and fall is largely useless for distance measurement. A pulse envelope shape according to a cosine bell, which is also known in signal processing as a von Hann or Hanning window, is a good compromise:

$$F_{\mathrm{u}}(t) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi \cdot t}{\tau_P}\right)\right) \text{ for } |t - t_0| < \tau_{\mathrm{P}/2},\ 0 \text{ else} \tag{17}$$

Although a pulse shaped in such a way loses 5/8 of the power compared to a square envelope with the same maximum amplitude, what remains is concentrated almost completely in the working band between

$$f_0 - \tau_P^{-1} < f < f_0 + \tau_P^{-1}; \quad \Delta f = 2\tau_P^{-1} \tag{18}$$

The required bandwidth $2\tau_P^{-1}$ of a pulse therefore corresponds to double the reciprocal value of the total pulse length. A further advantage of the band limitation, in addition to adhering to limit values, is the possibility of bandpass filtering on the receiving side which is useful for noise reduction. This is because the receiver bandwidth should be at least as large as the emission bandwidth so that no loss of resolution in the time of flight occurs as a consequence of receiving.

How short or rather how sharply restricted should a RADAR pulse be for use in driver assistance systems? For a long-range RADAR (LRR), at least two vehicles should appear separate at typical distances. Therefore, the pulse should have a length X_P of no more than 10 m or a corresponding maximum duration of $\tau_P = X_P/c \approx 33$ ns. When using the RADAR as short-range RADAR (SRR) with the capability of parking assist, a spatial resolution of 15 cm is required which is why the pulse should be no longer than double this, that is to say, $X_P \approx 30\ \text{cm}$, and consequently, the pulse duration should correspondingly be no longer than $\tau_P \approx 1\ \text{ns}$. Thus, the bandwidth requirements are at least 60 MHz for LRR and 2 GHz for SRR. These estimates are best-case considerations and must be increased by a factor of approximately 2 for practical purposes in order to exclude violation of the frequency band.

The unfavorable ratio of maximum output to mean output is one disadvantage of pulse modulation. To improve the signal-to-noise distance and thus to increase the sensitivity, pulse sequences, via which averaging is carried out, are “fired off.” Although the pulses can also be transmitted at shorter time intervals via so-called pseudorandom sequences, this does require very elaborate input electronics. It is easier to wait until we can rule out the possibility that a pulse from an earlier transmission can still be received. For this, a multiple of the maximum useful time of flight should be used (for LRR, this can be specified as 1 μs at a distance of 150 m, for SRR approximately 0.1–0.2 μs). This results in a pulse sequence frequency for SRR of approximately 1 MHz and for LRR of approximately 250 kHz.

3.4.2 Noncoherent Demodulation

Simple demodulation could be carried out similarly to the noncoherent demodulation used with ultrasonic sensors or LIDAR. The signal received is amplified as illustrated in Fig. 5, filtered by the carrier frequency f_0 by a bandpass corresponding to the pulse bandwidth. Then, rectification is carried out so that a direct component corresponding to the amplitude is formed from the alternating voltage, this direct component being available as an output signal in the subsequent low pass. The signal obtained is then sampled and compared in the microprocessor or directly with specified threshold values in the comparator, as show in Fig. 5 (block 7). This demodulation technique can be easily disrupted with external pulses and can only

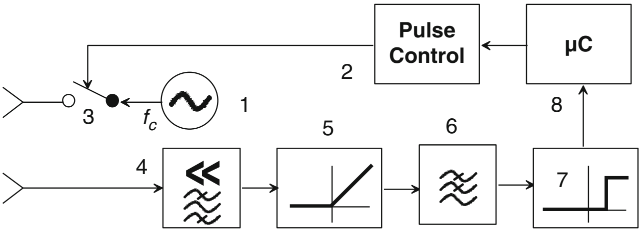


Fig. 5 Block diagram of a noncoherent RADAR; *1* oscillator, *2* pulse control, *3* pulse modulator, *4* amplifier and bandpass filter, *5* rectifier, *6* low pass filter, *7* comparator, *8* microprocessor

carry out time of flight measurements without, however, being able to use the Doppler effect that is very important for further signal processing.

3.4.3 Coherent Pulse Demodulation

The principle of the mixer is used in coherent pulse demodulation (also referred to as the pulse Doppler method). However, in this case, mixing is not directly down to the so-called baseband (which is around frequency 0) but rather an intermediate frequency is generated. This can be achieved either by a local oscillator which has a fixed frequency difference to the transmission signal or by the same oscillator if its frequency after transmitting the pulse is changed by a specific frequency difference. The intermediate frequency is around 100–200 MHz. Amplifiers, filters, and ADC can be implemented in this range with justifiable expense. In addition, it is still possible to map the pulse shape. The intermediate frequency can be sampled directly with an AD converter.

The real and the imaginary parts are formed in each case from the intermediate frequency, as described above. If the signal pair illustrated in Fig. 6 is sampled with 10 ns cycle time, for every sampling instance, we obtain a value pair that can be interpreted as the coordinates of a vector in a complex plane. With a later measurement (t_i), these vectors are rotated further by an angle $2\pi t_i(2\dot{r}/\lambda)$ according to Eq. 13 (see Fig. 7). The absolute value of the vectors represents the pulse intensity at the time-of-flight $t_{of} = t_{PC} - t_S$ specified with the sampling instance t_S, in relation to the time t_{PC} of the pulse center. This time of flight equates to the distance

$$r = \frac{1}{2} c \cdot t_{of},\ c: \quad \text{speed of light}, \tag{19}$$

such that the significance of so-called range gates is according to the individual sampling instances. If the signal arises, as in the example of Fig. 7, due to the reflection of the same object, the rotational speed of the vectors is identical because all exhibit the same Doppler shift. The range gates (and therefore the sampling cycle) should, commensurate with the pulse width, be so close that it is possible to create a center point and thus interpolate the distance, which may lead to distance

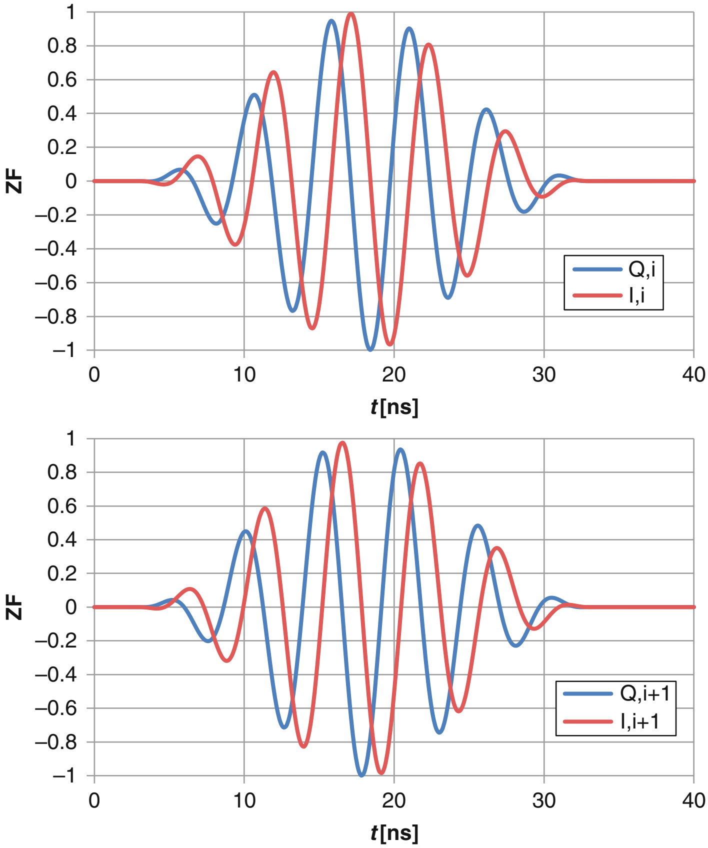
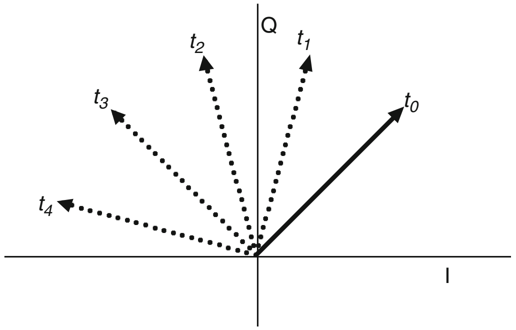


Fig. 6 Intermediate frequency signals (real part I and imaginary part Q) of two successive pulses (*top*, *bottom*) of an approaching single reflector (idealized)

Fig. 7 Rotation of the pointer brought about by Doppler shift in the complex Q/I plane



resolutions of significantly less than a tenth of the pulse length. To achieve this, the range gates should be no more apart than half the pulse length. Significantly shorter range gates are avoided for reasons of cost as the higher sampling frequency associated with them makes the ADC more expensive without actually achieving information of any higher quality.

It is necessary to repeat the pulses for two reasons. On the one hand, a single pulse contains only low energy, meaning that to increase the signal-to-noise ratio, repetition is both more cost-effective and also less critical in respect to the frequency license than increasing the pulse power. On the other hand, the Doppler frequency should be clearly sampled which results in at least one pulse repetition of 71.4 kHz according to Sect. 2.2. The overall length T_M of the pulse trains leads to resolution of the Doppler frequency of

$$\Delta f_\mathrm{Doppler} = \frac{1}{T_\mathrm{M}}, \tag{20}$$

and, therefore, to the relative velocity resolution of

$$\Delta \dot{r} = \frac{c}{2 f_0 T_\mathrm{M}}. \tag{21}$$

As a result, a measuring time of approximately 2 ms is required for $\Delta \dot{r} = 1$ m/s at 76.5 GHz. With an exact pulse repetition, it is possible for both phantom targets to arise due to overshooting and also for interference radiation to occur due to other RADAR sensors. A pseudorandom variation of the pulse repetition times can solve this problem (cf. Kühnke 2003), i.e., the follow-on pulse varies compared to the average cycle time by at least the duration of one range gate, so that the interference or overshoot falls in a different range gate when the pulse is repeated.

It is basically possible to measure even small distances below the pulse length using coherent pulse length modulation if the receive path is also available simultaneous to pulse transmission. If, unlike as shown in Fig. 5, the same antenna and the same oscillator are selected for the transmit and receive path, then it is not possible to switch to receive until the transmit pulse is complete. As a result, the full pulse cannot be observed for object distances up to half a pulse length. However, as parts of the pulse are still detected, it is at least possible to identify the object's presence within this zone, and while the distance cannot be determined, it is possible to determine the relative velocity as this can be determined in all areas of the pulse.

The strength of coherent pulse demodulation is the independent measurement of the distance and the relative velocity which is managed with a low average transmission power compared to other methods. Unfavorable aspects are the high receiver bandwidth required, which means that this principle is more easily disrupted than the methods described below, and the considerable effort required for the switching elements.

3.5 Frequency Modulation

In frequency modulation, the frequency f_0 is varied as a function of time, though it must be made clear that this is not an absolute and therefore constant frequency but rather a instantaneous frequency $f_0(t) = \omega_0(t)/2\pi$. In this chapter, frequency modulation is related to all methods in which the information about the time of flight is achieved by frequency variation.

Figure 8 shows the basic structure of FM RADAR. It is imperative for the manner of operation that the instantaneous frequency is varied by means of a voltage-controlled oscillator which enables the desired modulation via a control loop (e.g., phase-locked loop, PLL). The received signal is mixed with the signal currently being transmitted, filtered, sampled, and converted. Optionally, it is possible to use spatially separated supply lines for separating the signal of the

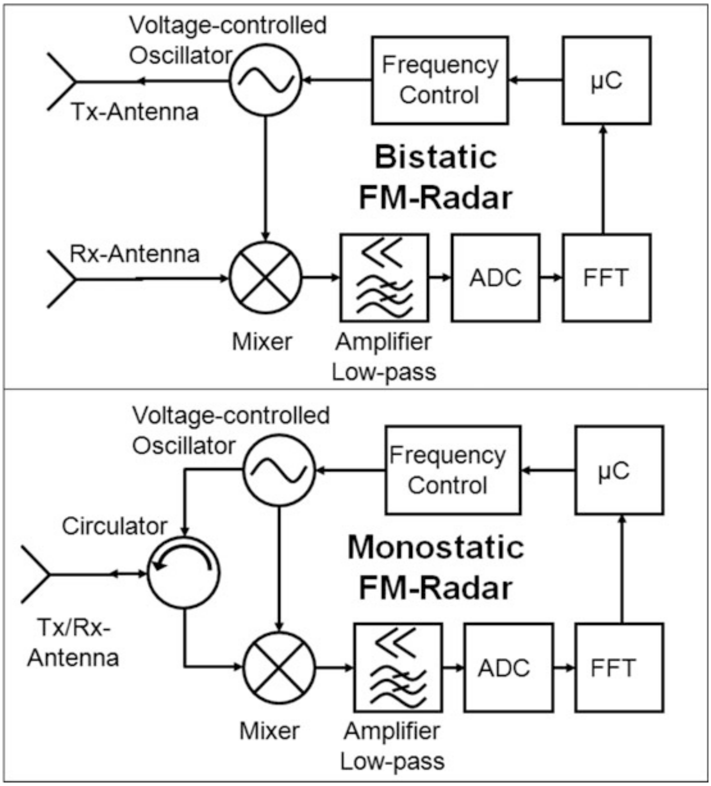


Fig. 8 Block diagram of a RADAR with frequency modulation. *Top*: in a bistatic version with separate antenna leads for transmit and receive beam; *bottom*: in monostatic version with circulator coupling

transmit path and receive path (Fig. 8 top) or special nonreciprocal couplers (Fig. 8 bottom) which couple direction specifically.

3.5.1 Frequency Shift Keying (FSK)

In frequency shift keying, the instantaneous frequency of the signal is varied in steps. In the simplest variant, two wave trains of length Δt with the instantaneous angular frequency ω_1 and ω_2 are transmitted one after the other, and the received signal is simultaneously mixed with a signal derived from the transmitted signal. According to Eq. 10, the following baseband frequency conversion products arise:

$$\overline{u}_{\mathrm{lt,r,i}}(t) = \frac{1}{2} A_r A_t \cos\left(\frac{\omega_{\mathrm{i}}}{c} 2\dot{r}t + \varphi_0 - \varphi_{\mathrm{r,i}}\right),\ i = 1,\ 2 \tag{22}$$

In this equation, $2\pi\lambda$ was substituted for ω_{i}/c with the result that the effects brought about by the frequency change become apparent. For simplification, it is initially assumed that there is no Doppler effect, that is to say, the detected object exhibits no relative velocity $\dot{r}$. The result, depending on the distance, is a phase change of

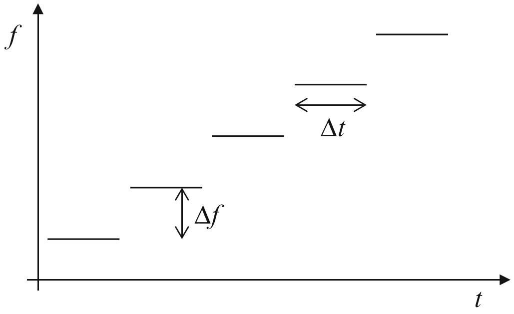
$$\Delta\varphi_{\mathrm{r,i}} = \varphi_0 - \varphi_{\mathrm{r,i}} = t_{\mathrm{of}} \cdot \omega_{\mathrm{i}} = \frac{2r}{c} \omega_{\mathrm{i}},\ \mathrm{i} = 1,\ 2, \tag{23}$$

and, therefore, in the differential consideration

$$\Delta\varphi_{\mathrm{r,2}} - \Delta\varphi_{\mathrm{r,1}} = t_{\mathrm{of}} \cdot \Delta\omega = \frac{2r}{c} \Delta\omega, \Delta\omega = \omega_2 - \omega_1. \tag{24}$$

The greater the phase difference, therefore, the longer the distance r and the higher the difference angular frequency. However, here too it is true that a phase is not clearly identifiable. Initially, only cosine values multiplied by the amplitude are measured, two values in this case. I/Q mixing, as shown in Sects. 2.3 and 2.4.3, would be helpful in this case but would also considerably increase the costs for the demodulation hardware. Alternatively, it is possible to make an initial statement about the distance with further jumps in frequency similar to Fig. 9. For the cosine arc to be identified as such, the n steps must together bring about a phase change of

Fig. 9 Principle of frequency shift keying (FSK) with several steps



at least 45° ($\pi/4$). Thus, the total travel of the frequency steps $n \cdot \Delta f$ is determined from the minimum measurable distance r_{min} as

$$\Delta\varphi_{\mathrm{r,n}} - \Delta\varphi_{\mathrm{r,1}} = t_{\mathrm{of}} \cdot n\Delta\omega = \frac{2r_{\mathrm{min}}}{c} n\Delta\omega = \frac{\pi}{4} \Rightarrow n\Delta f \geq \frac{c}{16r_{\mathrm{min}}} \tag{25}$$

This leads to a travel $n \cdot \Delta f$ of 625 kHz at 30 m or 18.75 MHz at 1 m. These values can serve as a starting point for the minimum bandwidth necessary for distance measurement. The number of steps results from the unambiguous criterion at the maximum postulated object distance r_{max}. Thus, the phase change between two steps must not be greater than 180° (π).

$$\Delta\varphi_{\mathrm{r,i+1}} - \Delta\varphi_{\mathrm{r,i}} = t_{\mathrm{of}}\Delta\omega = \frac{2r_{\mathrm{max}}}{c}\Delta\omega = \pi \Rightarrow \Delta f \leq \frac{c}{4r_{\mathrm{max}}} \tag{26}$$

This results in step heights Δf of maximum 188 kHz at 400 m. Although this distance value lies outside the distance target ranges considered, it cannot be ruled out that highly reflective objects will also be detected from this distance range. The minimum number of steps n_{min} results from the ratio $r_{\mathrm{max}}/r_{\mathrm{min}}$ of maximum to minimum distance:

$$n_{\mathrm{min}} = \frac{r_{\mathrm{max}}}{4r_{\mathrm{min}}} \tag{27}$$

All the statements continue to be maintained if we extend the consideration above to objects moving relative to the FSK RADAR. However, the signal of the individual step is not a direct signal but rather varies according to Eq. 22 with the Doppler frequency:

$$f_{\mathrm{Doppler,i}} = \frac{-f_{\mathrm{i}}}{c} 2\dot{r}. \tag{28}$$

Although the Doppler frequency of each step differs due to the varying fundamental frequency f_{i}, the changes are so small $\left(<10^{-5}\right)$ that in a Fourier analysis, the Doppler frequencies fall in the same frequency cell. Nevertheless, phase shifts can add up due to the differences although these can be predetermined and therefore also compensated.

In principle, the objects can be detected on the basis of the Doppler frequency alone, although the mathematical sign of the Doppler shift is not known. This can be derived from the phase difference between the steps for the Doppler signals found. If the phase increases when the transmit frequency rises, this indicates a positive Doppler frequency, that is to say, an approaching object. If, by contrast, the phase decreases, the only reasonable explanation is a negative Doppler frequency as a negative distance can be ruled out.

Resolution of the relative velocity depends only on the measuring time available for a step. If the steps are carried out one after another as described above, only a

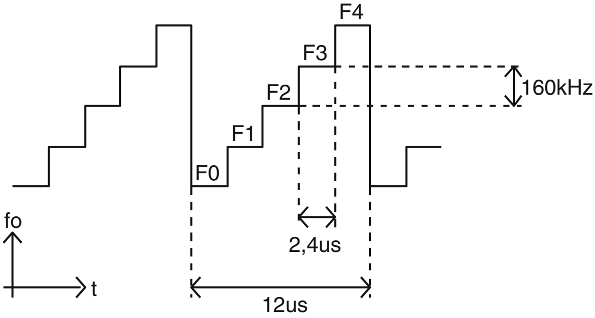


Fig. 10 FSK with five nested frequency steps (Source: TRW)

measuring duration of $T_{\mathrm{M}} = T/n$ is available for an overall measuring time T per step. When there are many steps, this leads to a considerable deterioration in the relative velocity resolution. When only a few steps are to be measured, it is beneficial to make use of the fact that the necessary sampling rate for the Doppler effect is so low that measurements can be carried out with other transmit frequencies in the measuring pauses between two sampling instances. In Sect. 2.2, a minimum sampling rate of 71.4 kHz was determined; hence, the pause is almost 14 μs. By contrast, the time of flight for an object 300 m away is only 2 μs. Theoretically, another six measurements could be squeezed in, although in practice, another four could be squeezed in as shown in a practical example in Fig. 10. The signals correspond to a step function in which the values for the same step height are combined into an analysis dataset for evaluation. In this way, it is not necessary to allocate the measuring time to the various steps, as a T-long dataset is evaluated for all steps and therefore according to Eq. 21 the result is a relative velocity resolution of $\Delta\dot{r} = c/2f_0T$ corresponding to $\Delta\dot{r} = (1\ \mathrm{m/s})/(510\ \mathrm{Hz} \cdot T)$ at 76.5 GHz. A velocity cell of approximately 1/20 m/s can be obtained with a measuring duration of 40 ms. This makes it possible to separate objects that exhibit three-cell difference, that is to say, speed differences of only 3/20 m/s or around 0.5 km/h. This high separation capability, however, is necessary with such a method because, due to the small frequency deviation, there is no separation capability regarding the distance. Thus if several objects have the same relative velocity such that they are arranged in the same relative velocity cell, it is no longer possible to recognize that more than one object is present. The distance value determined in such a case is very unreliable, although the strongest reflector in terms of value dominates the others. In the case of moving objects, it is highly unlikely that several objects will fall into the same cell together. By contrast, with stationary objects, this is always the case if their radial speed $\dot{r}$ cannot be distinguished by a different azimuthal approach angle ϕ of the RADAR vehicle moving at a driving speed v, if, therefore,

$$\left|\dot{r}_{\mathrm{i}} - \dot{r}_{\mathrm{j}}\right| < \Delta\dot{r} \Leftrightarrow v\left|\cos\phi_{\mathrm{i}} \cdot \cos\phi_{\mathrm{j}}\right| < c/2f_0T \tag{29}$$

applies. At a speed of $v = 10\ \text{m/s}$ of the RADAR vehicle ($f_0 = 6.5\ \text{GHz},\ T = 40\ \text{ms}$), all stationary obstacles within an azimuthal visual range of ±5.6° fall in the same speed cell as the stationary obstacles on the center line. Hence, such a method is unsuitable for detecting stationary obstacles.

Condensing several frequency steps enables even more signal improvement measures. Thus, the sampling instance for the received signal can be placed at the beginning of the step with a defined delay so that overshoots of objects with longer time of flights than this delay time can be excluded.

3.5.2 FMSK

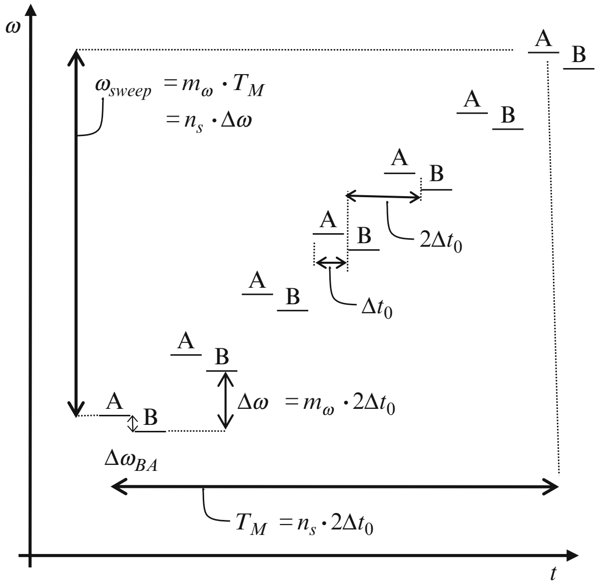
Another modulation based on frequency stairs is known as linear frequency modulation shift keying (LFMCW/FSK) (Meinecke and Rohling 2000). It is illustrated in Fig. 11. A frequency stair A with n_S steps is followed by a frequency stair B offset in time and frequency. For stair A, similar to Eqs. 22 and 23, a mixed signal arises:

$$u_{\overline{\text{lt,r,i,A}}}(t_{i,A}) = \frac{1}{2}A_\text{t}A_\text{r}\cos\left(\frac{2\omega_\text{i}}{c}\dot{r}t_\text{i,A} + \frac{2r}{c}\omega_\text{i,A}\right),\ i = 1,\ \ldots,\ n \tag{30}$$

where

$$\omega_\text{i,A} = \omega_\text{0,A} + i_\text{A}\cdot\Delta\omega = \omega_\text{0,A} + i_\text{i,A}\cdot m_\omega,\ t_\text{i,A} = t_0 + 2i\Delta t_0;\ i = 1,\ \ldots,\ n_\text{S} \tag{31}$$

Fig. 11 Frequency curve for linear frequency modulation shift keying (LFMCW/FSK) according to Meinecke and Rohling (2000)



having the sampling instances $t_{\mathrm{i,A}}$ and the stair slope of the angular frequency $m_{\omega} = \Delta\omega/(t_{\mathrm{i+1}} - t_{\mathrm{i}})$. Reinserted in Eq. 30,

$$u_{\overline{\mathrm{lt,r,i,A}}}(t_{\mathrm{i,A}}) = \frac{1}{2} A_{\mathrm{t}} A_{\mathrm{r}} \cos\left(\left(\frac{2\omega_{\mathrm{i}}}{c}\dot{r} + \frac{2m_{\omega}}{c} r\right) t_{\mathrm{i,A}} + \frac{2r}{c}\omega_{0,A}\right), \ i = 1, \ \ldots, \ n. \quad (32)$$

Similarly, the same result is obtained for the second stair, replacing index A for B. It should be noted that the sampling instances $t_{\mathrm{i,B}} = t_0 + \Delta t_0 + 2i\Delta t_0$ for $t_{\mathrm{i,A}}$ are offset by Δt_0 and the starting angular frequency $\omega_{0,\mathrm{B}}$ differs by $\Delta\omega_{\mathrm{BA}}$ for $\omega_{0,\mathrm{A}}$. For both cases, we obtain a discrete-time data series which, after the Fourier transform at the same angular frequency

$$\omega_{\mathrm{obj}} = \frac{2}{c}(m_{\omega} r + \omega_0 \dot{r}) \quad (33)$$

delivers a (complex) amplitude. The approximation of the prefactor for the Doppler frequency, the carrier frequency ω_{i}, with the starting frequency ω_0, which was carried out for simplification, only leads to errors $(\omega_{\mathrm{i}} - \omega_0)/\omega_{\mathrm{i}}$ in the per thousand range with modulation deviations of 100 MHz and carrier frequency 76.5 GHz. In both stairs, there is an amplitude of the same value at ω_{obj} but with a different phase:

$$\Delta\varphi_{\mathrm{BA}} = \frac{2}{c}(\Delta\omega_{\mathrm{BA}} r + \omega_0 \Delta t_0 \dot{r}), \quad (34)$$

from a speed-dependent portion due to the time offset and additionally from a distance-dependent portion due to the frequency offset. Both sets of information, the frequency of the signal (Eq. 33) and the phase difference between the complex amplitudes of both stairs (Eq. 34), are a linear combination of relative velocity and distance and can accordingly be represented as straight lines in each case in a $\dot{r} \div r$ diagram (see also Fig. 12):

$$\dot{r} = \frac{c}{2} \cdot \frac{\omega_{\mathrm{obj}}}{\omega_0} - \frac{m_{\omega}}{\omega_0} r, \quad (35)$$

$$\dot{r} = \frac{c}{2} \cdot \frac{\Delta\varphi_{\mathrm{BA}}}{\omega_0 \Delta t_0} - \frac{\Delta\omega_{\mathrm{BA}}}{\omega_0 \Delta t_0} r. \quad (36)$$

As long as the second stair does not lie exactly in the middle of the first stair, that is, $m_{\omega}\Delta t_0 \neq \Delta\omega_{BA}$, there is an intersection point of both straight lines which makes it possible to clearly determine both the distance and also the relative velocity:

$$r = \frac{c}{2} \cdot \frac{\Delta t_0 \cdot \omega_{\mathrm{obj}} - \Delta\varphi_{\mathrm{BA}}}{m_{\omega} \cdot \Delta t_0 - \Delta\omega_{\mathrm{BA}}} \quad (37)$$

$$\dot{r} = \frac{c}{2\omega_0} \cdot \frac{m_{\omega} \cdot \Delta\varphi_{\mathrm{BA}} - \Delta\omega_{\mathrm{BA}} \cdot \omega_{\mathrm{obj}}}{m_{\omega} \cdot \Delta t_0 - \Delta\omega_{\mathrm{BA}}} \quad (38)$$

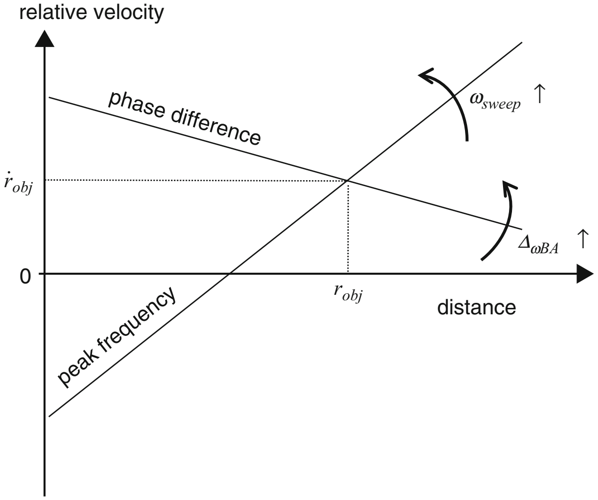


Fig. 12 Determination of the distance and relative velocity with the linear frequency modulation shift keying method (LFMCW/FSK) according to Meinecke and Rohling (2000)

As the duration of the stairs determines the measuring time $T_{\mathrm{M}} = 2n\Delta t_0$, it is possible, according to Eq. 21, to specify a relative velocity cell of

$$\Delta\dot{r} = \frac{c}{4f_0 n_{\mathrm{S}} \Delta t_0} \tag{39}$$

The distance resolution also depends on the measuring duration as the distance resolution is also determined via the frequency resolution in accordance with Eq. 33. However, the measuring time is cut again if the overall frequency deviation $f_{\mathrm{sweep}} = m_\omega T_{\mathrm{M}}/2\pi$ is used instead of

$$\Delta r = \frac{c}{2} \cdot \frac{\omega_{\mathrm{obj}}}{m_\omega} = \frac{c}{2} \cdot \frac{2\pi/T_{\mathrm{M}}}{m_\omega} = \frac{c}{2f_{\mathrm{sweep}}} \tag{40}$$

This expression also applies without constraint to other methods and corresponds to Heisenberg's uncertainty principle in which the product of time resolution and frequency resolution must result in at least the value 1. Therefore, a certain minimum bandwidth is necessary for a certain time resolution (here time of flight).

The step height $2m_\omega \cdot \Delta t_0$ mainly determines the maximum measurable distance according to the Nyquist theorem and $\Delta t_0 = T_{\mathrm{M}}/2n_{\mathrm{S}}$ as

$$r_{\max} = \frac{c}{2} \cdot \frac{\omega_{\mathrm{obj,max}} - \frac{2}{c}\omega_0 \dot{r}}{m_\omega} = \frac{c}{2} \cdot \frac{\frac{2\pi n_{\mathrm{S}}}{2T_{\mathrm{M}}} - \frac{2}{c}\omega_0 \dot{r}}{m_\omega} = \frac{\pi c}{4m_\omega \cdot \Delta t_0} - \frac{\omega_0 \dot{r}}{m_\omega}, \tag{41}$$

The number of steps n_{S} determines the ratio $r_{\max}/\Delta r$ between the maximum distance and the distance resolution. A Doppler shift leads, according to Eq. 33, to an elongation or shortening of the maximum measuring distance, in line with the second term of Eq. 41.

When applying Eqs. 37 and 38, care must be taken to ensure that the angular frequency ω_{obj} is signed. Without the use of an I/Q mixer, however, the sign of the frequency is not known, so that the mathematical sign must be determined via assumptions. In this case, positive distances may be assumed so that the object frequencies are positive with a positive gradient. This applies at least as long as $(m_\omega r + \omega_0 \dot{r}) > 0$. With a positive stair slope, it follows, therefore, that for objects below a

$$t_{\text{tc,min}} = (-r/\dot{r})_{\text{min}} = \frac{\omega_0}{m_\omega} \tag{42}$$

this condition is no longer met. t_{tc} stands for time to collision, the usual term for the quotient of distance and negative relative velocity. For an example with a $t_{\text{tc,min}} = 1$ s and 76.5 GHz carrier frequency, a slope of $m_\omega = 2\pi \cdot 76.5\,\text{GHz/s}$ is necessary corresponding to a frequency ramp of 76.5 MHz in 1 ms. Basically, this effect of a change of sign also occurs with a negative ramp slope with a corresponding “escape time,” there being no application in the area of driver assistance systems that has to identify objects “escaping” so quickly. Consequently, it is possible to operate a negative stair with a significantly lower ramp slope in terms of value.

$\Delta\omega_{\text{BA}}$ should be chosen as the last parameter. As a minimum requirement, a zero-value denominator must be avoided in Eqs. 37 and 38, i.e., $\Delta\omega_{\text{BA}} \neq m_\omega \Delta t_0$ must be chosen. It should further be noted that, according to Eq. 34, the phase difference clearly remains in the range of $0 \ldots 2\pi$ so that this range must at least be adequate for distances up to r_{max} at $(\dot{r} = 0)$ if the ambiguities are not to be resolved by other plausibility methods. This results in the condition for

$$|\Delta\omega_{\text{BA}}| \leq \frac{\pi c}{r_{\text{max}}} = 4 m_\omega \cdot \Delta t_0 \tag{43}$$

With some reserve margin for the change due to the Doppler effect $\omega_{\text{Doppler,max}} \cdot \Delta t_0$, the design which emerges is $|\Delta\omega_{BA}| < 10^6/\text{s}$, meaning a frequency jump of approximately 160 kHz, whereby in the case of a positive stair slope, a negative $\Delta\omega_{\text{BA}}$ leads to a higher slope difference than a positive $\Delta\omega_{\text{BA}}$. Distance and relative velocity being determined according to Fig. 12 as the intersection of two straight lines, orthogonality in respect to error robustness is optimal, i.e., the slope of one of the straight lines should be equal to the negative reciprocal value of the other slope, both variables being normalized to the resolution cell (Δr according to Eq. 40 and $\Delta\dot{r}$ according to Eq. 39). Using Eqs. 35 and 36, an optimum

$$\Delta\omega_{\text{BA,opt}} = -m_\omega \Delta t_0, \tag{44}$$

is then specified, i.e., the second stair is offset half a step downward (cf. Meinecke and Rohling 2000). Since, as described in Sect. 2.5.1, the sampling frequency

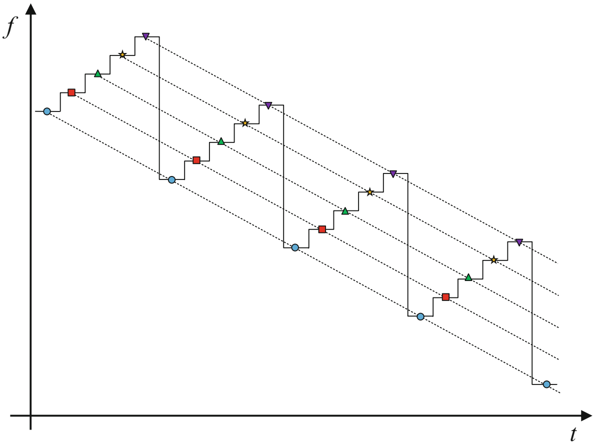


Fig. 13 Frequency over time curve of a combination of FSK and LFMCW/FSK, dash-lined: the measuring points combined into a dataset

necessary for the maximum Doppler frequency still permits sufficient time for intermediate measurements, it is possible to nest even further stairs. Thus the arrangement of Fig. 10 can be combined with a "macro stair," where, in line with the previous consideration, the offset (the small stair) is chosen contrary to the direction of the large stair (see Fig. 13). As a result, it is now significantly easier firstly to determine the phase difference $\Delta\varphi_{\mathrm{BA}}$ via four differences instead of one compared to the double-stair FMSK and secondly also to achieve a multitarget capability in the distance compared to the FSK because of the higher frequency deviation associated with the macro stair, so that the method also becomes suitable for stationary targets.

3.5.3 FMCW (Frequency-Modulated Continuous Wave)

Linear frequency-modulated continuous wave is a frequently used form of modulation. In this case, the instantaneous frequency is continuously changed in the form of a ramp:

$$\omega(t) = \omega_0 + m_\omega (t - t_0). \tag{45}$$

As a result, after mixing the receive and transmit signal, the following is obtained:

$$\overline{u}_{\mathrm{lt,r,i}}(t) = \frac{1}{2} A_{\mathrm{t}} A_{\mathrm{r}} \cos\left(\left(\frac{2\omega_0}{c}\dot{r} + \frac{2m_\omega}{c} r\right) t + \frac{2r}{c}\omega_0 + \left(\frac{2r}{c}\right)^2 m_\omega\right) \tag{46}$$

an expression similar to Eq. 32, where a constant phase displacement of $+(2r/c)^2 m_\omega$ is added due to the steadily rising transmit frequency but which is not otherwise significant. Although now the frequency is changed continuously compared to the FMSK stairs presented in the previous section, a signal sampled at discrete times

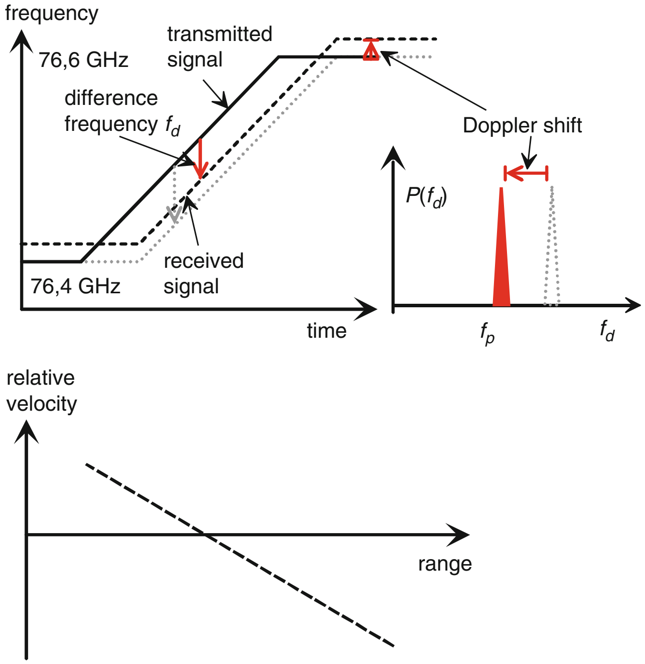


Fig. 14 FMCW with a positive ramp for an approaching object. *Top left*: transmitted and received signal; *top right*: spectral display of the difference frequency; *bottom*: distance and relative velocity values associated with a frequency

delivers the same difference frequency as in the stair form with the result that Eq. 33 remains valid and describes a linear combination of distance and relative velocity. The phase information, however, is not useful without comparing it to the phase of a different ramp.

As only the frequency information can be evaluated, the method can be illustrated clearly according to Fig. 14. With a positive ramp slope, the greater the frequency difference is, the longer the distance and the more the object is moving away. The ambiguity of the linear combination can be resolved if a further ramp with a different slope m_ω exists. With a negative ramp (see Fig. 15), the difference frequency is also greater the longer the distance. However, the difference does not increase as objects move away but rather as they approach. This is expressed in a linear combination which leads in a $\dot{r} \div r$ diagram to a negative slope. As shown in Fig. 15, the straight lines intersect at

$$r = \frac{c}{2} \cdot \frac{\omega_{\text{obj},1} - \omega_{\text{obj},2}}{m_{\omega,1} - m_{\omega,2}} \tag{47}$$

$$\dot{r} = \frac{c}{2\omega_0} \cdot \frac{m_{\omega,1}\omega_{\text{obj},2} - m_{\omega,2}\omega_{\text{obj},1}}{m_{\omega,1} - m_{\omega,2}} \tag{48}$$

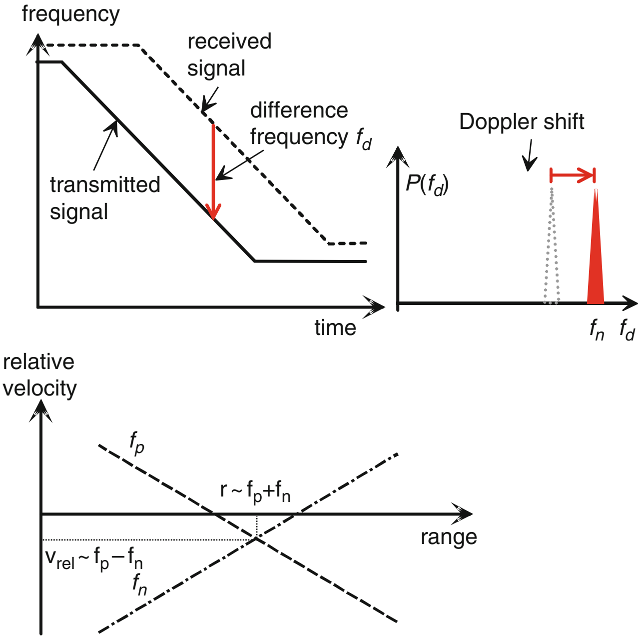
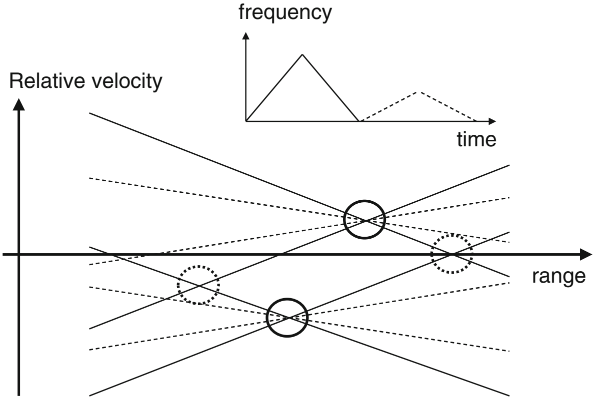


Fig. 15 FMCW with a negative ramp for an approaching object. *Top left*: transmitted and received signal; *top right*: spectral display of the difference frequency; *bottom*: distance and relative velocity values associated with the detected frequency for both ramps

Fig. 16 Ambiguity of the assignment with FMCW for two targets. *Crisscrossed solid line circles*: correct assignment; *dotted*: incorrect assignment and their resolution due to additional ramps; *dotted straight lines*: linear combination for the second double ramp



When applying these equations, care must be taken, as before, to ensure that the angular frequencies are signed. The constraint according to Eq. 42 applies here in an identical manner.

The multiramp FMCW method is not complicated as long as only one object is detected. In this case, the $\omega_{\mathrm{obj,i}}$ can be clearly assigned. This is no longer easily the case if several objects are detected. As illustrated in Fig. 16, misinterpretations are

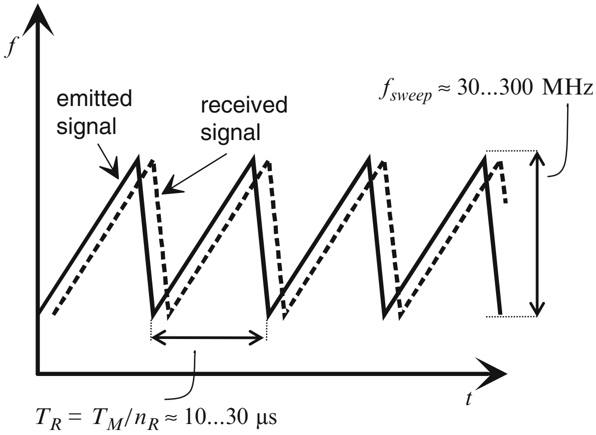


Fig. 17 Frequency time curve for chirp sequence modulation (pulse compression)

possible. The first pair of ramps (continuous lines) generates four intersection points from two objects, only two of which are correct. By means of one or several additional ramps with varying slopes, the ambiguity can be resolved, at least for a small number of objects, by only allowing those detections to apply that show an intersection of all ramps. In the example shown in Fig. 17 with two additional ramps of half the slope, there are four further straight lines in the $\dot{r} \div r$ diagram. However, all the straight lines of the four ramps only intersect for the correct objects. In scenes with a large number of desirable and undesirable targets, such as crash barrier posts, it may, nevertheless, happen that multiple intersection points are detected that do not correspond in reality. The equality of the amplitudes can be used as a further criterion for suppressing erroneous assignments, although it must be assumed here that the backscatter amplitude in the subsequent ramps is also really virtually identical. Although this assumption may not be correct in individual cases, the consequences are small as individual drop-outs are caught by the subsequent tracking. In spite of these measures, the assignment ambiguity remains the Achilles heel of this method.

The lack of coherence beyond the various ramps is a further weakness. The measuring duration T_R of the individual ramps rather than the overall measuring duration is relevant for the quality of the relative velocity. The smallest speed cell is determined according to Eq. 21 by way of the duration $T_{R,\max}$ of the longest ramp.

3.5.4 Chirp Sequence Modulation (Multichirp, Pulse Compression)

The modulation described below has several names. Here it is referred to as chirp sequence modulation because it consists of a sequence of identical linear frequency ramps (see Fig. 17). This method combines the advantages of all the methods described so far. At short intervals, n_R identical linear frequency ramps are repeated which, if they are frequency increasing (up-chirp), as illustrated in Fig. 17, would be heard in the acoustic range as chirps. The travel of the ramps is typically $f_{\text{chirp}} = 30\ldots300\ \text{MHz}$. The repeat rate depends on the Doppler frequency and should be approximately 80 kHz if ambiguities are to be prevented which, however, as mentioned several times previously, can also be eliminated by considering plausibility in the tracking, so that considerably lower repetition rates are possible. Although Eq. 33 applies for the individual ramps, there is nevertheless a clear assignment of the distance to the frequency cell

$$\omega_{\text{obj}} = \frac{2}{c} m_{\omega} r, \tag{49}$$

because the ramps are so short that a Doppler shift within the ramp duration becomes irrelevant and, therefore, a strong correspondence prevails between ω_{obj} and r. This relationship applies to all the following ramps as long as the target in the overall measuring time remains within the extension of the distance cells. This condition can definitely be violated at a high relative velocity and a long overall measuring duration T_{M}, if

$$|\dot{r}| > \frac{\Delta r}{T_{\text{M}}} = \frac{c}{2 T_{\text{M}} f_{\text{chirp}}} \tag{50}$$

With a high distance resolution of 1 m (corresponding to $f_{\text{chirp}} = 150\ \text{MHz}$) and a measuring duration of 20 ms, this occurs above $|\dot{r}| = 50\ \text{m/s}$. Despite such limits, it is possible to say that frequency cells correspond to distance cells which, in a similar manner to coherent pulse modulation (pulse Doppler), are understood as range gates. After the Fourier transform, as with the pulse Doppler evaluation in Sect. 2.4.3, a complex amplitude exists for each cell. In the same way as the pulse trains in the complex plane, in the following ramps, this amplitude describes a circle with the circular velocity ω_{Doppler} associated with the Doppler frequency. A Fourier transform of the complex amplitudes of the ramp sequence with the same distance cell therefore directly supplies the Doppler frequency, both for several targets in the same distance cell and different relative velocity and also with an algebraic sign, because now a complex dataset is transformed. The analogy to the pulse Doppler evaluation thus also leads to the term *pulse compression*, the whole energy of the ramp having now been concentrated on one range gate and, therefore, compared to a pulse duration that is approximately a thousand times smaller; a considerably better signal-to-noise ratio is achieved without increasing the peak output.

The approach described with two consecutive Fourier transforms is nothing more than a two-dimensional Fourier transform of the data field in which the measurement data of individual chirps form the gaps and the sequential chirps the lines. The result is present in a two-dimensional spectrum whose elementary cell is described by $\Delta r = c/2f_{\text{chirp}}$ and $\Delta \dot{r} = c/2f_0 T_{\text{M}}$. The extension of the field is determined by the sampling frequency f_S and the chirp sequence frequency $n_{\text{R}}/T_{\text{M}}$.

$$r_{\max} = \frac{\pi c}{4 m_{\omega}} f_S;\ |\dot{r}|_{\max} = \frac{n_{\text{R}} \Delta \dot{r}}{2} = \frac{n_{\text{R}}}{T_{\text{M}}} \cdot \frac{c}{4 f_0} \tag{51}$$

The chirp sequence modulation achieves the best possible utilization of the signal power, bandwidth, and measuring time. Along with the electrical noise of the receive path, the quality of the measurement is only defined by the quality of the frequency generation because nonlinearity, high phase noise, and inaccuracies in the ramp repetition (time and frequency errors) lead to "runout" of the detection peaks and diminish the detection capability, above all on the edge of the detection field, that is to say, with large distances and relative velocities.

The disadvantage of chirp sequence modulation is the high sampling rate $\left(>2\frac{r_{\max}}{\Delta \mathrm{r}}\cdot\frac{n_{\mathrm{R}}}{\mathrm{T_M}}\right)$ and the resulting large number of measured values $\geq 2\frac{r_{\max}}{\Delta \mathrm{r}}\cdot n_{\mathrm{R}}$ for the two-dimensional Fourier transform. In this way, almost empty data fields occur with more than 100,000 data points for the maximum 100 objects, and accordingly the desire here is to lower the data rate. This is possible at the cost of alias effects, for example, by reducing the chirp repetition frequency. The relative velocity ambiguity thus incurred can be eliminated for individual targets by comparison with differentiated distance. This measure fails, however, if targets having a velocity difference many times that of the velocity $v_{\mathrm{chirp}} = \frac{n_{\mathrm{R}} c}{2\mathrm{f_0 T_M}}$ corresponding to the chirp repetition frequency merge – such as a target that drives along a crash barrier at v_{chirp}. The consequences are incorrectly determined acceleration of the target object and also, following on from this, incorrect responses, e.g., of an ACC. Variable chirp repetition frequencies can provide a remedy in order to restore the uniqueness at least over several measuring cycles. In any case, it must be noted that with a reduced chirp repetition frequency, the frequency to distance assignment has to be corrected by a relative velocity share, similar to the FMCW method (cf. Fig. 14). In addition to reduction of the chirp repetition frequency, the data volume can also be reduced when sampling within a chirp with subsampling. Of course, corresponding alias side effects also occur in the process.

4 Angle Measurement

4.1 Preliminary Considerations Based on Antenna Theory

Before describing angle determination, we will first give introduction to the required principles regarding the beam shape of RADAR sensors. The beam characteristic of the electrical field intensity $E(\phi, \vartheta)$ in the far field, i.e., at distances that are much larger than the wave length, emerges (cf. Skolnik 2008) as the inverse Fourier transform of the antenna covering function $A(x, y)$, with the azimuth angle ϕ corresponding to the configuration in the x direction and the elevation angle ϑ corresponding to the y direction. The azimuth angle ϕ positive towards the left lies in the sensor horizontal plane of the sensor oriented in the Z_{S} direction and the elevation angle ϑ describes the angle to the $Z_{\mathrm{S}} - X_{\mathrm{S}}$ plane (positive upward). For a flat antenna parallel to the $X_{\mathrm{S}} - Y_{\mathrm{S}}$ plane, the result is as follows according to Skolnik (2008):

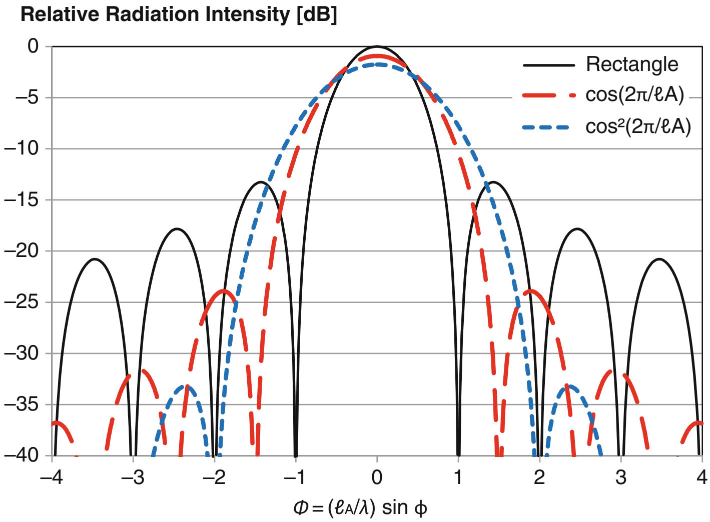


Fig. 18 Calculated one-dimensional antenna characteristic for a square covering function and a simple as well as a squared split cosine bell, normalized to the overall output. The abscissa variable $\Phi = (l_A/\lambda)\sin\phi$ is the sine of the beam angle normalized to the ratio l_A/λ of the aperture width to wave length

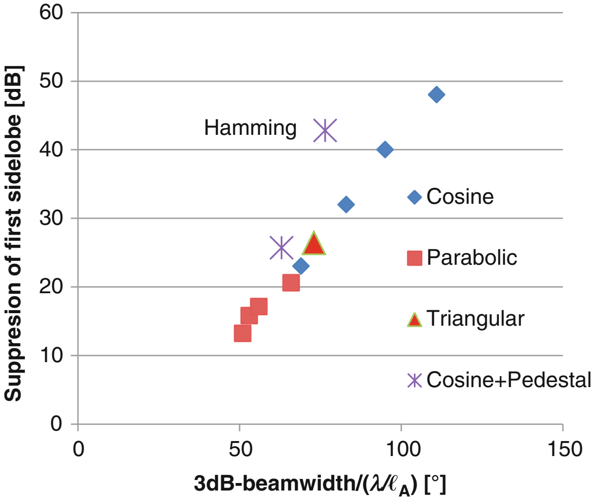
$$E(\phi,\vartheta) = \iint A(x,y)e^{j\frac{2\pi}{\lambda}(\sin\Theta\cdot(x\cdot\phi+y\cdot\vartheta))}dxdy;\ \text{with}\ \Theta^2 = \phi^2+\vartheta^2 \qquad (52)$$

This equation initially describes the field intensity distribution in the far field for a wave radiated with the covering function $A(x, y)$ but applies similarly for receiving. Thus, multiplication of the transmit characteristic with the receive characteristic applies for the angular dependency of a sensor. As long as the transmit antenna is not far away from the receive antenna, the two-way characteristic can be described as the (generally complex) product of the one-way characteristic. When using a monostatic radiation concept, that is to say, when the transmit beam runs through the same antenna unit as the received signal, the result is the square $E^2(\phi, \vartheta)$ (which is complex when the covering function is not mirror symmetrical about the antenna center).

The antenna characteristic resulting from this is illustrated in Fig. 18 for three simple, symmetrically one-dimensional cases of covering functions. The abscissa uses the normalized variable $\Phi = (l_A/\lambda)\sin\phi$ and is thus scaled by the ratio of aperture width l_A (antenna opening width) and wave length.

Based on these examples, it is already possible to anticipate the conflict between the strongest possible concentration of the main lobe and the lowest possible height of the side lobes. As specified in a table in Skolnik (2008), depending on the covering function, it is possible to choose a compromise appropriate to the angle evaluation concept (cf. Fig. 19). A characteristic optimal for suppression of the first

Fig. 19 Side lobe suppression vs. width of main lobe (at −3 dB, one way) according to Skolnik (2008)



side lobe is displayed by the Hamming window, in which 8 % of the amplitude prevailing in the center still remains at the margin. In spite of such an optimization strategy, the antennas must be approximately 80 times larger than the wave length multiplied by the reciprocal value of the main lobe width per degree; one degree of main lobe width requires an aperture width that is around $l_A = 80\,\lambda$ large, corresponding to 32 cm for 1° and 77 GHz.

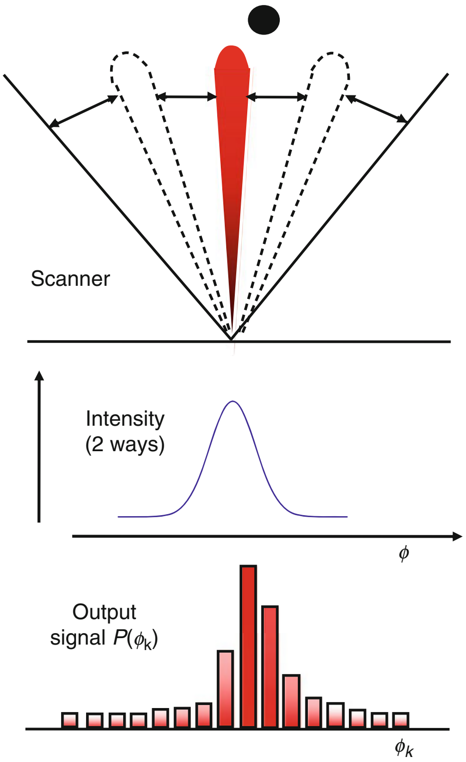
A further undesirable side effect of high side lobe suppression is the reduction of the antenna gain (see also Fig. 18) because the suppression is always brought about by a covering function which falls off towards the edge of the antenna. Accordingly, the effective antenna area decreases, and the main lobe becomes wider and thus distributes the output over a wider area which in turn leads to a decrease in intensity in the center of the beam.

For long-range RADAR applications such as ACC (see ▶ Chap. 45, "Adaptive Cruise Control"), an angle range $\Delta\phi_{\max}$ of approximately 10°. . . 20° azimuth and 3° elevation is required for total coverage. A separation capability in respect to the elevation would be desirable for differentiating a bridge from a stationary vehicle – height difference approximately 2 m. However, for this a separation efficiency in the long range of 1° (=2 m/116 m) and consequently an antenna of at least 30 cm would be necessary which is out of the question given the available installation space. Thus, the angle evaluation in the long range is limited to the azimuth. For the use of RADAR at close range, which has increased significantly in recent years, particularly for full-speed range ACC (see ▶ Chap. 45, "Adaptive Cruise Control") or for collision protection systems, stationary obstacles also have to be classified. Therefore, for these functions, not only is a considerably enlarged azimuthal range (30°. . . 60°) required but a resolution in elevation is also desirable. In this case, particularly with a planar antenna (see Sect. 3.6) with justifiable dimensions, it is possible to achieve measurement of the elevation and therefore to differentiate the much-quoted Coca-Cola can from taller objects.

4.2 Scanning

From the comprehension point of view, mechanical scanning is the easiest method of angle determination. To do this, a beam deflection unit or a planar antenna is mechanically pivoted so quickly that the entire azimuthal detection range is scanned within one measuring and evaluation cycle (50 … 200 ms). Figure 20 illustrates the principle. Due to the dependence on the aperture width described above, the RADAR lobe has at least 2° main lobe width if the aperture width is not supposed to be greater than 15 cm. The lobe is "pushed" over the measurement range in approximately 1° steps. Instead of really discrete step control, a continuous scanning movement takes place to prevent noise-generating accelerations and to manage with smaller load torques. The measured values are then nevertheless assigned to a discrete angular position, that is to say, the center of the scan positions within a measurement window, which is assigned to this angular segment. Although

Fig. 20 Scanner principle for angle determination. *Top*: closely concentrated beam scans the whole registration area and detects the point target; *center*: azimuthal angle characteristic of the concentrated beam; *bottom*: result for a point target



the uncertainty arising due to the lobe width increases by a “motion blur,” as the measured data are windowed to prevent leakage effects, i.e., are greatly reduced at the beginning and end of the measuring interval, the effective motion blur is reduced to approximately 30 %. A further benefit is that the blurs approximately (or exactly with a Gaussian characteristic of antenna and window function) add up geometrically so that the loss of definition is reduced to only approximately 10 %. Naturally, an even smaller step width may be chosen and thus the motion blur can be minimized. The argument against this, however, is that dividing the measuring times into many intervals assigned to the angular segments worsens the selectivity for the Doppler evaluation. It becomes clear, therefore, that owing to the principles involved, a mechanical scanner will be poorer in respect to relative velocity measurement than a multibeam arrangement measuring the same measuring time.

It should additionally be noted that the azimuthal evaluation area is smaller than the scan area, because on the edge at least it must be possible to identify a decrease when determining the focal point. Therefore, the actual angular area towards both edges is smaller than the scan area by approximately half a beam width. The huge advantage of the scanning method, in addition to a high level of accuracy because of the narrower beam by comparison with the other concepts, is also the ability to separate objects with regard to the angle. Determination of the lateral object extension is only practically possible at smaller distances as even a narrow beam of 2° width is expanded around 1.8 m at 50 m and is thus already as wide as a car.

However, it is still possible to achieve something if the antenna characteristic is known, e.g., by measuring at the end of manufacture. With the help of deconvolution algorithms, both the values for resolution and also the separation efficiency can, in the favorable case, be improved by a factor of approximately ½ (cf. Diewald 2013).

4.3 Monopulse

The monopulse method is based on a dual-antenna configuration (see Fig. 21), although this is mainly only used for receiving while the transmit beam is emitted by means of a single separate antenna.

The (receiving) antennas may differ due to the beam characteristics or simply because of the position which is displaced horizontally by $\Gamma \cdot \lambda$ for azimuthal angle measurement. For two adjacent, otherwise identical antenna fields, there is a phase difference of

$$\Delta\varphi = 2\pi\Gamma \sin\phi \tag{53}$$

depending on the azimuth angle. For the amplitudes of the difference signal, this means that instead of the original amplitudes A_1 and A_2 with $|A_1| = |A_2| = |A|$ for difference and sum signal, an amount weighted with the sine or the cosine of the phase difference is measured:

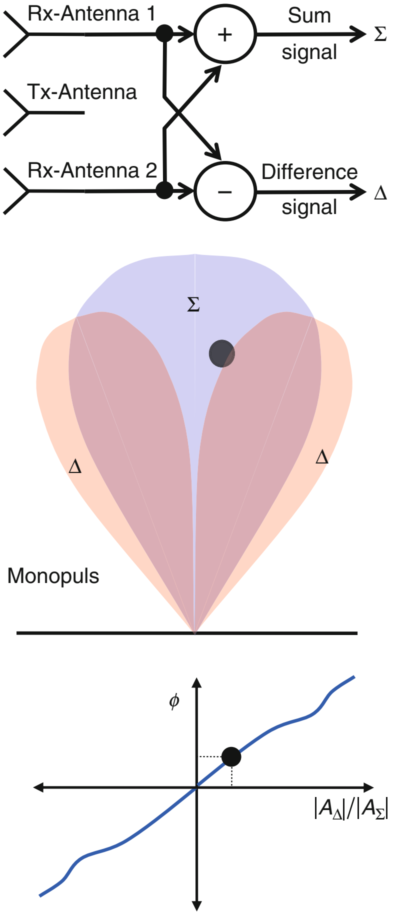


Fig. 21 Monopulse principle for angle determination. *Top*: formation of the sum and difference signals; *center*: the azimuthal angle characteristic of the beams thus formed; *bottom*: typical characteristic of the azimuth angle vs. quotient of the amplitude values of difference and sum signal at smaller angles

$$|A_\Delta| = 2|A| \sin \frac{\Delta\varphi}{2}; \; |A_\Sigma| = 2|A| \cos \frac{\Delta\varphi}{2} \tag{54}$$

Thus it is possible to determine the azimuth angle from the ratio of the difference to the sum signal without a phase-sensitive measurement being necessary for this:

$$\phi = \arcsin\left(\frac{\arctan \frac{|A_\Delta|}{|A_\Sigma|}}{\pi\Gamma}\right) \tag{55}$$

However, restriction to angles $\Delta\varphi = 2\pi\Gamma \sin\phi < \pi/2$ is necessary because of the uniqueness. From this follows the dimensioning specification of $4\Gamma \sin\phi_{\max} < 1$.

At a maximum azimuth of $\phi_{\max} = 30°$, the antennas would be exactly 0.5 λ away, at 6° approximately 2.5 λ.

A further possibility of the monopulse method consists of comparing the amplitudes with different beam characteristics. In the usual configuration that is symmetrical to the center, the beams outside the zero angle possess the maxima but have an identical amplitude at the zero angle because of the symmetry. The quotient of the amplitude quantities

$$\frac{|A_1| - |A_2|}{|A_1| + |A_2|}$$

can again be referred to initially as an approximately linear measure of the azimuth angle. When it is possible to assume a constant backscatter between two consecutive measurements, alternating sequential evaluation is sufficient. This method is, therefore, also called *sequential lobing*.

If, as previously illustrated in Fig. 21, the difference and the sum signal are generated directly, then the phase difference and amplitude difference overlap each other, resulting in an even steeper characteristic between azimuth angle and quotient $|A_\Delta|/|A_\Sigma|$.

The measuring method described is accurate for individual point targets. However, two targets can already generate unreasonable values in a way that is not identifiable in the same measuring cycle. Therefore, care must be taken when using this method that, due to good distance and/or relative velocity separation, the probability of the azimuth originating from two or more targets will be very low.

If the difference and the sum signal are measured simultaneously and if complex amplitude determination is possible, it is basically possible to verify the plausibility of signals via the difference phase between A_Δ and A_Σ. Separation of the influences (amplitude characteristic and phase difference) is used for this. As different amplitude characteristics are also mostly linked to phase differences, it is advisable to store the overall characteristic (amplitude ratio, phase difference) as a function of the azimuth angle. A further advantage is doubling of the uniqueness range of the phase evaluation to $\pm\pi$, as the algebraic signs of the complex amplitudes can also be used in arctan calculation.

4.4 Multibeam Antenna

Using multibeam antennas can improve the monopulse method. On the one hand, the measuring range is extended for a given individual beam width. On the other hand, in most cases, it is possible to identify multitarget distortion as described above. The basic principle is illustrated in Fig. 22. Angle evaluation is carried out by comparing with the sensor-specific standardized antenna characteristic which is stored in a nonvolatile memory. Figures 23 and 24 show examples of real angle characteristics.

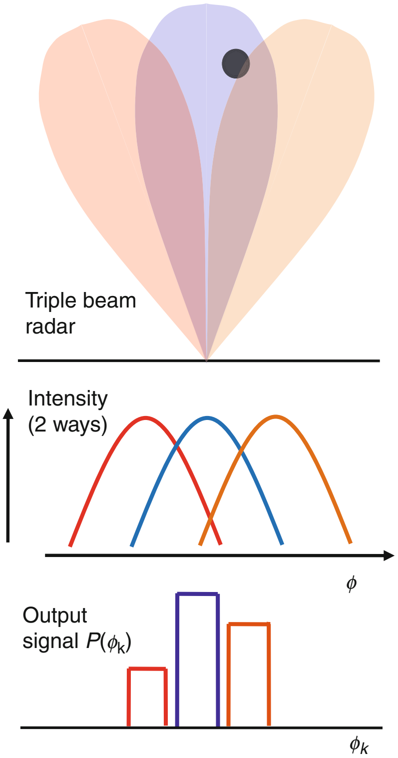


Fig. 22 Multibeam principle for angle determination. *Top*: overlapping lobes; *center*: the azimuthal angle characteristic of the individual beams; *bottom*: output in the individual beams resulting from a point reflector

Only the central beam of Fig. 23 shows strong side lobe suppression. The neighboring lobes each exhibit significantly raised side lobes towards the opposing side of their main orientation, indicating an asymmetrical covering function which originates from the off-center radiation (cf. also Sect. 8.3). With the quad-beam in Fig. 24, all the beams are asymmetrical, this is particularly true for the outer ones.

At the end of sensor production, target simulators are used to automatically build look-up tables to determine the angle characteristic. The signal outputs $|A_\mathrm{i}|^2$ of the ith beam measured $(i = 1 \ldots n)$ are standardized to the sum of the outputs of all the beams

$$a_i = \frac{|A_i|^2}{\sum_{j=1}^{n} |A_j|^2}$$

so that in the case of a point target which is located in the azimuth angle ϕ_0, for the cross-correlation

$$K(\phi_\tau) = \sum_{i=1}^{n} a_i \cdot a_{\mathrm{norm},\,i}(\phi_\tau) \tag{56}$$

with the correspondingly standardized angle diagram $a_{\mathrm{norm},i}$, a maximum at $\phi_\tau = \phi_0$ is reached with a value of close to 1. If the maximum value is significantly

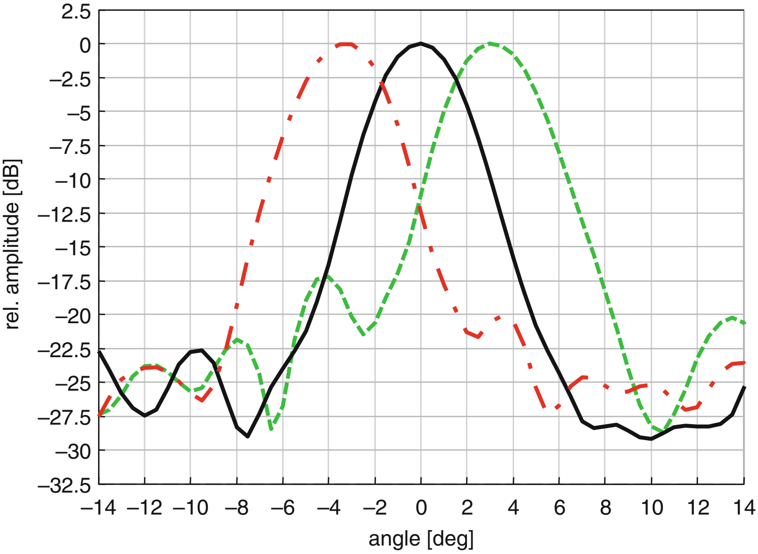


Fig. 23 Two-way antenna diagram of a triple-beam pulse Doppler RADAR (e.g., Continental ARS200) (Kühnke 2003)

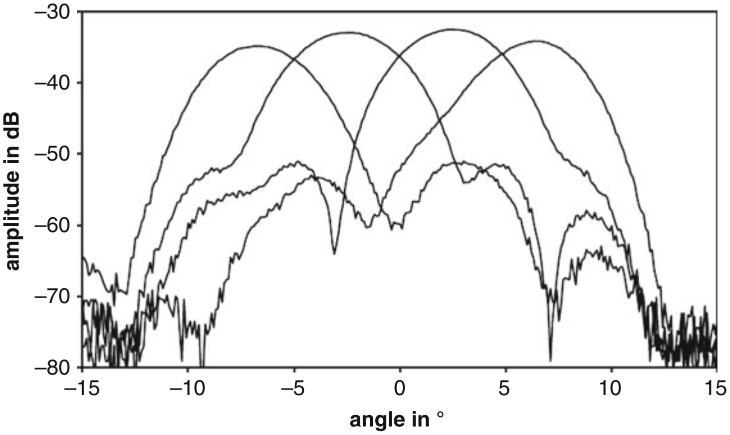


Fig. 24 Two-way antenna diagram of a quad-beam FMCW RADAR (e.g., Bosch LRR2) (Kühnle et al. 2002)

smaller than 1, it may be assumed that the assumption of a single reflector is not given and thus the determined angle must not be trusted. However, an evaluation according to Eq. 56 can also lead to distortions where $K(\phi_0) \approx 1$ if practically only one lobe has a high relative receiving power. Therefore, the so-called antenna matching in relation to dB can alternatively be carried out logarithmically, and then the correlation coefficient from this can be assessed. However, this requires an adequate signal-to-noise ratio of all values.

Even with multibeam concepts, it is possible to benefit from the phase differences according to Eq. 53 if the reflected signal is received simultaneously on several channels. In this way, additional information is received in the case of an n-beam antenna $n - 1$. One way of evaluating this information is to assign the reference phase to a beam k, e.g., the center beam or one of the two center beams in the case of an even number of beams. Then, from the differential phase $\Delta\varphi$, it is possible to determine real and imaginary parts $A_{\mathrm{Q},i} = |A_i| \cos \Delta\varphi/2$ and $A_{\mathrm{I},i} = |A_i| \sin \Delta\varphi/2$ for each of the other beams. Thus, during simultaneous receiving by an n-beam antenna, a total of $2n - 1$ information is available for the angle determination which can be evaluated in the manner described previously (Eq. 56).

Simultaneous receiving in multibeam antennas means that initially only the receiver side is multibeam while the transmit beam comes either from a separate transmit path or, as in the example from Sect. 8.1, from the superposition of several transmit paths. Basically the transmit paths can also be modified, e.g., by means of switches; however the mixers of the corresponding paths, which are needed for processing of the received signals, are also mostly tied up. Moreover, for such a modification on the transmit side, measuring time must be provided in a similar manner to the scanning method, meaning that measuring takes longer or the measuring time is divided between various beam configurations which leads to a deterioration in the relative velocity measurement.

The simultaneous operation of multibeam antennas with phase evaluation can also be described as a (simple) form of digital beam forming, because the sequential search for the highest correlation proceeds as if the antenna with its phase and amplitude identification is successively steered virtually in the search direction. The transmit characteristic, however, remains unchanged unless the transmit paths are also unmodified. This can also happen, in addition to switching of the transmit paths, as a result of targeted phase shift between the individual antennas of the transmit paths. Such antennas which are mostly designed as planar phased arrays enable a large number of evaluation methods which will be dealt with in greater detail in Sect. 3.6.

4.5 Dual-Sensor Concept

The concept presented in publication (Lucas et al. 2008) combines two RADAR sensors in one integral dual-sensor concept. In this case, two almost mirror-image asymmetrical antenna characteristics are used in which the side lobes responsible for a wide close-range illumination are directed towards the outside of the vehicle, while the more powerful central lobes are largely directed parallel and forward (cf. Fig. 25). Three advantages mainly arise: broad coverage from the outset (i.e., after the first distance cell), approximately ±20° vision at close range, and an overlap in the main range (cf. Fig. 26). The overlap can be used both for fault

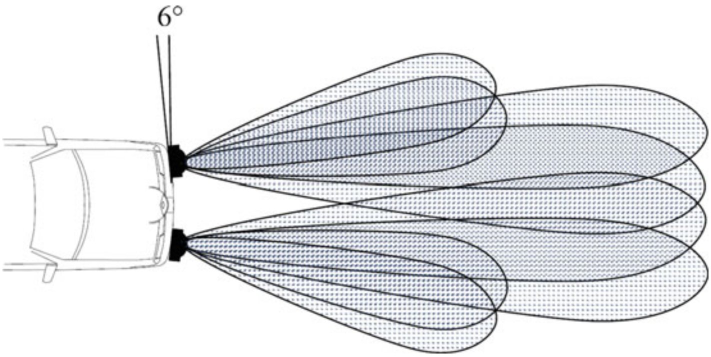


Fig. 25 Dual RADAR configuration with asymmetrical quad-beam RADAR sensors (Lucas et al. 2008)

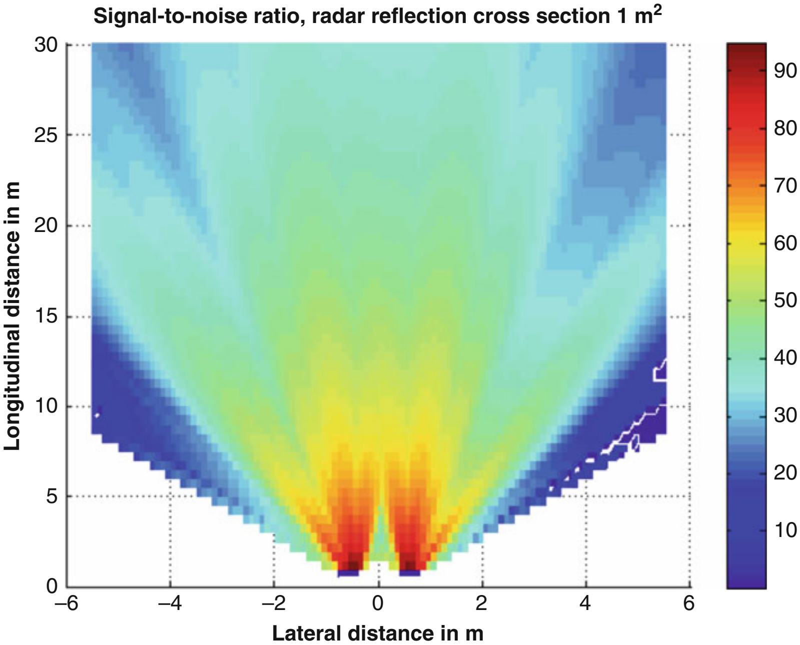


Fig. 26 Detection coverage of the dual RADAR configuration with asymmetrical quad-beam RADAR sensors (Lucas et al. 2008)

detection and also for improving the signal processing, primarily for the determination of the azimuth angle. The fact that two installation spaces will have to be found for such a configuration can be deemed both negative and also positive – positive in particular if the intention is to achieve "RADAR eye symmetry" with visible installation. The disadvantage is double the costs compared to the single sensor although dispensing with additional close-range sensors can improve the bottom line.

4.6 Planar Antenna arrays

Planar antennas have two positive properties relevant for practical use:

- The installed depth of the sensors is considerably reduced. It is no longer essentially dominated by the antenna but rather by the depth required for the other electronic and mechanical components (such as connectors) such that installed depths of 15. . .30 mm are obtained.
- Arrays may be created with which transmit and/or receive characteristics of the antenna can be controlled.

The most frequent configuration is the single transmit antenna surface which consists of mutually fed "patches" adapted to the wave length dimension, and several (≥ 4) receive antenna surfaces (also consisting of a large number of "patches"). Due to the repeated surface shape of the receive antennas, the receive characteristic for targets in the far field is identical. As with the monopulse principle (cf. Sect. 3.3), in the case of a single target (in the same distance and relative velocity cell), a phase of $\varphi = i2\pi\Gamma$ occurs between the individual receive antennas (index $i \in \mathbb{Z}$, offset $\Gamma\lambda$), plus a random phase offset which is the same in every case and can, therefore, be ignored. If another target is added, the complex amplitudes generated by the two targets interfere with each other so that the amplitudes of the individual array elements are also no longer equal in magnitude. It corresponds to a linear superposition of a complex signal that is "scanned" discretely by the individual antennas with the increment length $\Gamma\lambda$, with $\sin \phi$ as corresponding variable in the Fourier space.

On adapting the phase difference by means of variable phase shifters, antenna elements can be switched together to make electronically controlled antennas (similar to phased array) with high directivity. If these phase shifters are continuously activated, the result is electronic scanning. Controlled phase displacement is dispensed with in digital beam forming, and the data flow to the individual antenna elements is stored parallel or sequentially and only evaluated in respect to the phase difference in the digital postprocessing.

The boundaries can be deduced from the number and the distance of the individual antennas: the increment is used to specify the maximum unambiguous range ($\sin \phi$ between $\pm 1/2\Gamma$) and the overall width in relation to λ ($n\Gamma$, number n of the receive array elements) to specify the angle cell $\Delta \sin \phi = 1/n\Gamma$. A complex Fourier transform via the receive amplitudes which are present in relation to the individual antenna elements for the same distance and speed cell therefore delivers an angle range (strictly speaking, for this the arcsin function must still be applied to it). For an antenna with 1° angle cell, an overall width of 23 cm would be obtained even at $\lambda = 4$ mm (corresponds to 77 GHz). These dimensions considerably exceed the tolerance installation dimension. If two transmit antennas are each placed on the right- and the left-hand side at half the distance to the receive antennas, then the number of receive antennas required can be halved – thus instead of 1 Tx + 8 Rx now 2 Tx + 4 Rx. Naturally, this requires a multiplex system during the demodulation since either only one Tx antenna may be active in each case or they are put into an alternating sum and difference mode. In the latter case, a Tx antenna is sometimes activated with the transmit signal that is in phase with the other Tx antenna and sometimes with one that is inverted. If a bistatic element is placed on the edge of the Rx array, then one Tx antenna on the opposite side is sufficient (at the same distance as between the individual Rx antennas), as shown in the example of application in Sect. 8.2.

The width of the clear field of vision can be enlarged by means of a high number of antenna elements although, in addition to the costs for each new signal channel, space problems for the receiving surface also arise which, however, can only be partially compensated by "interlocking" or slanted fields. In practice, the number of arrays remains limited to values between four and eight. If the antenna elements are also offset in the vertical direction, it is also possible to measure the angle in elevation. Similar to the Nonius principle, with the help of an additional small offset for a subset of the Rx antennas, the uniqueness range can be extended allowing both a long-range measurement to be carried out with high directivity in the narrow angle range and also a short-range measurement in a wide angle range. Many of the approaches referred to above are described in further detail in Wintermantel (2010).

Every "spatial frequency" of the Fourier spectrum corresponds to a virtual antenna such that eight virtual "antenna lobes" can be formed from eight discrete antenna elements according to the discrete Fourier transform. However, it should not be forgotten that a discrete Fourier transform presupposes certain assumptions: in addition to the sampling theorem (uniqueness range), this also includes the assumption of the periodic continuation of the signals that are being transformed. This assumption is certainly not met here. Therefore, blurring occurs in the spectrum due to the signal running out (leakage) which is expressed as only gently sloping side lobes. The usual remedy, windowing with windows decreasing at the margin like the van Hann window, only leads to a lowering of the (effective) amplification of the outer antenna elements, whereby the effective width and the resolution decrease. In addition to the methods described below for the separation of several objects, it is possible to derive more easily interpretable angle interpretations from the original information (complex amplitudes for each antenna element) if the trick of "zero padding" is used in which the identical or an integer number multiple of zero elements is added. This acts like a spectral interpolation such that with the same number of zeros, twice as many virtual lobes are available although the actual resolution has not changed (every second lobe corresponds exactly to the original lobe without zero padding).

All multiple antenna configurations have a high level of sensitivity if the signal path is impaired by systematic or random errors. The systematic errors consist essentially of differences between the channels in amplification (absolute value) and phase as well as couplings of the antenna elements. These can be compensated by calibration at the end of the production line by means of laboratory equipment or by autocalibration in the field via the feedback of statistical variables (e.g., as described in Heidenreich (2012) or Massen and Möller (2012). The influence of electrical noise can only be reduced by using assumptions about the target behavior, whether about the maximum number of targets that have to be considered in the distance and velocity cell or about their temporal constancy so that the angle can be estimated over several measurements.

Due to the low basic resolution (large angle cell), the occurrence of a second target is often already a problem for determining the angular position. Parametric evaluation processes which specify a best estimation based on the hypothesis of a specific number of targets offer ways out here. Multiple signal classification (MUSIC) and estimation of signal parameters via rotational invariance techniques (ESPRIT) should be mentioned as familiar processes which do, however, need several datasets for calculation for this but then may also fail to reach the resolution limit referred to above. Nonlinear least square (NLS) methods can determine angles of the objects very efficiently with only one dataset as long as the targets do not lie within the resolution limit. These and other methods are described and compared in Stoica and Moses (2005). In Koelen (2012), a multiple target identification (MUTI) method is presented which first determines whether several targets are present in the elementary cell (cf. Sect. 4.4) with the same distance and relative velocity. In most cases, this will not be the case, so that for this large number, a determination method that is not computationally intensive can be used for the angular position, e.g., according to the straightforward phase monopulse principle. Only if the one-target condition is not met are more computationally intensive methods used, such as the NLS method already mentioned.

For transmit-side arrays, the phase centers between the antenna elements must be defined and controlled; phase networks can do this, depending on the feed-in points. The Butler matrix is one such network that makes it possible, in a quadratic arrangement with in most cases 2^n inputs and exactly as many outputs, to generate a defined phase difference between the adjacent antenna elements (outputs) if power is applied to just one of the inputs. As a result, by switching the transmit path over to one of the inputs, the transmit beam can be realigned like a scanner with discrete angular steps.

5 Main Parameters of Performance

Even though the most important variables of performance emerge from an understanding of the functions, particularly modulation and angle evaluation, they are summarized here in a brief overview.

5.1 Distance

The performance of the distance measurement is mainly specified by the frequency bandwidth f_{Bw} of modulation (cf., e.g., Eqs. 18 and 40) and determines the size of the distance cell

$$\Delta r \geq \frac{c}{2f_{\mathrm{Bw}}} \tag{57}$$

and therefore the separation capability. The measuring limit for the maximum distance is determined essentially by the sampling rate (cf. Eq. 51) for RADAR with frequency modulation, while it is specified in pulse Doppler RADAR by the length of the sampled received signal.

The maximum distance in relation to a standard target also depends, in addition to the modulation parameters, on the transmit power, the antenna quality (amplification at 0°), and the signal-to-noise distance of the receiving electronics (cf. Eq. 4, Sect. 1). Note that in practice, the reflectivity of the objects fluctuates by several orders of magnitude and, in addition, multipath interferences make this limit appear anything but sharp.

The minimum distance can only be smaller than the separation capability interval if multitarget capability at a distance is dispensed with. “Turnover effects,” as described in Sects. 2.5.2 (Eq. 42) and 2.5.3, may lead to an enlargement of the minimum distance that is dependent on the relative velocity. In pulse RADAR systems using the same antenna paths for transmitting and receiving, it is only possible to measure after the transmit pulse has decayed which results in an range corresponding approximately to the pulse length in which the distance cannot be correctly determined. At above ca. 25 % of the pulse length, however, it is possible to detect that there is a target.

5.2 Relative Velocity

For the cell size $\Delta \dot{r}$ and therefore for the separation capability and also for the accuracy of the relative velocity, the uninterrupted measuring time T_{M} is decisive (cf., e.g., Eqs. 21 and 39). Sampling of the Doppler effect is significant for the maximum and minimum relative velocity. However, an ambiguity due to too low sampling frequency can definitely be compensated if an assignment to the ambiguity areas via distance differentiation is successful.

5.3 Azimuth Angle

No simple relationship is specified for the performance of determining the azimuth angle. An azimuthally narrow beam that electronically or mechanically scans an azimuthal sector that is as wide as possible would be ideal. With monopulse and multibeam concepts, wide illumination is only possible by means of individual beams that are also wide. The overall measuring range is used here as a quality characteristic

$$\Delta\phi_{\max} = \phi_{\max} - \phi_{\min} \tag{58}$$

and the azimuth cell size relevant for the separation capability

$$\Delta\phi_{\min} = \frac{\Delta\phi_{\max}}{N_{\text{azimut}} - 1} \tag{59}$$

defined via the number of independent pieces of information N_{azimuth}. For a scanner, $\Delta\phi_{\min}$ results from the beam width of the individual beam, for a sequential n-beam concept $\Delta\phi_{\min} = \Delta\phi_{\max}/(n-1)$ and for a simultaneous concept with phase evaluation $\Delta\phi_{\min} = \Delta\phi_{\max}/(2n-2)$. For sequential lobing and monopulse, $\Delta\phi_{\min} = \Delta\phi_{\max}$, as there is no multitarget information present unless both signals are measured simultaneously for monopulse and a separation of phase difference and amplitude difference is used (then $N_{\text{azimuth}} = 3$).

5.4 Performance and Multitarget Capability

A RADAR for use as a surroundings sensor in automobiles must have multitarget capability. For this, suitable separation capability is necessary in at least one of the dimensions distance, relative velocity, and azimuth angle. Depending on the concept, the separation capability is sometimes prioritized for distance and sometimes for relative velocity. In the figurative sense, a smallest possible “cell volume” is aimed at obtaining a multitarget capability that is high in practice, the cell volume being the product of the cell sizes of the three dimensions even if they have different units. Established conversion and, therefore, weighting factors needed for a consideration of the volume are not known and probably not always appropriate. This applies above all to cases lying far apart if, for example, a sensor that has uniformly small cell sizes is to be compared to a sensor that only resolves one dimension but can do this very accurately. The following section specifies guide values for the required cell size of a long-range RADAR which result in an adequate multitarget capability on their own. In this case, we assume the longitudinal extension of a small passenger car that is standing at a distance of 100 m from the sensor. Moreover, for a separation, it is assumed that a distance of three cells is required. In theory, the distance of two cells would also be sufficient, but the windowing and the blurring of the beam mean that this is not possible:

$$\Delta r \approx 1.5\ \text{m}, \Delta\dot{r} \approx 0.1\ \text{m/s}, \Delta\phi \approx 0.7^\circ. \tag{60}$$

It can be seen here that multitarget capability based only on angle is not possible with installation-compatible antennas (aperture width would have to be >45 cm). Separation efficiency based on the distance alone reaches limits if several objects are at virtually the same distance; separation efficiency according to relative velocity fails with stationary objects. Therefore the aim is for separation according to distance and relative velocity. Figure 27 illustrates in diagrammatic form the solid rectangular area $\{r_{\min} \dots r_{\max}, \dot{r}_{\min} \dots \dot{r}_{\max}, \phi_{\min} \dots \phi_{\max}\}$ which is obtained from the minimum and maximum values and which consists of the single cell volumes $\{\Delta r, \Delta\dot{r}, \Delta\phi\}$. The qualitative statement that can be derived from this is that performance is greater the larger the area volume and the smaller the cell volume.

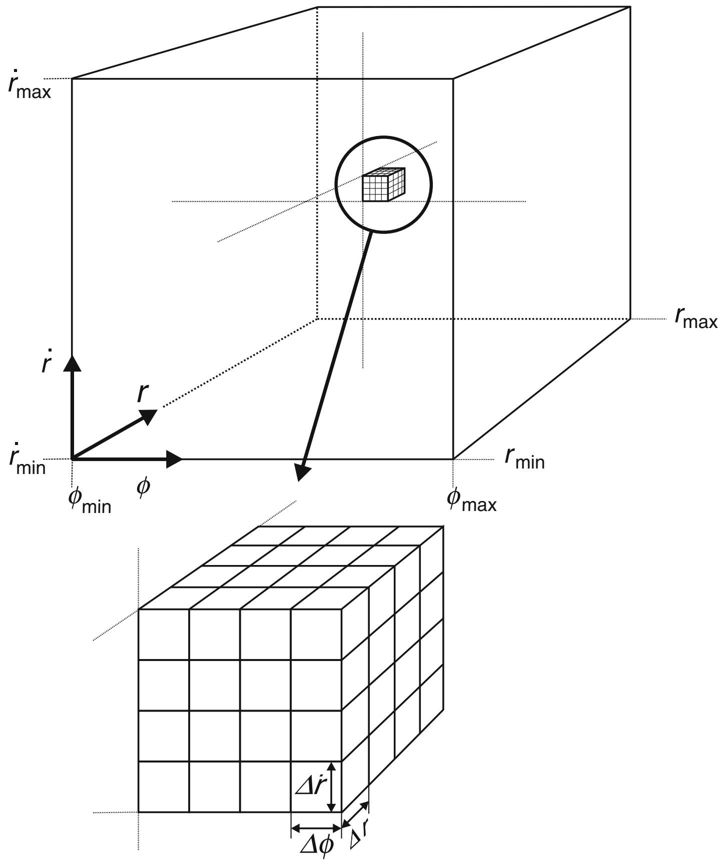


Fig. 27 Visualization of the separation capability as cell volume in the dimensions distance, relative velocity, and azimuth angle

However, it must be noted that other reasons for deterioration exist in addition to the principle-related limits. Frequency generation and modulation in particular may contribute to deterioration. Both a nonconstant amplitude over the uninterrupted measuring time and also phase noise or nonlinearities lead to widening of the frequency peaks and reduce the separation efficiency.

Alongside separation capability, resolution also plays a big role in quality. For this, adjacent cells are used for a peak focal point determination by means of which resolutions of approximately 1/10 of the cell width can be obtained. On the other hand, however, this only applies for point targets. Real targets by comparison cause variances that lie significantly above this value. Alternating reflective focal points lead both longitudinally and also laterally to jumps of several meters, but even in the relative velocity, variances occur if relative movements are detected, such as moving parts or transport goods with a relative degree of freedom, e.g., cars on car trailers. These variances can be so severe that an object is detected in several cells.

It is then necessary, mainly by means of heuristic approaches, to cluster these cells back into a common object.

In addition to the efficiency of detecting objects, robustness in relation to artifacts plays an important role. Thus neither so-called ghost targets – that is, object identifications without an existing object – nor distorted values are desirable. Nonlinearities in the signal chain can lead to representatives of the first group, unresolved ambiguities, and interferences to the second group. Another expected ability is the suppression of targets that can be driven over or under, such as manhole covers or bridges, at least in the area where the activation of an emergency stop in response to a stationary object is no longer excluded (cf. ▶ Chap. 46, "Fundamentals of Collision Protection Systems"). Unfortunately, the robustness referred to cannot be predicted by specifying design parameters, because the countermeasures are buried "deep" in the evaluation algorithms.

5.5 24 GHz Versus 77 GHz

The frequency band from 24.0 to 24.25 GHz also allows RADAR use in road traffic in addition to the 76–77 GHz band. The advantages are lower loss cable routing and more reasonably priced components even if the gap will decrease with the increasing use of SiGe components at 77 GHz. The increase in the relative velocity cell may be considered a disadvantage because the Doppler frequency scales proportionally to the carrier frequency. The biggest difference arising from the lower frequency results from the higher wave length ($\lambda \approx 12\,\text{mm}$) which, in turn, leads to widening of the beam characteristic if the size of the antenna is to be retained. The antenna gain is smaller and the angle resolution deteriorates. Therefore, the use of 24 GHz is ideal for the midrange up to 100 m. It is also well suited for the close range if a wide beam characteristic is desirable. However, detection is hardly possible below 0.5 m due to band limitation with the result that a 24 GHz RADAR remaining in the band cannot make the parking sensors superfluous.

The ultrawide band (UWB) technology offered a way out that was only tolerated temporarily. Although a carrier frequency of 24.15 GHz is also used for this technique, only very short low-energy pulses are transmitted. The pulses, which are only around 0.5 ns long, lead to an effective bandwidth of 5 GHz (→ UWB band 21.65–26.65 GHz). Although they remain below the permit limits of the adjacent bands with the energy distributed over the entire width, this is still not deemed acceptable. For example, UWB RADAR must be switched off in the vicinity of radio astronomy stations resulting in forced coupling with a positioning system.

Since 1 July 2013, 24 GHz UWB RADAR is only permitted to commission with a reduced band (24.25–26.25 GHz); as of 1 January 2018, it will longer be permitted at all. Thus, only the 77–81 GHz range is still available, even though this is linked to costs that are initially even higher and is also not permitted in all countries. This new band is so generous that modulation methods other than UWB are possible in order to achieve a small distance cell. Details can be found in the interface description of the German Federal Network Agency (German Federal Network Agency 2005).

The 24 GHz band offers a middle route with the wide band low activity mode (WLAM) (cf. (ETSI EN 2013)), providing a total 450 MHz bandwidth between 24.05 and 24.50 GHz to offer better functionality for a short time for automatic calibration, for emergency stop situations, and for reverse parking than with narrowband RADAR devices that are continuously offering 200 MHz.

6 Signal Processing and Tracking

Signal processing also takes place for various modulation and antenna concepts in mainly the same order as shown in Table 1.

It begins with **signal forming**. In all concepts, this includes signal modulation, e. g., stair generation according to Figs. 9, 10, 11, and 13 or ramp generation according to Figs. 14, 15, 16, and 17. If the antenna characteristic is also dynamically changed (e.g., by scanning according to Fig. 20), then this must also be counted as signal forming.

The first processing step with received signals consists of **preprocessing and digital data acquisition**. This step combines demodulation and digital data acquisition and often contains adjustment filters, e.g., in order to compensate the drop in receiving power linked to the distance. The analog signals are sampled after demodulation and amplification and converted into digital values. In this case, both classic parallel converters and also Σ-Δ converters may be used. The latter are 1 bit converters with oversampling and downstream digital filter. However, they

Table 1 Generalized work steps of RADAR signal processing

Processing step	Explanation
Signal forming	Modulation (frequency stairs or ramps, pulse generation), beam switching or forming
Preprocessing and digital data acquisition	Demodulation, amplification, digital data acquisition
Spectral analysis	Mostly one- or two-dimensional (fast) Fourier transform of the digital data; in this case the frequency location and the complex amplitudes contain information about distance, speed, and azimuth angle
Detection	Identification of peaks in the spectrum, mostly by comparison with an adaptive threshold
Matching	Assignment of detected peaks to an object
Determination of azimuth angle	Determination of the azimuth angle by comparing the amplitudes of different receive paths with antenna characteristic
Clustering	Combining of detections that probably belong to one object
Tracking	Assign current object data to previously known objects (association) to obtain a chronological data track that is filtered and from which the object data for the next assignment are predicted

are not suitable if the input channel is switched over during the measurement (multiplexing).

The data volume corresponds to the number of cells according to Fig. 27, i.e., one measured value per cell. It may be between one thousand and one million values depending on the concept.

In all modern ACC RADAR sensors, the **spectral analysis** performed by Fourier transform plays an important role in preprocessing the signals. Put simply, the Fourier transform is a calculation-intensive conversion from the time domain to the frequency domain and vice versa. A sequence of measured values defined in discrete time steps becomes a sequence of discrete "measured values" defined in frequency steps which determines the frequency spectrum. Leading-edge signal processors are powerful enough to perform this transformation, even with many measurement points (order of magnitude 1,000) in just a few milliseconds. However, this high transformation speed is only achieved if the number assumes certain values. In the classic fast Fourier transform (FFT) algorithm, it must be a power of 2 (e.g., 512, 1024, 2048).

Windowing is normal in conjunction with the spectral analysis in order to prevent artifacts due to limiting of the measuring window (so-called leakage errors). Even if different window functions optimized to different criteria can be used for this, it leads to an effective cell enlargement by approximately 1.5 times in every dimension with correspondingly worsened accuracy and separation efficiency.

Detection is the search for special features in the measured data series. Often these are peaks in a spectrum, be it a frequency or time-of-flight spectrum. The aim is to identify the reflection signals of individual objects and to differentiate them from those of other objects. Due to the very different signal strengths of the various objects, but also of the same objects at different times, a threshold algorithm has to be found which finds all the peaks possible that originate from real objects but is nevertheless insensitive to peaks that have arisen due to noise or disturbance signals. Therefore, mainly adaptive thresholds are used, as in the example spectrum in Fig. 28. If systematic peaks occur that are not attributable to external reflections, they must be masked in the same way as any ground reflections. Unfortunately, strong reflections of a real object also make detection difficult. On the one hand, they may obscure objects that reflect more weakly in adjacent frequency ranges, particularly if the transmit frequency does not ideally follow the modulation curve. The reasons for this are phase noise of the oscillator and linearity errors in the FM method. On the other hand, deviations from the mixer characteristic (see Sect. 2.3) lead to harmonics, as can also be seen in Fig. 28. In fact, the "harmonic targets" have greater distances but also correspondingly multiplied relative velocities, making it possible to calculate from such artificial object data significantly larger object delays for the approximation than for the original object.

Matching is understood to be the assignment of detected peaks to an object, where this can mean both the assignment of various spectra (of, e.g., different measuring ramps of an FMCW RADAR) of a beam and also the assignment of peaks of different beams. At the same time, object data of past measurement series

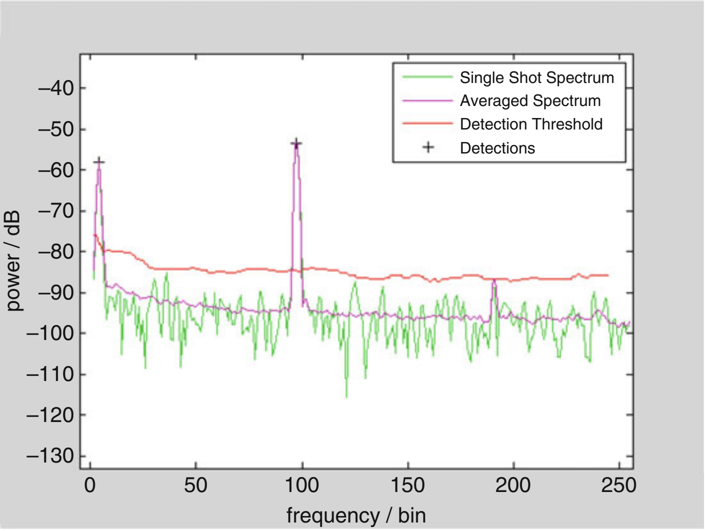


Fig. 28 Example spectrum for an FMCW measurement. In addition to the actual target at the approximately 95th frequency line, there is also a close-range echo (in the first lines) of the target simulator's antenna horn and the harmonic of the target (at line 190) (Source: Bosch)

may also be consulted so that during assignment, it is possible to become more selective as a result of plausibility considerations for potential ambiguities. This applies particularly to the matching of FMCW.

Determination of the azimuth angle is carried out via the (complex) amplitudes of the peaks that have been measured in various beams of an object. The angular position of the object can be determined if the angle characteristic and the beam direction are known. For a scanner concept with continuous angular velocity, however, the angle can also be determined during detection.

"Too much" information is created particularly with high-resolution RADAR sensors. For example, with a small distance cell, 5–10 reflections are quickly detected from a truck, or, with high relative velocity resolutions, the relative movements of connected objects (tractor, trailer, load, or limbs of pedestrians). Therefore, by means of heuristic **clustering**, attempts are made to link the detections of the same object and to show it as only one object in the measuring list.

Tracking is understood to be the formation of a chronological relationship between individual measuring events and quasicontinuous "tracks" of individual objects. Detection and subsequent clustering lead initially to individual object hypotheses which are provisionally only applicable to this one cycle. In tracking, initial attempts are made to assign to hypotheses of previous cycles (association). These object hypotheses are normally organized in lists and have object identifiers as "individual" tags. For the association, the state variables of the previously known objects (e.g., distance or lateral position) are predicted to the time of the current measurement. The current object hypotheses are then assigned to the existing hypotheses, a search window being placed around the predicted values because both measuring and also prediction errors are to be assumed. The track is continued if a currently identified object can be assigned to a previous object in this search window. At the same time, the object quality is increased or remains at a high level. If objects of the current measurement are left over, then new objects are generated in the object list and initialized with the measured date of the current measurement. However, this object starts its track with a low object quality that is generally so slight that one or more congruities are required in subsequent measurements before this object qualifies for use (e.g., ACC) as a target object.

If no hypothesis for an existing object can be assigned from the current measurement, the object quality decreases. After failing several times, the quality falls below a defined threshold value whereupon this object is removed from the object list. In addition to these basic cases, possible ambiguities also have to be considered, for example, that a current object falls in the search window of other objects or that several individual objects of the list belong to a single real object.

In addition to association, a mostly very application-specific state-estimation filter takes place with the tracking, often combinable with the association step as a Kalman filter which already implicitly contains the prediction step needed for association. The state variables of objects detected with active sensors always contain the distance in x and y direction, the relative velocity, the acceleration in longitudinal direction, and sometimes the transverse speed. If the state variables of the ego vehicle are consulted, it is also possible to create the absolute object variables for velocities and accelerations and to record them in the object list. Therefore, it is also possible to perform a differentiation between moving (in same direction), stationary, or oncoming objects. As the objects are based on a history due to the tracking, this history can also be used to differentiate between stationary and "stopped" objects. These differentiations represent the main classes of a classification even if it is a simple one. The nonreaction of conventional ACC systems to stationary objects is based specifically on this classification and not on the frequently referred to but nevertheless incorrect assertion that no stationary objects can be detected with RADAR.

In modern RADAR sensors, the rough classification referred to has long ceased to be adequate. Even if the receive amplitudes of an individual measurement have little meaning, the observation of the backscatter amplitudes over time, particularly if the distance changes significantly in the process, provides helpful information on classification of the targets. As mentioned in Sect. 1, the fluctuation arising due to the multipath reflection can be used specifically to draw conclusions about the target's height.

Following on from the signal processing steps referred to, interpretation of the situation begins which in the simplest manner takes over the selection of a RADAR target from the object list. Target selection and also wider interpretation of the situation depend heavily on the application and must be described as a part of that application. Target selection for ACC can accordingly be found in ▶ Chap. 45, "Adaptive Cruise Control," Sect. 7.

7 Installation and Adjustment

Basically, two concepts are feasible for the installation of the RADAR sensor: invisible with an optical cover of the antenna and visible without any optical cover. An optical cover is certainly more design-friendly than direct visibility of the RADAR sensor even if it can be argued that this is the only way that the "RADAR sensor status symbol" can come into its own. An important feature for the cover, also known as a radome, is that the RADAR beams are only slightly weakened and that the angle characteristic does not lead to any unexpected change. Plastic materials as a cover are not considered problematic. With a thickness several times half the wave length $\lambda'/2$; $\left(\lambda' = \lambda/\sqrt{\mu_r \varepsilon_r}\right.$; at 77 GHz $\lambda/2 \approx 2$ mm, for plastic material $\mu_r \geq 1; \varepsilon_r \approx 2 \ldots 2.5$), the fractions reflected on the exit surface are amplified if they are likewise also thrown back again by the entrance surface. For larger angles, however, the difference in the time of flights decreases between the direct and twice-reflected signals such that the superposition can lead to changes in the resulting phase. Moreover, the transmission and reflection rates on the boundary surfaces themselves depend on the angle (and the polarization), which means that the changes in the signal processing brought about by the cover must be taken into consideration at least in the case of larger angles. Nonmetallic paint poses no problem; metallic paint on the other hand can lead to significant problems. In this case, the repainting specification, which permits three coats of paint, is particularly problematic. A metal cover is completely unsuitable, of course, if the penetration depth is less than the material thickness. Very thin layers (<1 μm) may again be transparent for mm waves without losing their metallic reflecting property for optical waves. This is utilized to reproduce metallic structures (radiator grill, brand logo) on plastic surfaces. Thus it is possible to design a RADAR cover that is quite hard to recognize.

The RADAR sensors are usually attached at three points, easily visible in the example shown in Fig. 29. In this case, a holder functions as a coupling element to the car body or the chassis. The sensor can be rotated both in azimuth ϕ and also in

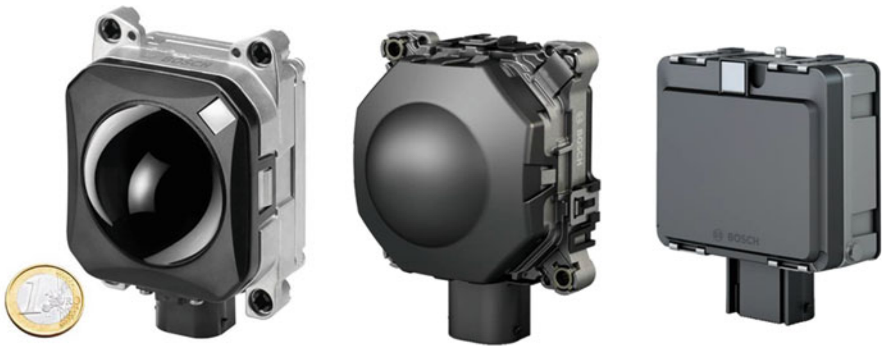


Fig. 29 Bosch long-range RADAR sensors of the third and fourth generations (LRR3, LRR4, MRR) (Source: Bosch)

elevation ϑ via the screw connection with the holder which can be used for alignment at the end of the vehicle assembly process or in the workshop.

In all, three error sources must be considered for azimuth ϕ_{err} and elevation ϑ_{err}:

- Errors in the alignment within the sensor ($\phi_{\mathrm{err,internal}}$, $\vartheta_{\mathrm{err,internal}}$)
- Errors during alignment of the sensor on the vehicle ($\phi_{\mathrm{err,fitting}}$, $\vartheta_{\mathrm{err,fitting}}$)
- Alignment error of the sensor holder = ego vehicle due to a pitch angle $\vartheta_{\mathrm{err,veh}}$ deviating from the construction position or a sideslip angle $\phi_{\mathrm{err,veh}}$ ("crabbing") occurring even when driving straight ahead

In the RADAR sensors used nowadays, misalignments in the elevation "only" lead to the reduction of the detection range or to reduced accuracy of the azimuth angle but not to systematic measuring errors. Consequently, there is no need to continuously check the alignment. As a result, it is sufficient for the sensor-axis alignment to "horizontality" to be carried out at the end of production or in the workshop. It is possible to compensate $\vartheta_{\mathrm{err,fitting}}$ with a sensor housing-side mirror or a spirit level mounted on the sensor. With a reference measurement using a metal mirror, in which the mirror level is moved into three positions, it is even possible to compensate the sum error $\vartheta_{\mathrm{err,fitting}} + \vartheta_{\mathrm{err,internal}}$.

Presetting in the assembly or in the workshop must also be carried out for the azimuth angle. For this, a reference is created for the X_{V} direction of the vehicle using the methods customary in chassis measurement. Via mirroring with housing-side reflectors (cf. Fig. 29), it is possible to compensate $\phi_{\mathrm{err,fitting}}$ as long as the supplier has guaranteed that the sensor axis has been appropriately aligned with the housing axis. The sum error $\phi_{\mathrm{err,\,fitting}} + \phi_{\mathrm{err,\,internal}}$ can be compensated in the direct method of measuring the zero azimuth angle by means of the sensor while using a mirror positioned in the $Y_{\mathrm{V}} - Z_{\mathrm{V}}$ plane. Determination of the offset in operation is essential in the case of the azimuth angle, because the static sideslip angle ("crabbing") only shows up when driving, unless it has been determined in advance using a chassis dynamometer. Moreover, uncertainty still remains with regard to the other angle errors. Azimuth offset estimation methods are necessary due to the target selection's high sensitivity to azimuth errors. Basic information provided by these estimators are the averaged gradients of the lateral offset as a function of the longitudinal distance, corrected by the apparent movement of the object caused by the rotation $\dot{\psi}$ of the ACC vehicle about the center of the circle M:

$$\phi_{\mathrm{err}} = \overline{-\left(\frac{\partial Y_{\mathrm{S}}}{\partial r}\right) - \frac{\dot{\psi} \cdot (r - X_{\mathrm{MS}})}{v}} \tag{61}$$

X_{MS} is the distance from the sensor to the instantaneous center of rotation projected onto the X axis and takes the sideslip angle into account. In drift-free travel, X_{MS} is equal to the negative distance from the rear axle to the sensor. Otherwise, X_{MS} must be determined from dynamic driving considerations.

Alternative approaches to determining the azimuthal sensor offset, e.g., with the help of the assumption that the target objects travel on average without lateral offset to the ACC vehicle, can be added to this. They suffer, however, from the severe simplification of these assumptions.

If the RADAR sensor has wide azimuthal coverage, it may be possible to dispense with precision adjustment ex factory or workshop. The concurrent estimate of azimuthal offset would then have to converge quickly and safely.

An online estimate of elevation offset is known from Continental ARS 300, Sect. 8.3. In this case, the measured distance of the ground echo is utilized specifically. An adjustable elevation pivoting device briefly tips the RADAR beams 7° towards the ground and measures the distance of the ground echo. If the installation height is known, it can be used to determine the elevation and therefore correct it. However, here too it is stated that only an elevation loss angle of 0.5° will be tolerated.

8 Electromagnetic Compatibility

Basically, the requirements that are applicable to control units in the car also apply to a RADAR sensor. In addition to conformity to the frequency regulation, attention must be paid to robustness against disturbances due to other RADAR sensors.

The disruption by other sensors may overdrive the input stage. They must, therefore, be designed so that this disruption (clipping) has no effect or is at least detected and possibly displayed to the driver as a fault indication. There is virtually no fear of false measurement resulting in ghost targets as no target would be selected as a relevant output variable that has only been detected once (see Sect. 5 about tracking).

The probability of being disturbed synchronously by a different RADAR sensor is extremely small and is substantiated by the next example. With measurements repeated at approximately 100 ms intervals, for a successful association, the relative velocity and the distance would only have to differ very slightly from the predicted values (max. 5 m and 2 m/s, respectively). This would require a reproduction accuracy of the disturbance of approximately 20 ns in time and 1 kHz in frequency. Even normal quartz time bases can barely provide the high relative accuracy $\left(\Delta t/t = 2 \times 10^{-7}, \Delta f/f = 1.2 \times 10^{-8}\right)$ required for this. If the cycle time then fluctuates as well due to desired jittering or to asynchronous cycles, the probability of repeated errors drops to such low values that this type of disturbance does not lead to “ghost targets” in practice. However, other RADAR sensors may interfere with their radiation similarly to noise and thus contribute to a loss of sensitivity.

The information referred to here can also be found as findings of the European MOSARIM project carried out between 2010 and 2012 (MOSARIM 2013) during which the sensitivity in relation to radiation from other sensors was investigated both experimentally and also by means of simulation. As RADAR sensors have since achieved wide distribution and are aligned both forward and backward in relation to the vehicle’s direction of travel, the probability of interference disturbance has also increased, which is why countermeasures appear necessary (cf. Work Package 5.1 of MOSARIM (2013)). Subbands within current bandwidths that are either dynamically accessed or separated according to alignment are particularly effective, although only suitable for frequency-modulated RADAR sensors. The latter address the particularly critical case of mutual disturbance of cars which are traveling one behind the other that can disturb each other in a different way to oncoming cars over time. The choice of polarization could also be used for preventing interference, although a corresponding standardization would be too late for the solutions already on the market. For pulsed methods, which also include the chirp sequence modulation and FSK variants with repeated stairs, a timing jitter in the repetition offers an opportunity for suppression (see also Massen and Möller (2012)). So a shift of 100 ns in the repetition period changes the target position in another system by 15 m but at 77 GHz causes just a relative velocity error of only 5×10^{-3} m/s.

9 Examples from Industry

9.1 Bosch LRR3

The third generation of Bosch long-range RADAR sensors has been in use since 2009. As in previous generations, this is a 76.5 GHz RADAR with integrated control unit. The high integration of the necessary components meant that it was possible to implement a housing with a volume of only 1/4 l (Fig. 30). A die-cast aluminum subcarrier accommodates an HF and an LF circuit board.

Generation of the high-frequency transmit power is initially monolithically integrated, based on silicon-germanium (SiGe) MMICs. Although other manufacturers had used monolithic mm-wave ICs earlier, based until then on the expensive gallium-arsenide (GaAs), SiGe nevertheless offers more cost-efficient production conditions due to its wider range of uses. The integration gives rise to a variety of new possibilities in the transceivers' field which was utilized particularly for the receiving electronics in the LRR3. In this case, Gilbert cell mixers are used. On the one hand, they keep the conversion losses small and therefore permit a lower peak power. On the other hand, they make it easily possible to modify the mixer amplification of the individual receive paths and thus to adjust an adapted antenna characteristic. If the sensor is to be used as an individual RADAR sensor, the aim is a symmetrical radiation characteristic with equally high inner and outer lobes in each case. For a dual arrangement of RADAR sensors according to Fig. 25, the lobes of one sensor are designed asymmetrically in order to achieve extended coverage when the fields of view of both sensors overlap.

Apart from the MMICs, a RADAR ASIC which is responsible for generating the signal modulation and scanning of the four receive channels is also located on the HF circuit board. In this case, analog-to-digital conversion is based on an

Fig. 30 Construction of the Bosch RADAR sensors MRR and LRR3 (Source: Bosch)

overclocking sigma/delta converter and a decimation filter. Further signal processing, including spectral analysis, then takes place on the low-frequency circuit board on a μ-controller which also has the usual functions of an automotive controller. Moreover, a multifunctional ASIC is also present on this circuit board which takes over monitoring, diagnostic, and voltage supply functions. The sensor hardware architecture of the LRR3 is illustrated in Fig. 31.

Other advantages over previous sensors have also been achieved particularly in the area of angle determination where the angular range has been expanded by improving the algorithms and specific changes to the antenna characteristic. As a result, it is possible to measure 20° overall in the medium range (30–100 m) and as much as 30° in the close range at less than 30 m. Other parameters of this sensor can be taken from Table 2. A further hardware innovation is a FlexRay transceiver, meaning that another bus interface now exists in addition to the CAN.

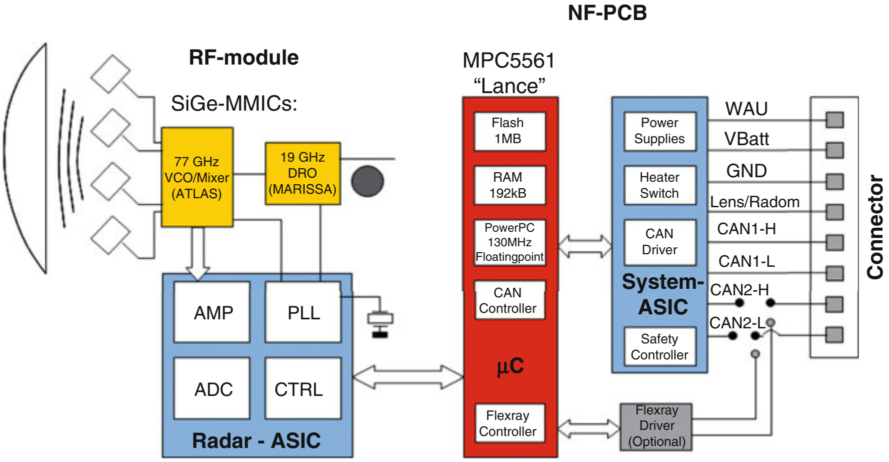


Fig. 31 Hardware architecture of the Bosch LRR3 sensor (Source: Bosch)

9.2 Fourth-Generation Bosch RADAR Sensors

A modular system (Hildebrandt et al. 2013) has been developed to cover the sharp rise in the number of different application domains and the increasing variety of functions. In addition to the purely sensory aspects, a high level of flexibility is needed with regard to integration in a comprehensive system architecture (Classen et al. 2012). By way of example, Fig. 32 illustrates a configuration for detecting the vehicle environment with five RADAR sensors. In the fourth generation of radars, the product variants LRR4, MRR, and an MRR dual mode (rear/corner) are offered for this. Corresponding product images can be seen in Fig. 29.

The LRR4 is therefore the logical further development of the long-range RADAR LRR3 with lens antenna and very long range. To take account of the increasing spread of RADAR-based assistance systems, the MRR variant was additionally developed with a planar antenna system and reduced range. To implement future requirements in the area of pedestrian protection (Schubert et al. 2013), it is possible to switch over as necessary to an alternative transmit antenna with wide opening angle (see Fig. 33 left). The principle of two switched transmit antennas was also implemented in the MRR rear, the main beam directions in this case being configured in different directions. In this way, for example, a rear sensor can be implemented with extended visual range which, on the one hand, can detect traffic approaching from the rear up to a distance of approximately 80 m and, on the other hand, can detect traversing objects with the same range (Fig. 33, right). Further areas of use can be covered by varying the antenna layout. The technical parameter of the product variants is illustrated and compared to the values of the previous generation in Table 2.

Figure 30 shows the individual assemblies and the housing concept of the MRR. The modified antenna system with planar patch arrays without lens is obvious

Table 2 Technical data of the third and fourth generations of Bosch long-range RADAR sensors

	Generation 3	Generation 4		
Bosch product variants	LRR3	LRR4	MRR	MRR rear/ corner
	2009	2015	2013	2014
General characteristics				
Dimensions (W × H × D)/ mm^3	74 × 77 × 58	78 × 81 × 62, dimensions incl. mounting lugs	60 × 70 × 28	
Mass	285 g	240 g	190 g	
Cycle time	<125 ms	60 ms	60 ms	
High-frequency module				
Frequency generation	MMIC/ SiGe, bonded 18.9 GHz reference oscillator, PLL	SiGe MMIC, eWLB package standard soldering process 18.9 GHz reference oscillator, PLL		
Beam forming	Patch + dielectric lens, monostatic	Patch + dielectric lens, monostatic	Bistatic, patch arrays, including digital beam forming	
Radiated power (EIRP peak)	33 dBm (EIRP)	<29 dBm	<34 dBm	<18 dBm
Signal characteristics				
Frequency range	76–77 GHz	76–77 GHz		
Modulation process	FMCW	FMCW		
Ramp height (typical)	500 MHz	425 MHz	425 MHz	425 MHz/700 MHz
Ramps (number/typical duration)	4 (6.5/1/7/ 11.5 ms)	5 (2.1..9.6 ms)	5 (1.3..5.9 ms)	4 (1.1..4.6 ms)
Number of measurement ranges	1	1	2	2
Type of angle measurement (azimuth)	Quad-beam concept	Six-beam concept	Four-channel with phase evaluation	
Type of angle measurement (elevation)	–	Amplitude monopulse (2Tx)		
Detection characteristics				
Distance range	0.5 … 250 m	0.36 … 250 m	0.36 … 160 m	0.36/0.23 … 80 m
Distance cell	0.3 m	0.36 m	0.36 m	0.36/0.23 m
Relative velocity range	−80 … +30 m/s	−80 … +30 m/s	−80 … +80 m/s	

(*continued*)

Table 2 (continued)

	Generation 3	Generation 4		
Bosch product variants	LRR3	LRR4	MRR	MRR rear/ corner
	2009	2015	2013	2014
Relative velocity cell	0.2 m/s	0.2 m/s	0.33 m/s	0.43 m/s
Az. measuring ranges	12° Long range 20° Midrange 30° Short range	12° (200 m) 20° (100 m) 30° (30 m)	12° (160 m) 18° (100 m) 58° (60 m) 90° (25 m)	See field of view
Az. angle cell (defined via 1/2 of the separation efficiency of two point targets)	2°	2°	3.5°	–
Az. accuracy of point target	0.1°	0.1°	0.2°	0.3°
Elev. accuracy (typical)	–	0.2°	0.6°	–
Elev. lobe width (6 dB)	5°	4.5°	13°	13°

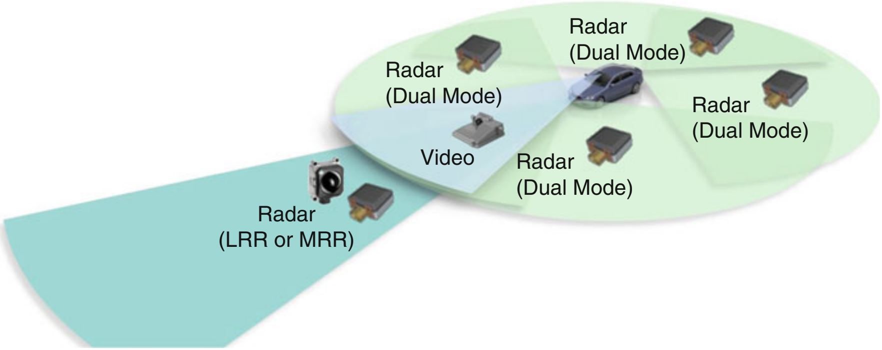


Fig. 32 Areas of use for RADAR sensors in vehicle environment sensor technology (5R1V) (Source: Bosch)

compared to the LRR3 sensor. The block diagram of the MRR can be seen in Fig. 34. The μ-controller is now also accommodated on the underside of the enlarged HF circuit board along with the RADAR ASIC.

Furthermore, the technologies and production methods used in the fourth generation have been enhanced in line with the steep rise in production figures: this also includes, for example, the transition to standard soldering processes for the SiGe MMICs. Figure 35 shows the HF circuit board with the antenna structure and the soldered transmit and receive MMICs. The receive antenna fields to be seen in the

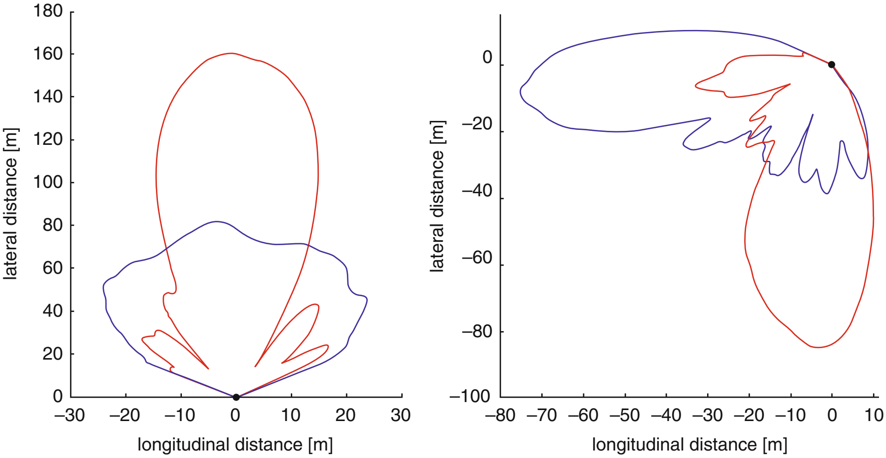


Fig. 33 Typical detection areas of the MRR sensor as a front sensor (*left*) or rear sensor (*right*). Both antenna switching conditions are illustrated in each case (Source: Bosch)

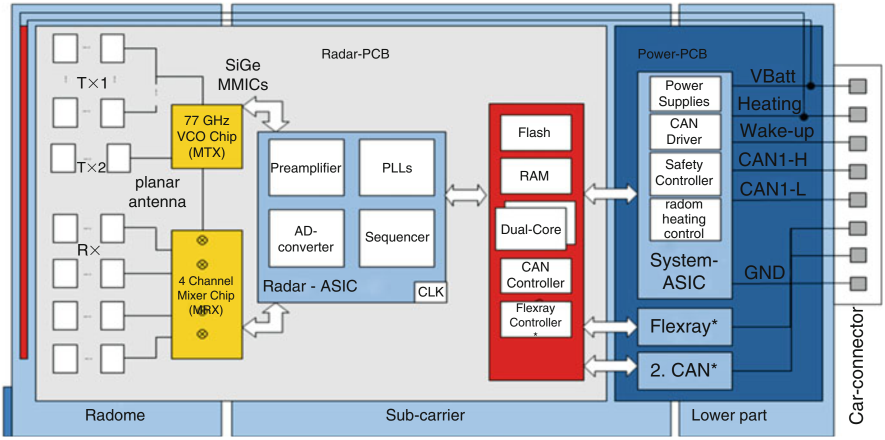
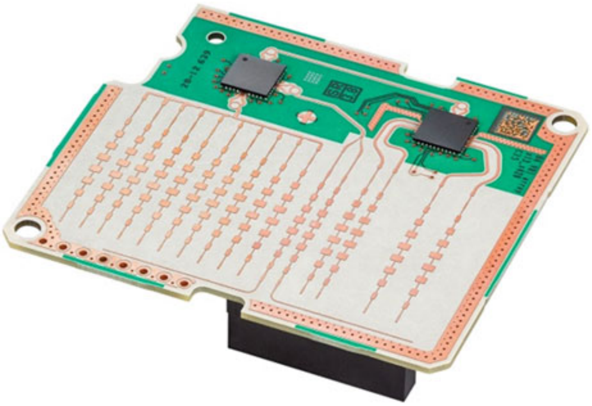


Fig. 34 Hardware architecture of the Bosch MRR sensor (Source: Bosch)

diagram on the right (four columns in total) form a "thinned out" array with the spacings of 0.5, 2.0, and 3.0 times the wave length (in relation to the left-hand side column in each case). Thus, in spite of the restrictions to four columns, an angle uniqueness of $\pm 90°$ and a beam width (separation efficiency) of $\arcsin(1/3) \approx 20°$ are achieved, as would otherwise be achieved by seven columns evenly spaced by $\lambda/2$. The parametric deterministic maximum likelihood method is used for the estimation of the angle so that the separation efficiency of two targets can be improved on significantly smaller angle ranges – in this case to 7°. The different sizes of the patches in the vertical arrangement (tapering) are therefore used to reduce side lobes in the elevation.

Fig. 35 HF circuit board of the Bosch MRR sensor with antenna structures and SiGe MMICs, from *left*: 10 strips for Tx1, then two strips for Tx2, and then four strips for Rx1 (Source: Bosch)



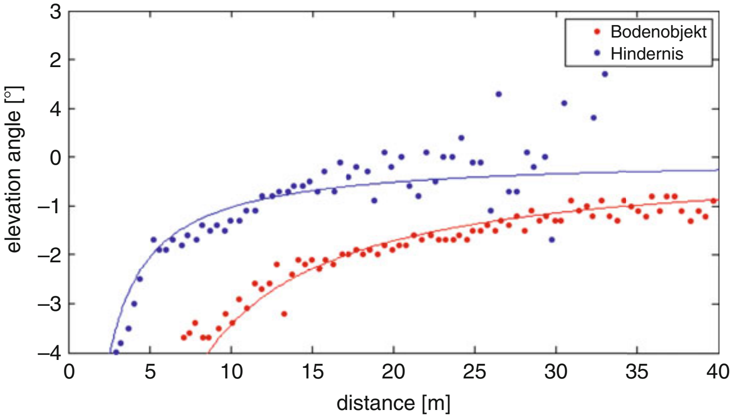


Fig. 36 Angle of elevation measurement at raw signal level on approaching a ground object (*red*) or a target a little below the sensor level (*blue*). The measurements were performed with an MRR sensor (Source: Bosch)

For long-range application, a total of ten transmit columns are used; for close-range application, two, each spaced by $\lambda/2$. In the MRR rear, five transmit columns fed by phase-shifted transmit signals are used in each case, resulting in a main radiation direction of approximately −45° for one antenna and approximately +45° for the other.

When implementing extended safety systems (cf. ▶ Chaps. 46, "Fundamentals of Collision Protection Systems," and ▶ 47, "Development Process of Forward Collision Prevention Systems"), the classification of stationary objects plays a major role. Measuring the object angle in the elevation direction as well is one possibility. In the LRR4 and MRR variants, the second transmit antenna can be used for this if its main direction of elevation deviates from that of the other antenna (amplitude monopulse principle). In this context, Fig. 36 illustrates the raw measured values on approaching two objects of different height, the distance in each case being related to the object concerned. The red measurement points represent the elevation angles of a target lying on the ground and should be classified as being traversable. The blue measurement points belong to a target slightly below the sensor height and should therefore be seen as an obstacle. In this measurement, in the close range at least, the two object classes can be clearly separated based on the elevation angle. It should be noted, however, that the measurement of raised targets is subject to the systematic effects of ground reflections described in Sect. 1.

9.3 Continental ARS 300

The ARS 300 (Fig. 37) uses a mechanical scanning principle that works with a roller and both a transreflector and a twist reflector. The latter, as shown by Fig. 38, is also used for focusing as with an offset parabolic antenna. The twist reflector is movable and therefore can also pivot in the elevation direction which is used for elevation offset identification and correction. This reflector arrangement is not fed by individual feeds but by a dielectric leakage wave guide. This wave guide guides the microwaves without escape unless they experience scattering on the underside due to the roller's grooves. The scatter amplitude of all grooves leaving the wave guide forms together a level wave front that is aligned by the gap between the grooves. If the gap between the grooves is smaller than the wave length in the wave guide, the diverted wave will point towards the left if fed from the left-hand side; with wider gaps, the radiation turns to the right. Thus, the gaps between the grooves vary in the peripheral direction. This type of scanning does not require any back rotation as when pivoting an antenna and generates negligibly low mass forces.

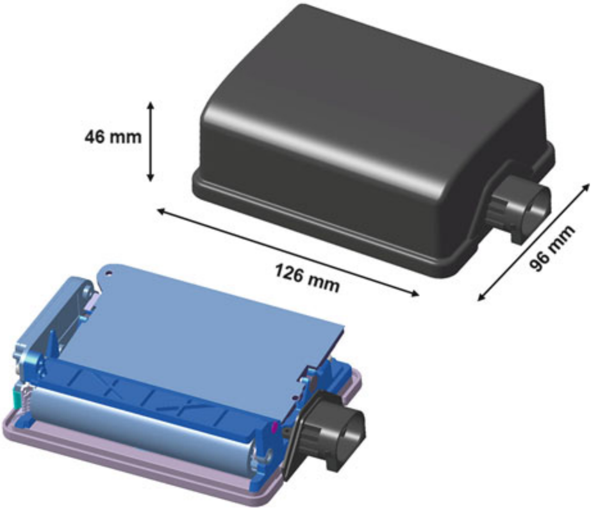


Fig. 37 External view and construction of the RADAR sensor ARS 300 by Continental (Source: Continental)

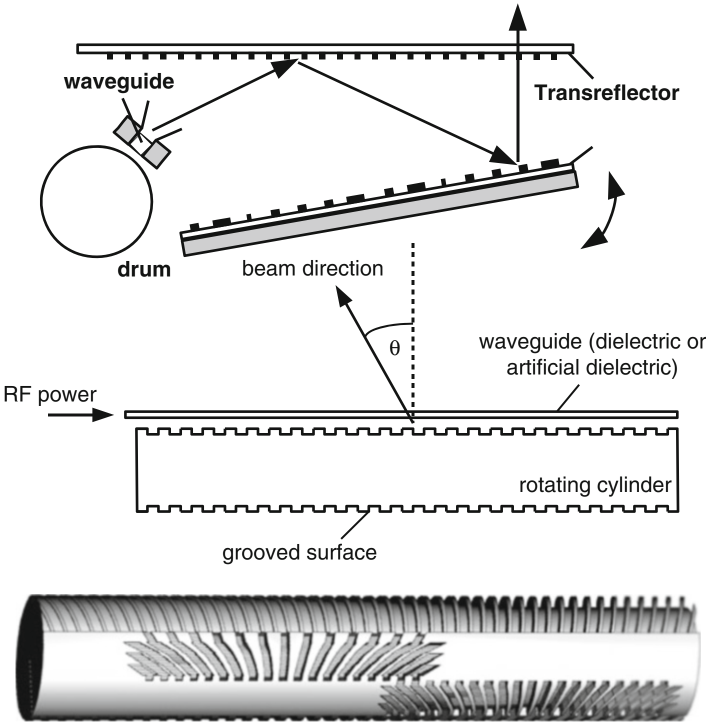


Fig. 38 Antenna concept of the RADAR sensor ARS 300 by Continental, *top*: lateral view (elevation), bottom roller arrangement for azimuthal scanning (Source: Continental)

In addition to scanning for the central field of view (17°), the fields of view at higher angles are also detected by scanning; further grooved areas on the drum are milled in asymmetrically (see Fig. 38 bottom). The offset on the drum allows for the offset needed for an approximately central beam path with illumination to the side, i.e., the area is displaced to the opposite side of the illuminated area. The beam width in the azimuth is 2.5° (in the short-range scan 8°), in the elevation 4.3°. With the pivotable twist reflector, the ARS 300 can be optimally aligned with regard to the elevation and can compensate stationary pitch angles or maladjustments.

Unlike earlier generations, the modulation is a chirp sequence modulation, also described as pulse compression (cf. Sect. 2.5.4). This approach makes maximum use of the measuring time, although costs are incurred for a high sampling rate ($\approx$40 MHz), the approach having already been needed for the ARS200. As stated in Table 3, a distinction is made between a near-range scan (up to 60 m) and a far-range scan (up to 200 m). In this case, with the near-range scan, evaluation is only carried out up to 60 m, meaning that the sampling rate only has to be one third. The central area is traversed by both scans which facilitates assignment of the two measuring series. The speed cell is smaller in the central visual range – i.e., in the far range – as a longer measuring period is available for this measurement. Owing

Table 3 Technical data of the RADAR sensors by Continental

Continental	ARS 300	SRR 200
General characteristics		
Dimensions (W × H × D) (with mounting and bushing)	$141 \times 96 \times 47\ \text{mm}^3$	$112 \times 83.2 \times 25\ \text{mm}^3$
Mass	<500 g	290 g
Cycle time	66 ms	40 ms
Measuring duration in cycle	35 ms for FRS, 16 ms for NRS	20 ms
High-frequency module		
Frequency generation	GaAs MMIC (ARS 300) or SiGe MMIC (ARS301), free running VCO	SiGe MMIC
Radiated power (peak, average)	3 mW average	12.7 dBm (EIRP)
Signal characteristics		
Frequency range	76–77 GHz	24.05…24.25 GHz (ISM band)
Modulation process	Chirp sequence	
Pulse duration/ramp height and duration	187.5 MHz (750 MHz at low speed), 16 μs	187.5 MHz, 9 μs
Pulse/ramp repetition rate	50 kHz	12.5 kHz (per channel)
Number of measurement ranges	3 (1 × FRS, 2 × NRS)	1
Type of angle measurement	Scanning	Digital beam forming, eff. eight channels
Detection characteristics		
Distance range $r_{\min} \ldots r_{\max}$,	1…200 m FRS 1…60 m NRS (0.25…50 m at low speed)	0.3…100 m at ±40° 0.3…65 m at ±60° 0.3…35 m at ±90°
Distance cell Δr	1 m (0.25 m at low speed)	1 m
Relative velocity uniqueness range $\dot{r}_{\min} \ldots \dot{r}_{\max}$	−74… + 25 m/s	−40… + 40 m/s
Relative velocity cell $\Delta\dot{r}$	0.77 m/s FRS 1.53 m/s NRS	0.3 m/s
Measuring range azimuth $\Delta\phi_{\max}$	17° FRS 56° NRS	±75° Measurement ±90° Detection
Lobe width ϕ_{lobe} (3 dB one-way)	2.5° FRS 8° NRS	20° (Virtual lobe)
Angle cell $\Delta\phi$	1° FRS 3.125° NRS	14° (≈arcsin(1/4))
Elevation ϑ_{spec}	4.3° (Lobe width)	±12° at −6 dB ±16° at −10 dB
Accuracy point target (azimuth)	0.1°	±2° for ±30° ±4° for ±60° ±5° for ±75°

(*continued*)

Table 3 (continued)

Continental	ARS 300	SRR 200
Special features		
ARS 300: self-adjustment capability in azimuth and elevation		
SRR 200: self-adjustment capability in azimuth (no self-adjustment for elevation due to large elevation opening)		

Abbreviations: *FRS* far-range scan; *NRS* near-range scan

to the principles involved, the measuring time for angle determination with scanning must be divided up between the individual angle segments – in this case between 17 angular steps. The result is therefore a measuring time of 35 ms/17 ≈ 2 ms per segment. However, the RADAR lobe is wider so that a longer measuring period is available for the relative velocity. Therefore, the specified cell size of 0.8 m/s (corresponding to 2.56 ms) is also smaller. In the close range, there is a rougher segmentation of 3.125°, so that despite the lower overall measuring duration of 16 ms, a speed cell of 1.5 m/s can be maintained.

A range of less than 50 m is needed at low speeds. Consequently, it is possible to focus on the close range and by increasing the modulation bandwidth (quadrupling here) to achieve better distance resolution and separation capability.

The ARS 300 is an efficient RADAR that makes additional close-range sensors for stop-go or emergency stop functions superfluous due to very wide illumination. Good lateral resolution is achieved by means of scanning, and a lateral object separation capability is possible up to distances of 80 … 100 m. It goes without saying that this performance cannot be achieved without the price of a wider aperture and consequently wider housing (see Fig. 37), while on the other hand, the compactness in view of function is highly acclaimed.

The ARS 300 has been further enhanced without changing the basic principle of the antenna and frequency modulation. For the ARS301 variant produced since 2012, the high-frequency generation was changed from GaAs to SiGe, and the sensing algorithms were further optimized.

9.4 Continental SRR 200

The outer dimensions of the short-range RADAR SRR 200 working in the 24 GHz range are shown in Fig. 39. It is based on a planar antenna array concept. Three Rx, one Tx, and one Rx/Tx antenna fields are used, the transmitter elements being on the outside, as described in Sect. 3.6 and illustrated in Fig. 40, and thereby effectively providing eight-channel capability for digital beam forming. Digital beam forming is implemented via a DFT of length 16 – that is to say, another eight zero channels are added; while this zero padding does not bring any information gain from a signal theory consideration, the “oversampled” beam spectrum

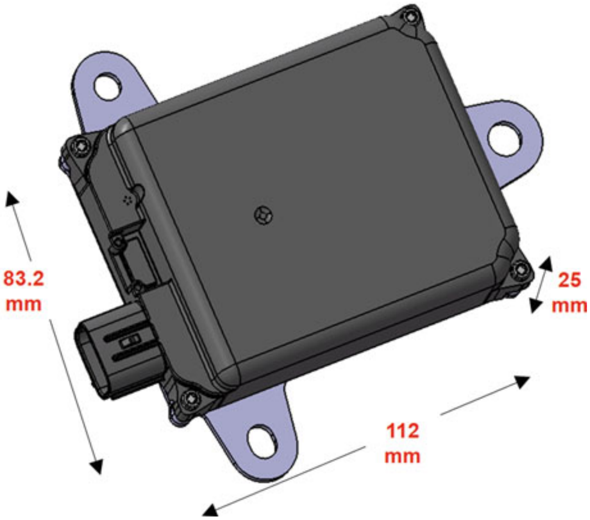


Fig. 39 Housing and dimensions of the Continental SRR 200 (Source: Continental)

resulting from it makes evaluation easier to implement. Figure 41 illustrates the 16 virtual lobes arising from the DFT. Conclusions about the signal processing can be drawn from the block diagram shown in Fig. 42. Modulation takes place as a chirp sequence similar to the ARS 300. Four chirps are transmitted alternately in each case on Tx1 and Tx2. Although the Rx paths are mixed in parallel, only one is further evaluated subsequently via a multiplex system, the selected channel being changed from chirp to chirp. Thus all eight antenna channels are acquired sequentially and this is repeated 256 times. The chronological shifts arising due to the sequential acquisition of the antenna channels can be compensated during further processing. As with the other SRR sensors, the SRR 200 is suitable for the close and medium range to the front, side, and rear.

9.5 Hella 24 GHz Midrange RADAR

The sensors developed by Hella based on the 24 GHz RADAR narrowband technology are now represented in the market in generations 2 and 3 (cf. Table 4). It is also basically possible to implement long-range sensors with this technology; the sensors introduced in this section exclusively address rear functions and accordingly have a midrange attribute of the antenna characteristic. In addition to the lane change assistance, the rear functions mentioned also include the rear cross traffic alert for the identification of cross traffic when reverse parking.

The linear frequency modulation shift keying (LFMSK/FSK) method illustrated in Sect. 2.5.2 as a modulation principle is used above generation 3 in a modified variant with non-time-constant transmit bursts. For worldwide use that is problem-free in respect to frequency regulation, the modulation bandwidth is limited to no

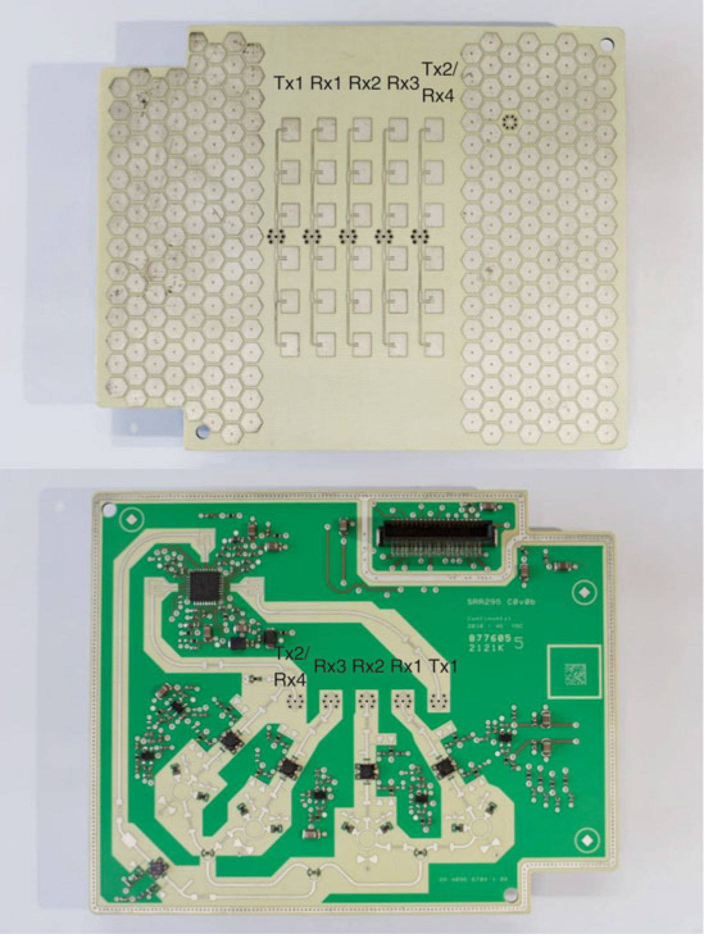


Fig. 40 RF board with antenna layout (*top*) and conductor structures (*bottom*) of the Continental SRR 200 (Source: Continental)

more than 200 MHz (24.05–24.25 GHz at transmit power <13 dBm avg. EIRP). Angle determination relies on the simultaneous monopulse method.

Generation 2 shown in Figs. 43 and 44 (left-hand side in each case) reflects the basic construction of the RADAR sensor. A radome that is neutral in respect to the beam path covers the antenna. The antenna consists of a transmit and receive path and is implemented in microstrip conductor technology. As part of the printed circuit board structure, the antenna is situated on the upper side of the RADAR front end circuit board (RFE) and consists of 1 Tx and 3 Rx elements. The remaining HF electronics are located on the underside – sometimes still in a discrete design and sometimes in highly integrated form as a GaAs MMIC. Essentially, this is where the

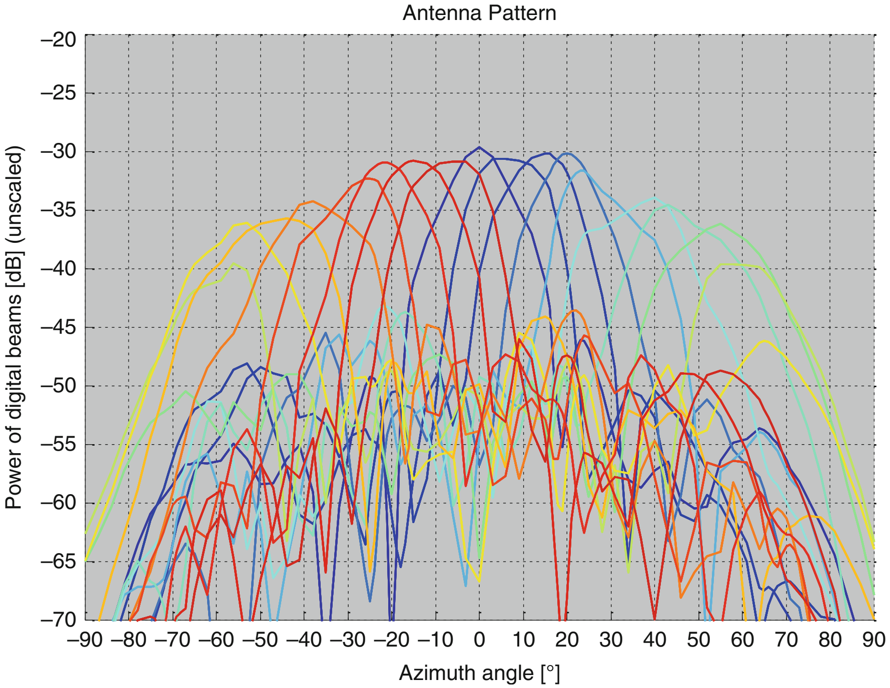


Fig. 41 Antenna diagram for the 16 virtual lobes after digital beam forming from the eight real channels and the eight datasets filled with zero, Continental SRR 200 (Source: Continental)

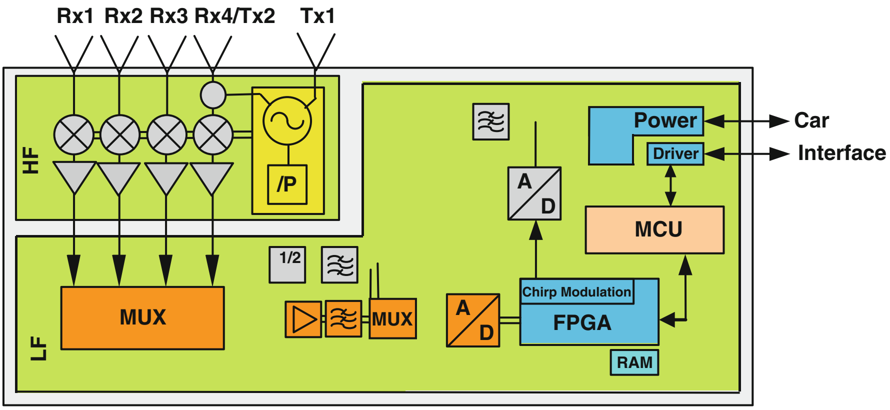


Fig. 42 Block diagram of the Continental SRR 200 (Source: Continental)

LNAs, the complex mixers, and bandpass filters are located. The HF screening ensures that HF radiation is only emitted as desired towards the radome side. The evaluation unit with patch board is mounted underneath this. A DSP on this signal processing circuit board implements the RADAR signal processing steps such as

Table 4 Characteristics of Hella 24 GHz RADAR generations 2 and 3

Hella	Generation 2	Generation 3
General characteristics		
Dimensions (W × H × D)	$105 \times 89 \times 34\ \text{mm}^3$	$98 \times 78 \times 26\ \text{mm}^3$
Mass	270 g	160 g
Cycle time	50 ms	
Measuring duration in cycle	36 ms	
High-frequency module		
Frequency generation	24 GHz VCO MMIC	
Radiated power	<13 dBm av.	
Signal characteristics		
Frequency range	24.05 GHz–24.25 GHz	
Modulation process	FMSK	Modified FMSK
Ramp height and chirp duration	Approximately 100 MHz, 36 ms	Approximately 200 MHz, 36 ms
Chirp types	Up, down, constant	
Type of angle measurement	Bistatic, simultaneous monopulse concept	
Detection characteristics		
Distance range	See Fig. 45	See Fig. 45
Distance cell	1.5 m	0.75 m
Relative velocity range	−70 m/s. . . + 70 m/s	
Relative velocity cell	0.16 m/s	
Measuring range azimuth	See Fig. 45	See Fig. 45
Lobe width	28° at −10 dB	100° at −10 dB
Specified elevation	16° at −10 dB	18° at −10 dB
Accuracy of point target	0.5°	

raw signal processing, angle determination and tracking, and monitoring and diagnostics of the RFE. In addition, a microcontroller implements basic software concerns and functions. A switch-mode power supply delivers the necessary operating voltages.

Generation 3, shown in Figs. 43 and 44 (right-hand side in each case), is based on the same RADAR principle. Unlike generation 2, the equipment design here has been consistently progressed towards size reduction and cost optimization. As a result, the entire electronic circuitry, including switch-mode power supply, communication, evaluation electronics, and HF circuitry, is located on only one circuit board. The RADAR antenna, as already shown in generation 2, is accommodated on the upper side of the printed circuit board as a microstrip conductor structure. This compact construction has been made possible using highly integrated HF components (SiGe MMICs) and by appropriately designing the layout and positioning the components on the printed circuit board. To meet the growing importance of functions that measure in the lateral region of the vehicle's environment, the transmit antenna characteristic has been designed with an all-round view in mind (see Fig. 45). It is possible to detect objects not only in the

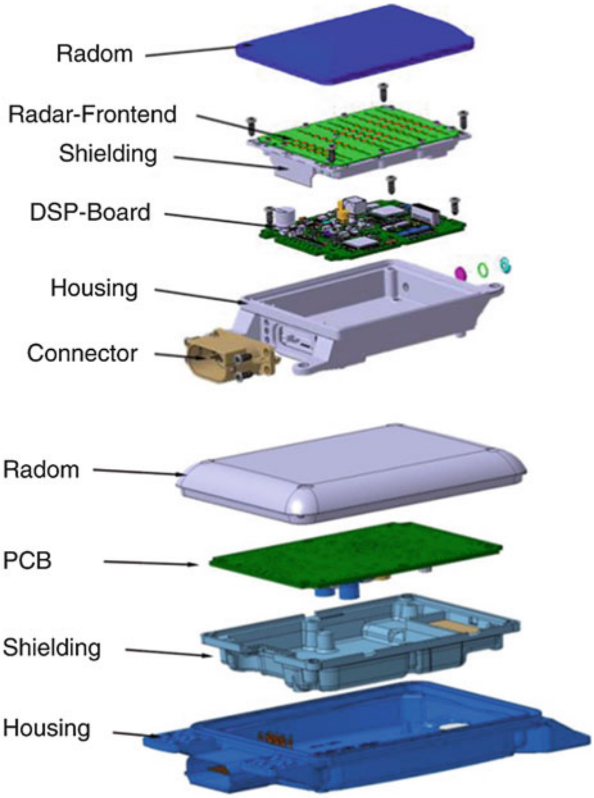


Fig. 43 Exploded view Hella RADAR 24 GHz – second (*top*) and third generations (*bottom*) (Source: Hella)

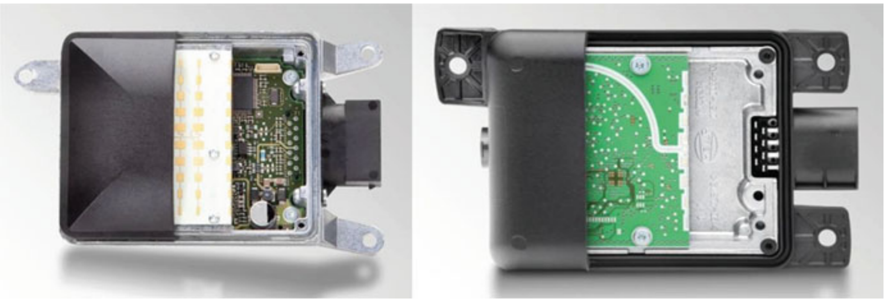


Fig. 44 Hella RADAR 24 GHz – Comparison of second and third generations. *Left*: second gen., visible: radome, RFE board (antenna side), DSP board. *Right*: third gen., visible: radome, printed circuit board (antenna side), shielding (Source: Hella)

longitudinal and lateral region of the vehicle but also under 45°. Such antenna characteristics mean that all-round vision systems can be implemented with an appropriate sensor configuration. Two Rx antenna elements are sufficient on the receive side. Naturally, such an extension of the sensor's visual range leads to

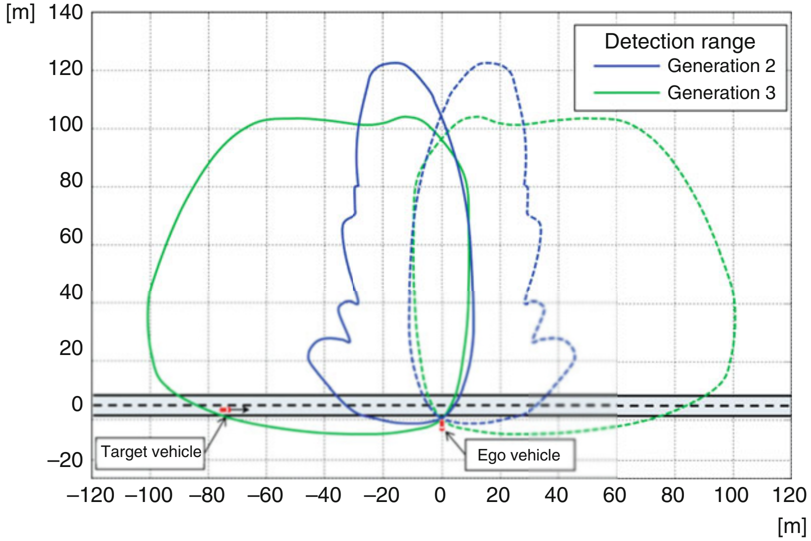


Fig. 45 Comparison of detection areas (e.g., detection of rear cross traffic). Integration of two RADAR sensors in the rear of the ego vehicle in each case (*right/left*) and sensor alignment diagonally backward (Source: Hella)

adaptation of the software, particularly in the area of raw signal processing, tracking, and formulating hypotheses.

9.6 TRW AC1000

The modular RADAR family AC1000 by TRW (Fig. 46) works in the 77 GHz frequency band and uses the silicon-germanium (SiGe) technology. Figure 47 shows the detailed illustration of the sensor, and Table 5 shows the specification. The angle evaluation is based on the principle of digital beam forming with four receive channels (cf. Sect. 3.6). As a result, several objects with the same relative velocity and at the same distance – but at different angles – can be separated and tracked simultaneously. As can be seen in Fig. 48, the digital beam forming (DBF) method permits the adaptation of the beam to the driving speed and therefore to the relevant driving conditions or application. The current design of the forward-looking AC1000 works with three modes: At higher speeds (>70 km/h driving speed), the sensor selects the long-range mode with reduced opening angle and maximum range, for example, for adaptive speed regulation on the motorway. At average speeds, a wide opening angle is set to take account of the complex urban environment and, for example, to be better able to identify pedestrians who suddenly step into the road. The third operating mode with a dedicated transmit antenna is specifically designed for low speeds and offers a particularly large

Fig. 46 AC1000 by TRW (Source: TRW)

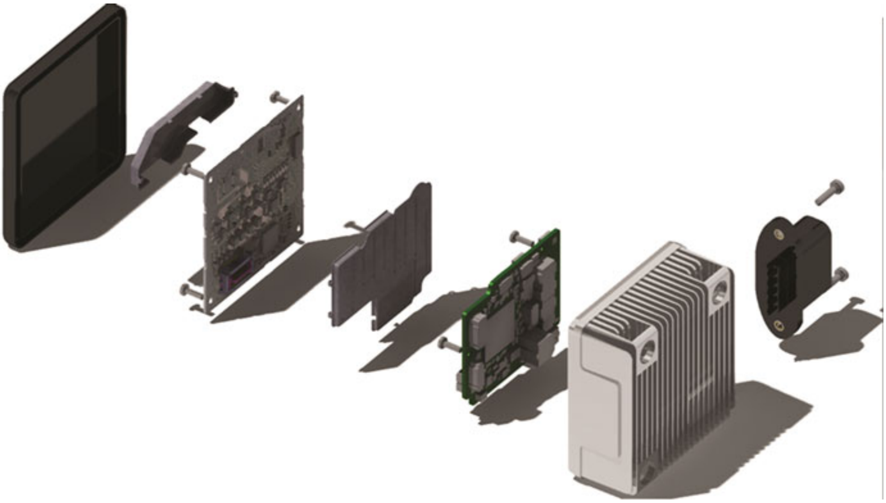


Fig. 47 Three-dimensional exploded view of the AC1000 by TRW (Source: TRW)

opening angle. A method combining FMCW and FSK is used for modulation as described in Sect. 2.5.2, Fig. 13. Thus unambiguous, direct assignment of relative velocity and distance is achieved, making it possible to dispense with subsequent removal of ghost targets.

Table 5 Technical data of the RADAR sensor AC1000 by TRW

TRW	AC1000 (forward-looking version)
General characteristics	
Dimensions (W × H × D)	75 × 77 × 38 mm^3
Mass	<300 g
Cycle time	40 ms
Measuring duration in cycle	39 ms
Signal characteristics	
Frequency range	76–77 GHz
Modulation process	FMFSK
Modulation deviation	200/400 MHz
Burst deviation and duration	0.32 MHz, 2.5 μs (1.25 μs possible)
Number of measurement ranges	3
Type of angle measurement	Digital beam forming
Detection characteristics	
Distance range $r_{\min} \ldots r_{\max}$	0.5…180 m
Distance cell Δr	0.375 m (city mode)/0.75 m (ACC mode)
Relative velocity range $\dot{r}_{\min} \ldots \dot{r}_{\max}$	> ±61 m/s
Relative velocity cell $\Delta \dot{r}$	<0.06 m/s
Measuring range azimuth $\Delta\phi_{\max}$	>70° (in low-speed mode)
Fitting tolerance elevation	±5°

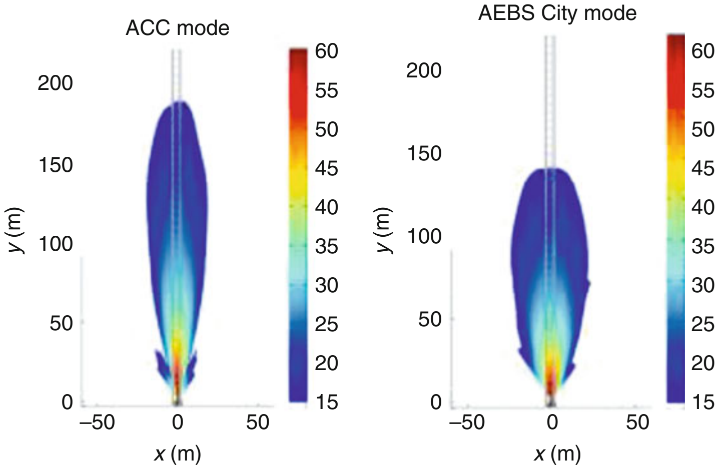


Fig. 48 Field of vision adjustment by means of digital beam forming in the AC1000 by TRW, details of color scale in dB (Source: TRW)

Due to the scalability achieved with the new AC1000 technology, radars of this type can be used individually or in groups for 360° – i.e., front, rear, and lateral regions – depending on configuration and installation location.

9.7 Valeo MBH

The multibeam high-performance (MBH) RADAR sensor, offered by Valeo since 2012, for lane change assist and cross traffic alert applications works according to the LFMCW principle in the 24 GHz range. The special feature of the Valeo MBH sensor is its multibeam transmit antenna (see Fig. 49) which divides the detection range into seven areas (cf. Fig. 50). The precise target angles are measured on the receive side according to the phase monopulse method, meaning that high phase sensitivity becomes possible due to the selectivity of the transmit beams since the uniqueness requirement for the phase difference can be dropped. This can be seen in Fig. 49 by the large distance of $3/2\lambda$ between the receive antennas and the resultant uniqueness range of $\pm\arcsin(1/3) \approx \pm 20°$ in the central region.

The LFMCW principle is a modulation similar to the chirp sequence method. Between 16 and 64 frequency ramps are transmitted with a duration of ¼ ms and a bandwidth of $\approx$190 MHz in a window of constant transmit beam direction. The number is specified dynamically so that the angle ranges that are of great interest can be better resolved. Thus, in the 70 ms measuring cycle, for example, three beams can be operated with 64 bursts and four with 16 bursts. The uniqueness range of the relative velocity of an individual measurement is around a chirp rate of 4 kHz at approximately ±12 m/s, which is why further measures are used to establish the uniqueness.

The transmit antenna consists, as Fig. 49 shows, of eight patch rows each with six patches. The eight rows are fed by an eight-channel passive Butler matrix phase network, that is, with eight inputs and eight outputs. On switching the transmit path over to one of the inputs, a total of eight transmit beams of different direction can be generated.

The receive antenna consists of two patch rows each with six patches. The received signals of the two patch rows are evaluated separately by two parallel receive channels. In signal processing, the phase difference between the two channels is determined for each target echo and the target angle is estimated from the phase difference. The received signal is downmixed, filtered, and A/day converted. All other processing steps including the alert function take place in a DSP.

To implement the target function referred to above, the approximately 28 mm thick sensor (Fig. 51; for construction, see Fig. 52) is installed as a pair on the right- and left-hand rear corner of the vehicle, respectively, behind the bumper. Table 6 summarizes the characteristics.

10 Summary and Outlook

RADAR technology has significantly advanced the development of driver assistance systems. RADAR technology itself has also been influenced by its use in cars. Nowadays, experiences made in the volume production of complex RADAR

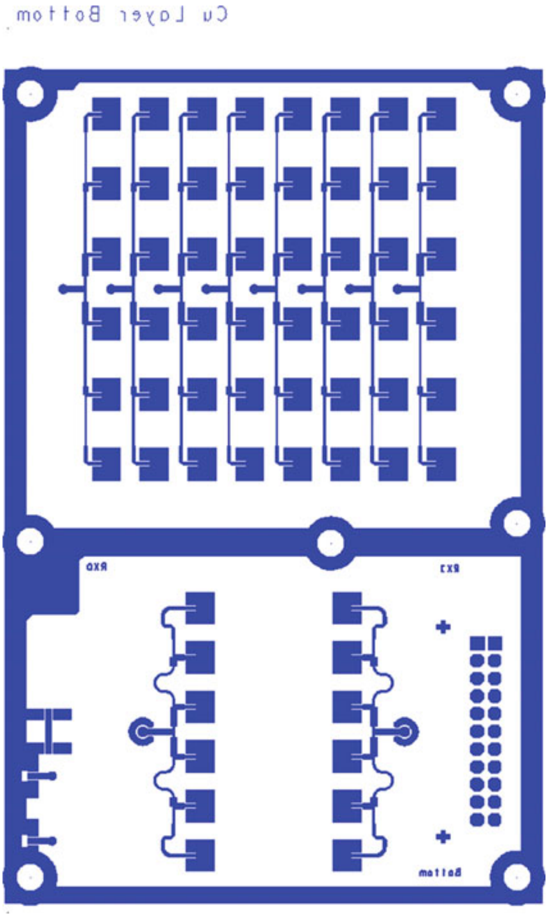


Fig. 49 Antenna of the Valeo multibeam high-performance RADAR (Source: Valeo)

devices are supporting current developments with experience in the field. Whereas the first-generation RADAR sensors by various suppliers were marked by a high level of diversity, initial convergences are now visible in respect to modulation and frequency generation. Nevertheless, it will still be a long time until a status similar to that of ABS is reached, where it is barely possible to distinguish between the devices of different manufacturers.

The constant “threat” of RADAR technology in the car, the LIDAR technology (see ► Chap. 18, “Automotive LIDAR”), has also made huge progress in recent years. But complementary performance differences continue to exist: RADAR benefits particularly from the ability to measure the Doppler effect and from its greater robustness to weather conditions. On the other hand, RADAR permits only low solid angle resolution with an acceptable antenna size. While the terahertz technology promises to improve this situation, technologically it is still in its infancy at the present time. Lobe widths of 1° could be achieved with half the

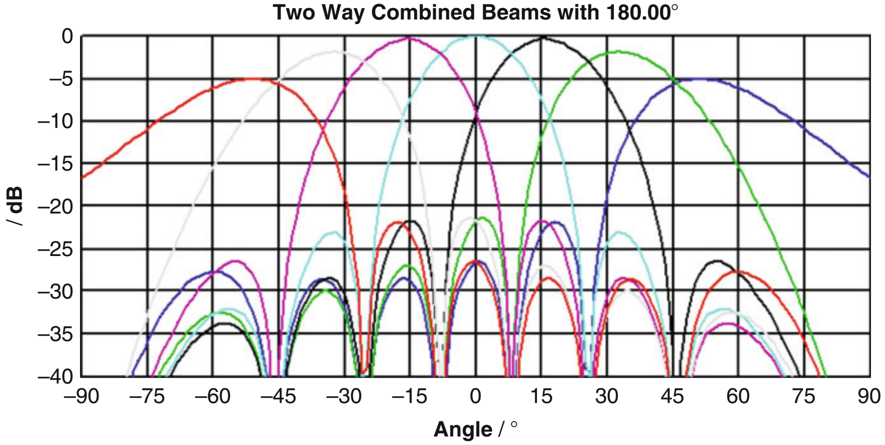


Fig. 50 Phase diagram of antenna (calculated) (Source: Valeo)

Fig. 51 External view of the Valeo MBH RADAR (Source: Valeo)

wave length, and these lobes could determine the borders of objects with a thoroughly satisfactory level of quality.

Perhaps technology will also take a different route if the one-sensor concept is abandoned and approaches are made towards diverse multisensor technology with sensor data fusion. Here, the combination of RADAR plus camera could become a “dream team” that leaves hardly anything to be desired that could be fulfilled by a LIDAR, for example. In a combination solution, the requirements on the RADAR could change and possibly lead to “disarmament” of the RADAR with overall cost optimization.

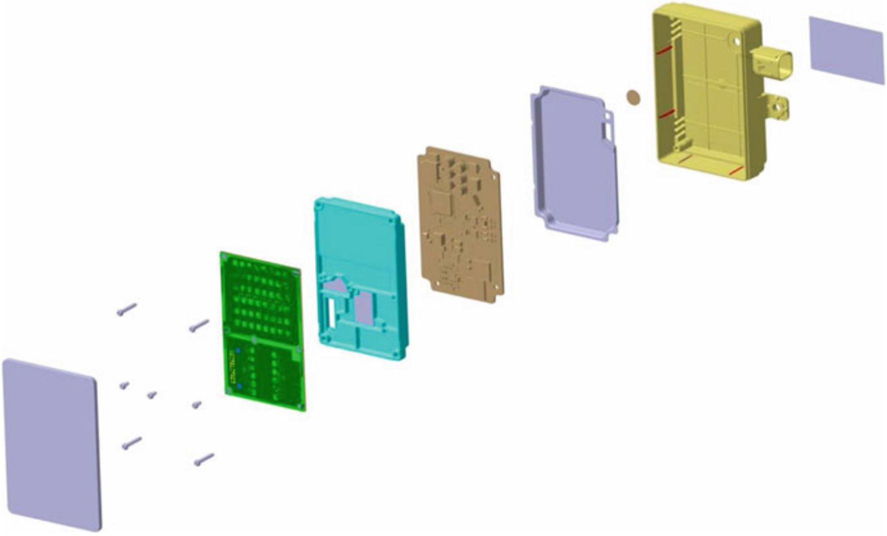


Fig. 52 Exploded view of the Valeo MBH sensor (Source: Valeo)

Table 6 Technical data of the Valeo multibeam high-performance RADAR

Valeo	MBH
General characteristics	
Dimensions (W × H × D)	$70 \times 102 \times 28.1$ mm^3
Mass	240 g
Power consumption	Typically: 2.5 W at 13 V
Cycle time	70 ms
Measuring duration in cycle	39 ms
Signal characteristics	
Frequency range	24.05–24.25 GHz (ISM)
Transmitting power	<= 20 dBm EIRP
Modulation process	FMCW
Modulation deviation	190 MHz
Burst number and duration	16 to 64, 250 μs
Number of measurement ranges	1
Type of angle measurement	Monopulse, seven transmit lobes each 30°
Detection characteristics	
Distance range $r_{\min} \ldots r_{\max}$	>70 m
Distance cell Δr	0.8 m
Relative velocity range $\dot{r}_{\min} \ldots \dot{r}_{\max}$	95 m/s
Relative velocity cell $\Delta \dot{r}$	1.6…0.4 m/s
Measuring range azimuth $\Delta\phi_{\max}$	150°
Accuracy of angle	1°…2° (depending on angle)
Lobe width elevation	±10° (10 dB)

Acknowledgment The author would like to thank Dr. Hermann Buddendick of Bosch, Dr. Markus Wintermantel of Continental, Martin Mühlenberg of Hella, Dr. Alois Seewald of TRW, and Dr. Urs Lübbert of Valeo for providing their expertise and for the technical diagrams.

References

Classen T, Wilhelm U, Kornhaas R, Klar M, Lucas B (2012) Systemarchitektur für eine 360 Grad Fahrerassistenzsensorik (System architecture for a 360 degrees driver assistance sensor system), Tagungsband 8. UniDAS Workshop Fahrerassistenzsysteme

Diewald F (2013) (Object classification and free space detection on the basis of imaging radar sensors for the perception of the surroundings), Dissertation Ulm University Ulm

Diewald F, Klappstein J, Sarholz F, Dickmann J, Dietmayer K (2011) Radar-interference-based bridge identification for collision avoidance systems. In: Proceedings IEEE intelligent vehicles conference 2011, Baden-Baden

ETSI EN 302 858–1 V1.3.1 (2013) Electromagnetic compatibility and Radio spectrum Matters (ERM); Road Transport and Traffic Telematics (RTTT); Automotive radar equipment operating in the 24,05 GHz up to 24,25 GHz or 24,50 GHz frequency range; Part 1: Technical characteristics and test methods, http://www.etsi.org/deliver/etsi_en/302800_302899/30285801/01.03.01_60/en_30285801v010301p.pdf (2013–11)

Fonck KH (1955) Radar bremst bei Gefahr (Radar brakes when in danger). Auto Motor Sport Heft 22:30

German Federal Network Agency (2005) SSB LA 144 – Schnittstellenbeschreibung für Kraftfahrzeug-Kurzstreckenradare (Short Range Radar, SRR) (Interface description for short range radars on motor vehicles), July 2005

Heidenreich P (2012) Antenna array processing: autocalibration and fast high-resolution methods for automotive radar, Dissertation Technische Universität Darmstadt

Hildebrandt J, Kunert M, Lucas B, Classen T (2013) Sensor setups for future driver assistance and automated driving. In: Proceedings IWPC Frankfurt, Germany

Koelen C (2012) Multiple target identification and azimuth angle resolution based on an automotive radar, Dissertation TU Hamburg-Harburg, Shaker Verlag

Kühnke L (2003) 2nd generation radar based ACC – a system overview, Workshop on environmental sensor systems for automotive applications, European Microwave Week, Munich, Oct 2003

Kühnle G, Mayer H, Olbrich H, Swoboda HC (2002) Low-cost long-range-radar für zukünftige Fahrerassistenzsysteme. (Low cost long range radar for future driver assistance systems) Aachener Kolloquium Fahrzeug- und Motorentechnik 2002, p 561

Lucas B, Held R, Duba GP, Maurer M, Klar M, Freundt D (2008) Frontsensorsystem mit Doppel Long Range Radar, (Front sensor system with dual long range radar) 5. Workshop Fahrerassistenzsysteme (Workshop Driver Assistance Systems), Walting

Ludloff A (2008) Praxiswissen Radar und Radarsignalverarbeitung (Practise knowledge radar and radar signal processing), 4th edn. Vieweg + Teubner Verlag, Wiesbaden

Massen J, Möller U (2012) Final report of BMBF project: "RoCC" radar-on-chip for cars – Teilvorhaben Continental 79 GHz SiGe Nahbereichsradarsensorik (– Part Continental 79 GHz SiGe short range radar sensors), Förderkennzeichen 13N9824 (Reference number)

Meinecke MM, Rohling H (2000) Combination of LFMCW and FSK modulation principles for automotive radar systems. German Radar Symposium GRS2000, Berlin

MOSARIM (2013) More safety for all by radar interference mitigation, EU-Projekt 248231, 2010–2012. Accessed Aug 2013

Schneider R (1998) Modellierung der Wellenausbreitung für ein bildgebendes Radar (Modelling of wave propagation for an imaging radar), Dissertation Universität Karlsruhe

Schubert E, Kunert M, Menzel W, Fortuny-Guasch J, Chareau JM (2013) Human RCS measurements and dummy requirements for the assessment of radar based active pedestrian safety systems, International Radar Symposium IRS

Skolnik M (2008) Radar handbook. Introduction to radar systems, 3rd edn. McGraw-Hill, New York

Stoica P, Moses RL (2005) Spectral analysis of signals. Prentice Hall, Upper Saddle River

TC204/WG14, ISO (2008) ISO 22179 Intelligent transport systems – full speed range adaptive cruise control (FSRA) systems – performance requirements and test procedures

TC208/WG14, ISO (2002) ISO 15622 (transport information and control systems – adaptive cruise control systems – performance requirements and test procedures)

Wintermantel M (2010) Radarsystem mit Elevationsmessfähigkeit (Radar system with elevation measurement capability), patent application WO2010/000254 A2

Wolff C (2014) Radargrundlagen – Winkelreflektor (Radar fundamentals – Angle reflector), http://www.radartutorial.eu/bauteile/bt47.de.html. Accessed Nov 2014

Automotive LIDAR

18

Heinrich Gotzig and Georg Geduld

Contents

H. Gotzig (✉)
Valeo Schalter und Sensoren GmbH, Driving Assistance Product Group, Bietigheim-Bissingen, Germany
e-mail: heinrich.gotzig@valeo.com

G. Geduld
Eichberg, Switzerland
e-mail: georg@geduld.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_18

Abstract

Light Detection And Ranging (LIDAR) is an optical measurement principle to localize and measure the distance of objects in space. Basically it is similar to a RADAR-system, but instead of using microwaves LIDAR uses ultraviolett, infrared or beams within the visible light spectrum. Besides distance measurements, which is the basic task, LIDAR sensors can be used for a limited visual detection of objects by analyzing the light intensity, visibility measurement by analyzing the shape of the reflected LIDAR pulse, day/night detection as background illumination is significantly different between day and night, pollution detection and speed estimation. As several research vehicles for autonomous driving vehicles like e.g. Google car use LIDAR as basis sensor technology for scanning the environment there is an increase in development activities for LIDAR sensors meeting automotive requirements (cost, performance, reliability).

Parking Assistance are systems which support the driver during parking manoeuver. This is achieved either by providing distance information to relevant obstacles, camera images or additionally by steering assistance. For the different system configurations requirements regarding sensors, signal processing, HMI, interfaces to the vehicle network need to be met. While first parking systems only provided informations to the driver, recent systems provide lateral steering support and current plus next gerneration parking systems will take more and more both lateral and longitudinal controll of the vehicle during the parking manoeuver. In the near future Valet Parking systems will be available, where the system searches for a suitable parking slot and performs the entire parking process automatically.

1 Function, Basic Principles

1.1 Terminology

LIDAR: Light Detection and Ranging is an optical measurement principle to localize and measure the distance of objects in space. Basically, it is similar to a RADAR system, but instead of using microwaves, LIDAR uses ultraviolet, infrared, or beams within the visible light spectrum (see Fig. 1).

1.2 Measurement Method Distance Sensor

There are several distance measurement methods when using infrared sensors. Most common in automotive is “time-of-flight distance measurement.” The time elapsed between transmitting and receiving of the light (laser) pulse is directly proportional between the measurement system and detected object.

With “time-of-flight measurement,” one or several light pulses are transmitted; they are reflected by a possible existing object. The elapsed time until reflected

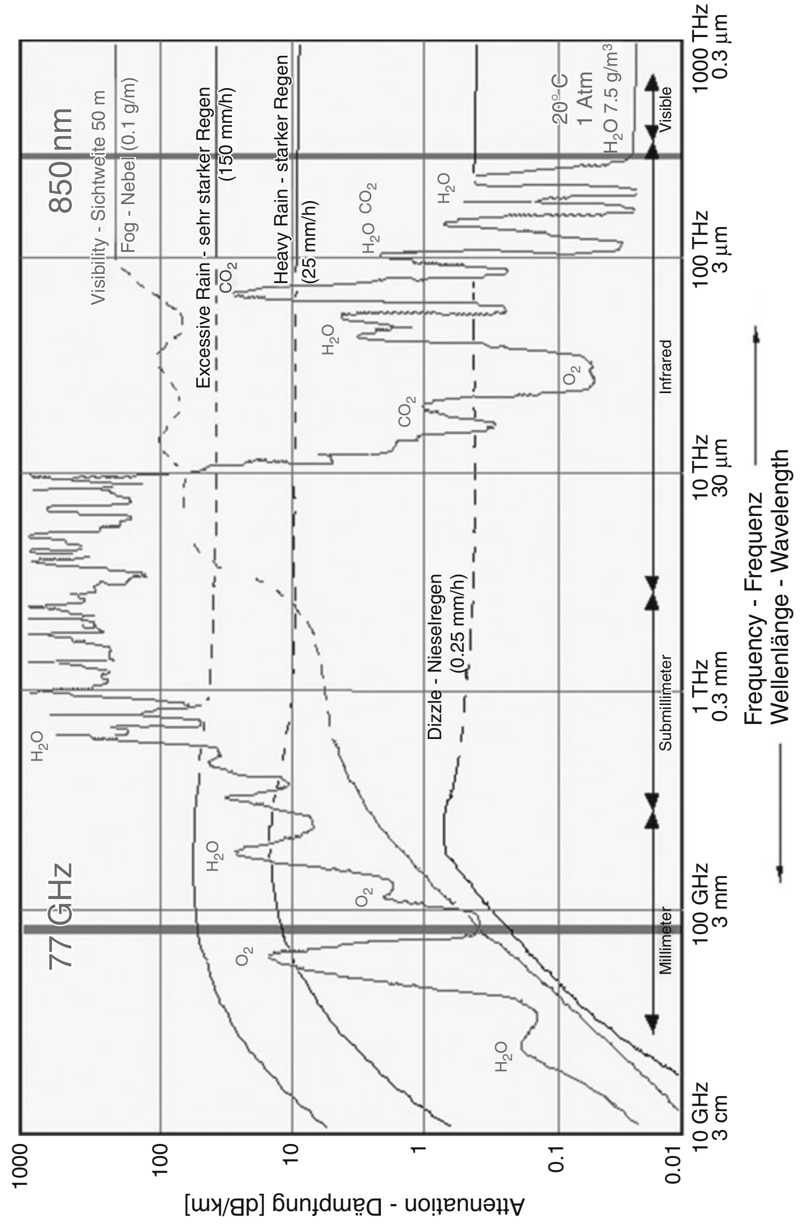


Fig. 1 Frequency spectrum

signal is received is proportional to the distance. With a speed of light of 300,000 km/s (in air), the time to be measured with an object at a distance of 50 m is approx. 3×10^{-7} s or 333 ns. This is a typical driving situation where "speed = 100 km/h and distance = half tachometer" (see Fig. 2):

$$d = \frac{c_0 \cdot t}{2} \tag{1}$$

d = distance in m
c_0 = speed of light (300,000 m/s)
t = time in s

A reflected pulse of a fixed and single object (e.g., vehicle) has the shape of a Gaussian curve.

Since the emitted light pulse travels twice the distance between the sensor and obstacle, the elapsed time between sending and receiving represents twice the distance to the object (see Figs. 2 and 3).

If there are several objects in the detection area of the sensor and therefore in the measurement channel, they can be recorded with an appropriate evaluation if the distance between the obstacles is sufficiently large enough. This is called multi-target capability of the system (see Fig. 4).

If there is an increased attenuation of the atmosphere due to fog, rain, etc., then individual pulses are reflected on the water droplets in the air (see Fig. 5). Depending on the optical design of the system, this can lead to saturation behavior in the receiver. A measurement is no longer possible.

But state-of-the-art sensors work with a dynamic adjustment of the sensitivity, and together with the multi-target capability of the measurement channel, obstacles within or behind "soft" atmospheric disturbances can be measured. With "soft" objects, like fog, many signals are recorded from different distances. This leads to a merge of the single pulse responses, and the total signal response is a flat and extended echo (see Fig. 6). Thus, both the elapsed time of the received pulse response (a single Gaussian curve in the simplest case) and the shape of the echo are used for a more detailed analysis on the kind of the detected echoes and the environmental conditions.

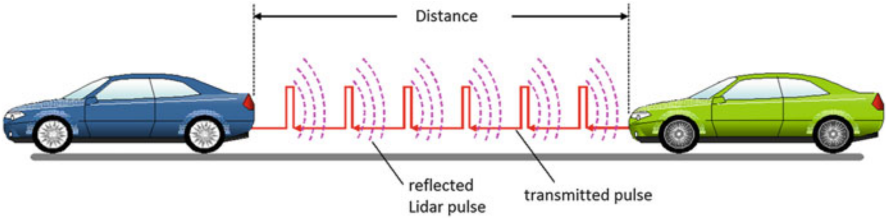


Fig. 2 Travel time measurement

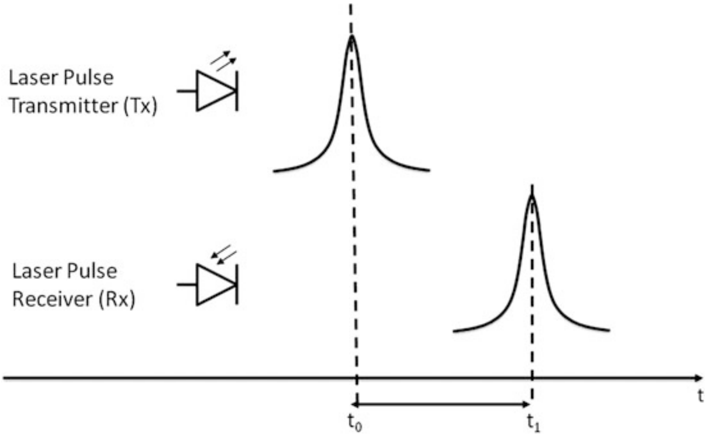


Fig. 3 Impulse response of an object

Fig. 4 Impulse response of two objects

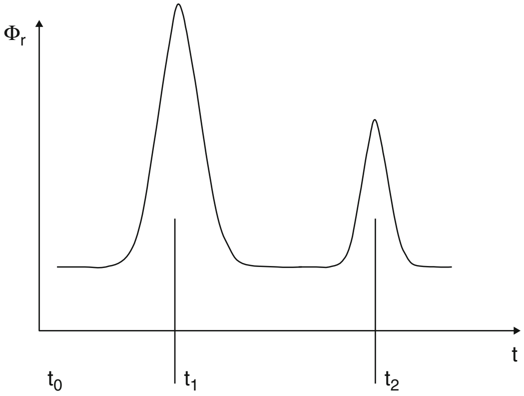
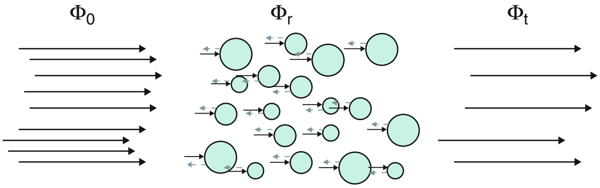
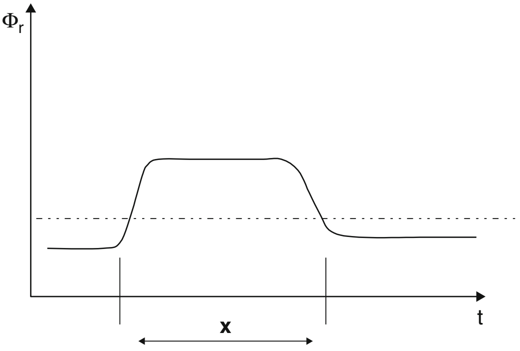


Fig. 5 "Soft" object (fog)



Thus, for example, the signal of mist or spray differs from the one of a vehicle (see Fig. 7). The shape of this signal contains information about the absorption degree of atmospheric disturbance. From the measurement of "signal length" x and analysis of the temporal decrease of $\Phi_r(t)$ (see Fig. 6), the prevailing visibility can be estimated.

Fig. 6 Signal response “soft” object (fog)



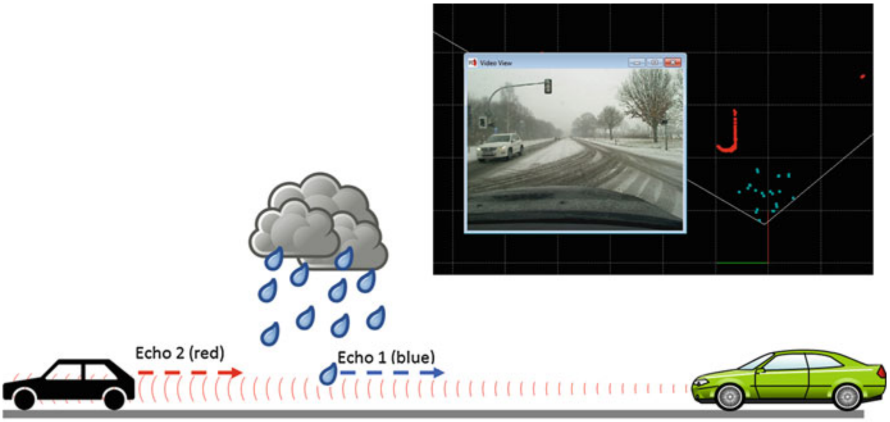


Fig. 7 Distinction rain – vehicle

The range performance is significantly influenced by the intensity of the emitted light pulse and the receiver sensitivity. The pulse power is limited by the eye safety requirements. Other parameters, such as transmission of the atmosphere and the size or reflectance of the object, however, are not influenced.

The received light intensity is described by the following equation (only valid if size of beam is smaller than size of object) (Eq. 2):

$$P_r = \frac{KK \cdot A_t \cdot H \cdot T^2 \cdot P_t}{\pi^2 \cdot R^3 \cdot (Q_v/4) \cdot (\Phi/2)^2} \tag{2}$$

In the case that the target (at a greater distance) is smaller than the beam surface, the following equation is valid:

$$P_r = \frac{KK \cdot A_t \cdot H \cdot T^2 \cdot P_t}{\pi^2 \cdot R^4 \cdot (Q_v \cdot Q_h/4) \cdot (\Phi/2)^2} \tag{3}$$

with:

P_r = intensity of the received signal (W)
KK: reflectivity of the measured object
Φ: angle of reflection object (rad)
H: object width (m)
A_r: target size (m^2)
T: transmission of the atmosphere
Q_v: vertical beam divergence (rad)
Q_h: horizontal beam divergence (rad)
A_t: receiving lens surface (m^2)
P_t: laser power (W)

1.3 Additional Functionality

Basically, LIDAR sensors can also be used for a limited visual detection of objects in addition to pure distance measurement. Here, the light intensity is then additionally evaluated accordingly. But due to the physical and technical principle, the performance is worse compared to a camera, as they have a higher resolution and are capable to detect a wide frequency. Furthermore, the performance depends largely on the contrast between objects to be detected and the environment (e.g., lines on roads).

1.4 Construction

In principle today's LIDAR distance measuring devices are constructed according to the same construction. There are differences in the way how to generate multiple measurement channels (beams) and the implementation of a beam deflection at "sweeping" (depending if, e.g., the curve radius is tracked) and scanning methods.

1.4.1 Transmit Path

LIDAR uses for active distance measurement a laser source, which typically emits in the range between 850 nm and 1 μm. In order to ensure an optimal target separation of multiple echoes, the measurement pulse should be kept as short as possible. During the design of LIDAR sensor, eye safety has to be guaranteed, since the integral of the pulse is emitted energy. The radiation peak power of the high-power diodes used may reach 75 W (OSRAM 2015; Laser Components 2015a) or even more. The pulse length is typically in a range of about 4–30 ns, with an active

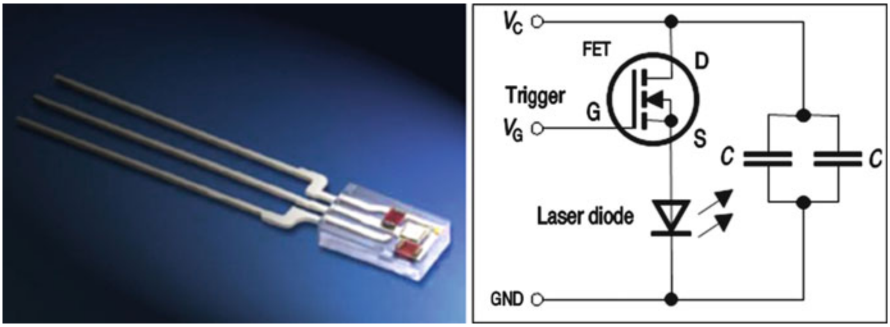


Fig. 8 OSRAM SPL LL90 – driver stage in the housing of the laser diode (OSRAM 2015)

surface on the semiconductor laser of about 200 × 10 μm^2. This corresponds to a laser power at the surface of about 35 GW/m^2.

In order to achieve the high electrical radiated power interference-free, the driver stage of the laser (see Fig. 8) has to be as close as possible and thus directly built to the housing of the semiconductor laser. Other challenges include the temperature of the boundary layers to the plastic housing and the power supply (operating voltage 12 V) to the whole component.

1.4.2 Receiving Path

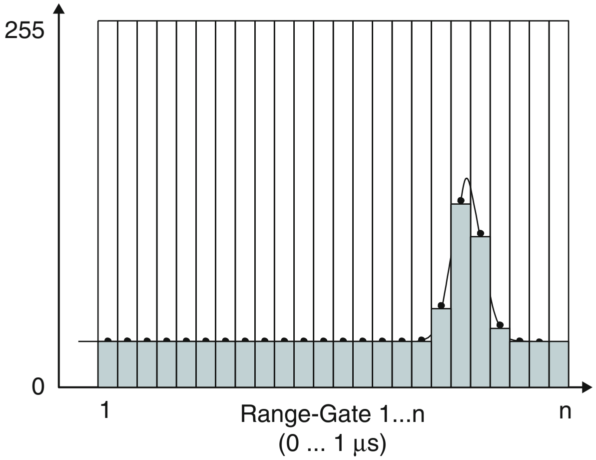
The sensitivity of the receiver decisively impacts on the achievable performance of the sensor. Basically, the sensitivity of a sensor component can be reached via the size of the receiver surface. Limitation of this is the aperture and the quality of the optics. To achieve the required accuracy in the centimeter range, high measurement speed is required. In a measuring range from about 10 cm to about 150 m, the light travel time for different LIDAR systems is in a range from 0.1 ns to 1.0 μs. Another challenge is the "glare" by the ambient light. During the day several orders of magnitude causing the spread of the sunlight spectrum, which also includes a significant share in the infrared range, more light output than the LIDAR distance sensor. By suitable filter measures, the light component (caused by solar radiation) is suppressed. These measures are carried out mainly in hardware.

When receiving positive intrinsic negative (PIN) diodes or avalanche photodiodes (APD) used (Laser Components 2015b).

Avalanche diodes are used as a photodiode semiconductor detector for counting single photons. Typically they are operated with a large series resistance in the reverse direction. Due to the high field strength, a single photon is able to release an electron, which is accelerated by the field in the barrier layer, and triggers an avalanche effect. The resistance prevents that the diode remains broken through (passive quenching). The diode returns back to the locked state. This process is repeated periodically, and their measuring frequencies up to 100 MHz are possible.

The PIN diode is mainly used in optoelectronics for optical communications technology for optical waveguides such as photodiodes. PIN diodes are due to the

Fig. 9 Digitizing by means of parallel gating



thick i-layer (lightly doped, conductive-intrinsic conductivity) of thermally more stable and cheaper, but less sensitive, since in here more charge carriers can be stored. Peak values for sensitivity are between –40 and –55 dBm at a wavelength of 850 nm.

To improve performance, the APD is often implemented as an ASIC.

In order to achieve the spatial resolution of a few centimeters, the so-called "parallel-gating" process of the amplification of the signal among other things is applied. Here, the received signal is digitized via a time-controlled multiplexer and stored in individual "memory cells" (range gates) (see Fig. 9). Each memory cell corresponds to a "range gate" of, e.g., 1.5 m. The addition of multiple transmit pulses results in a Gaussian distribution of the range gates around the actual measurement point.

To increase the accuracy of time measurement in the processor unit of the sensor, the expected received pulse is reconstructed, and its apex is determined. Thus, from a still relatively coarse time (distance) and amplitude resolution, range resolution in the centimeter range can be achieved. A further increase in accuracy is possible by the temporal analysis of the calculated distances.

With the number of transmission pulses per measurement, the sensitivity of the sensor system can be increased or be controlled. With the described method, a dynamic range of more than 50 dB with respect of the optoelectronic current can be achieved. This is necessary to detect bad reflective objects in the target distance.

Figure 10 shows the basic structure of today's typical LIDAR distance sensor. Depending on the required numbers of individual sub-elements are integrated into an ASIC.

In addition to the hardware structure of a distance sensor, individual functions are realized by software. As mentioned above, a large part of the signal evaluation such as the determination of distance and relative speed detection is done by the software. Also on the block "signal processing," information on the visibility and

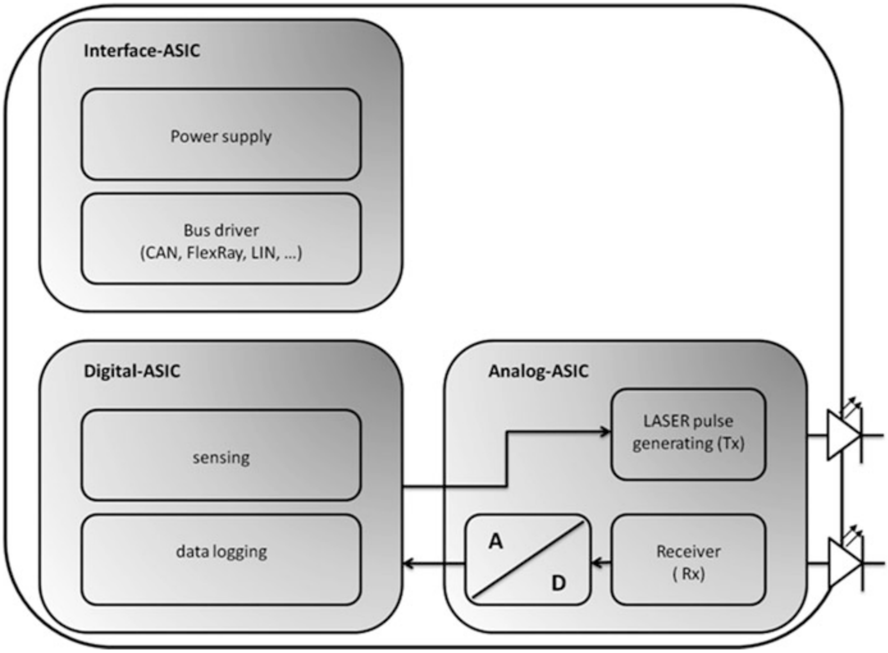


Fig. 10 Basic structure of a LIDAR distance sensor

system limitations are calculated. The single arrows represent the exchange of data between different software blocks (see Fig. 11).

The laser driver provides the timing of measurement and receiving channel.

1.5 Transmission and Reflection Properties

As with all the active and passive measurement methods, the transmission or the attenuation of the atmosphere plays an important role in the system design and the performance of the sensor system that can be achieved. While with passive methods, such as cameras, the distance from the object to the sensor has to be put back only once, with active technology, the distance from the object to the sensor has to be put back twice.

The transmission properties of the atmosphere are significantly affected by their constituents and their physical conditions (see Fig. 1).

Figure 12 shows a simplification of the test section in the atmosphere (l). It should be noted here that this represents only one way of the distance sensor relative to the object. The light must travel the same way again in the reverse direction back after reflection on the object (see Fig. 13).

The transmitter radiates the luminous power Φ_0. Through the atmosphere (containing water droplets, dust particles, etc.), parts of the light are reflected diffusely (Φ_r). In addition a part of the total energy is absorbed (converted into heat) until at the end of the track only a reduced optical power is available Φ_t:

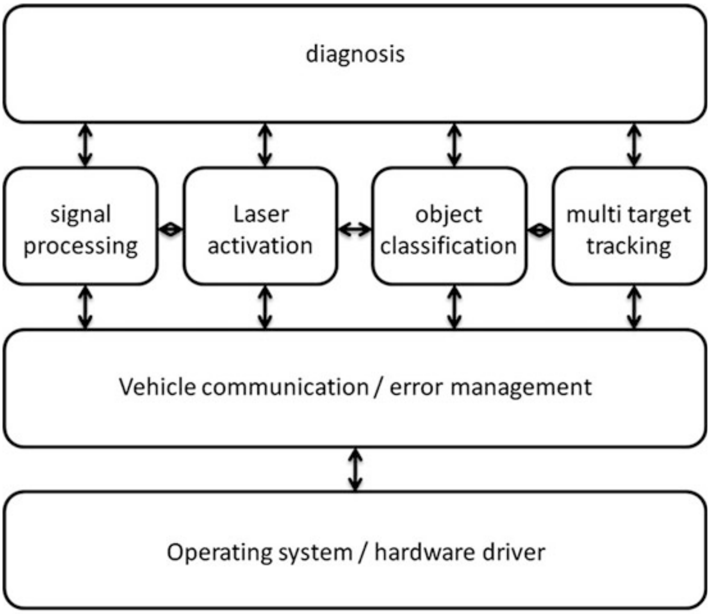
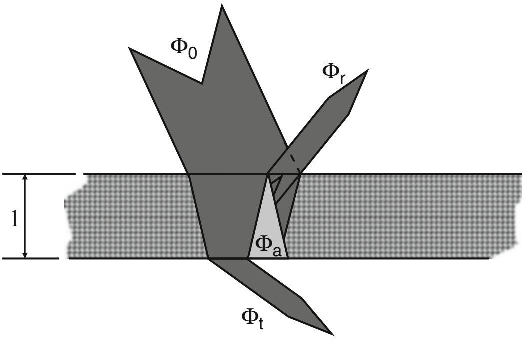


Fig. 11 Basic software functions

Fig. 12 Absorption, reflection, and transmission



$$\Phi_0 = \Phi_r + \Phi_a + \Phi_t \tag{4}$$

$$\text{Degree of reflection} \cdot \varrho_r = \Phi_r / \Phi_0 \tag{5}$$

$$\text{Degree of absorption} \cdot \varrho_a = \Phi_a / \Phi_0 \tag{6}$$

$$\text{Degree of transmission} \cdot \tau = \Phi_t / \Phi_0 \tag{7}$$

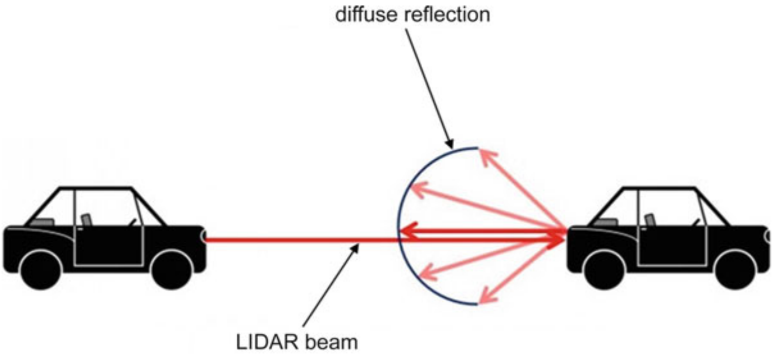


Fig. 13 Lambert reflector

where

Φ_0 – emitted luminous power
Φ_r – reflected luminous power
Φ_a – absorbed luminous power
Φ_t – received luminous power
l – path through the atmosphere

The proportion of the transmitted radiation is referred to as transmittance (τ) (see Eq. 7). The attenuation is generally composed of absorption, scattering, diffraction, and reflection and is wavelength dependent (see Eq. 4).

A major challenge in laser measurement is the requirement for eye safety which results in a very limited receiving energy after reflection from an object. It should be noted that usually the object (vehicle) is similar to a Lambertian reflector which radiates its energy diffusely in the half solid angle (180°) (see Fig. 12).

With the Lambert reflector the backscattered energy is not focused, but is distributed inhomogeneously in the solid angle (within a "ball"). Therefore, only the part of the backscattered energy that is radiated back directly into the sensor's receiver can be used for detection. This is, in practice, at best, 20 % (usually much less) of the reflected energy to the object.

As mentioned, the average transmission power is limited, but as an improvement the beam can be bundled to increase the energy density or use a high gain receiver. The bundling has the disadvantage that at too small solid angles, the beam may, at a homogeneous surface on the vehicle (e.g., only the bumper) and thereby of the entire beam, be reflected away by total internal reflection.

Total reflection (see Fig. 14) occurs when narrow beams (see also "Lambert reflector" – Fig. 13) are used, which impinges on a sloping surface. This can be remedied by widened beam or several beams. The ideal situation is

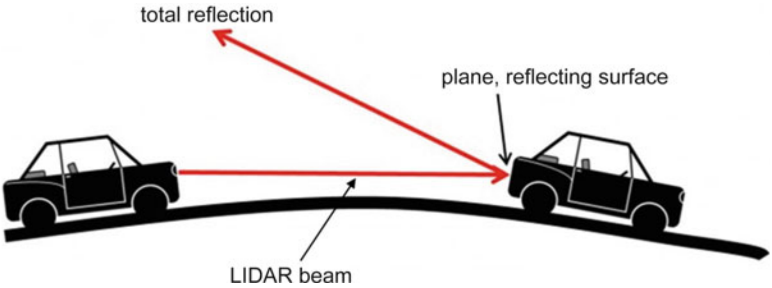


Fig. 14 Total reflection

to shed light on edges or directed perpendicular to the transmitter parts in the detection area.

These measures are sometimes counterproductive (see energy density problem); the problem of multiple receive beams is compensated by the use of scanning systems with multiple transmit/receive channels (several hundred) but sometimes also leads to higher costs.

1.6 Speed Motion Detection

Driver assistance systems require information about the own speed of the vehicle, the relative speed to objects, and the motion of objects in the relevant environment. The determination of the own speed (value and direction) is generally carried out by evaluating steering angle sensor and wheel rotation sensor. In principle, the Doppler effect to determine the relative speed of the detected objects is used by the LIDAR system. However, the increased demands and the associated costs in measuring the Doppler frequency in the spectrum of light prevent its implementation.

For this reason, one uses the differentiation of two one following ideally several successive distance measurements (see Eqs. 8 and 9).

$$\vec{v}_{rel} = \frac{\mathrm{d}\vec{R}}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta \vec{R}}{\Delta t} \tag{8}$$

A prerequisite is that the distance information is definite, i.e., the same object/reflection point. Depending on the type of LIDAR, distance information R is either a purely radial distance value, or it contains additionally direction information. The horizontal angular resolution for scanning systems is typically in the range of $\leq 0.5°$. Neglecting vertical information, the relative velocity is given by

$$\vec{v}_{rel} = \frac{\vec{R}_2 - \vec{R}_1}{t_2 - t_1} \tag{9}$$

$\vec{v}_{rel}$: relative velocity in m/s
R: distance in m
t: time in s

This method is possible only when a very accurate distance measurement can be guaranteed. Accuracy increases can be achieved through suitable filter such as state observer or Kalman filter.

In order to achieve a better prediction of how the environment situation develops in the future, change in the movement of the relevant objects must be known. By analyzing the change of speed, the relative acceleration can be determined (see Eq. 10):

$$\vec{a}_{rel} = \frac{d^2 \vec{R}}{dt^2} \tag{10}$$

$\vec{a}_{rel}$: relative acceleration in m/s^2

As a possible error in distance measurement, inaccuracy in calculation of relative acceleration increases. Thus, for a possible control task, the signal must be filtered accordingly.

1.7 Tracking Method and Selection of Relevant Objectives

The term tracking includes all processing steps that serve the tracking of objects. The aim of this tracking is, on the one hand, the extraction of information about the history of the movement and the position of an object and, on the other hand, the reduction of negative influences, mostly originating from random measurement errors (measurement noise). The accuracy of the determined position and motion information depends on the used tracking algorithm as well as on the accuracy of the measurements, the measurement error, and the sampling rate of the cyclic measurements.

In general, the target selection can be performed in two different ways. Either one covers the entire area and then selects a relevant target (see ▶ Chap. 45, “Adaptive Cruise Control”) with lane assignment and/or further selection features, or one is limiting the detecting of objects from the beginning on to the relevant range of expected travel trajectory. Both methods have advantages and disadvantages. Table 1 shows a comparison (Fig. 15).

The performance of the ACC sensor is determined not only by the sensitivity primarily by the quality of the determination of the relevant object. This requires a powerful calculation of the trajectory of the driving path (see ▶ Chap. 45,

Table 1 "Tracking" methods

	Option 1	Option 2
Figure	Figure 15a	Figure 15b
Description	Detection of objects in the entire detection area	Detection of objects in the entire detection area
	Discrimination of targets based on the determined driving path	Information about distance and direction of the measurement only in the relevant field
Advantage	Detection of all objects	Low computing power
Disadvantage	Computing and memory requirements also for nonrelevant objects	

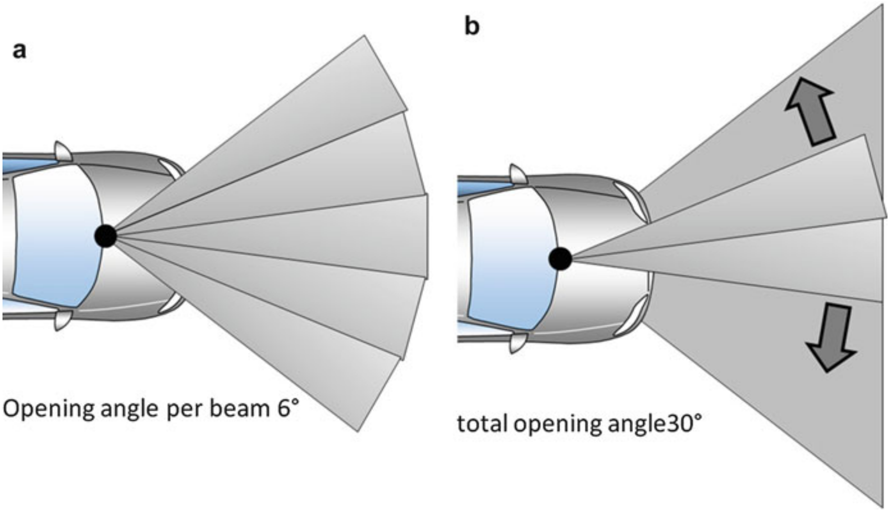


Fig. 15 Comparison of different tracking methods. (**a**) Multibeam rigid (number or value of the opening angle can vary). (**b**) Multibeam sweep (total open angle, number, and value of the single opening angle can vary)

"Adaptive Cruise Control"). The tracking can basically be divided into the following processing steps:

1.7.1 Prediction (Extrapolation/Estimation)

In this processing step, the (computational) prediction of the position and movement information on a basis of the known history as a function of physical properties of the relevant object (dynamic) and assumptions on how the objects will behave in the future. Of basic interest as well are other active road users (cars, people, etc.) and static objects (stationary vehicles, lane boundaries, etc.). For the prediction, it is important that the position and movement information with sufficient accuracy depending on the environment (highway, urban environment, etc.) and own vehicle velocity is precise enough to realize a corresponding application (e.g., avoidance assistant).

1.7.2 Association (Linkage of Objects)

In particular, when there are multiple objects observed in the detection area (multi-target tracking) and the system is able to clearly differentiate themselves through various measuring cycles, the component maintains a mapping of an observed object in previous measurement cycles at a current measured object. Errors in this processing step may have a very negative effect on the results of impact (v_{rel}, a_{rel}, etc.). Scanning systems offer significant improvements which have an angular resolution in the range of a few 0.1°.

1.7.3 Innovation (Linking Real Measurement and Prediction)

The determination of the current position and other relevant information takes place on the one hand by the prediction and the other by current measurements. The innovative step leads both results together, in which the individual results are weighted. This weighting can be done either dynamically or statically. If the prediction is weighted higher, then the result is smoother, while actual measurements yield results that adapt quickly to changes in the measured values. Depending on the function or situation (security system/emergency braking or comfort system/ACC), the filters are then adjusted accordingly. The quality of the models or the degree of approximation to reality decisively determines the result of tracking. Usually for a LIDAR distance sensor, a Kalman filter is used.

1.7.4 The Kalman Filter: Function Principle

The Kalman filter (Wikipedia 2015a; Filos et al. 2015; Jiang et al. 2015) is used to estimate states and parameters of a system due to partly redundant distance and relative velocity measurements, which are superimposed by noise (Fig. 16).

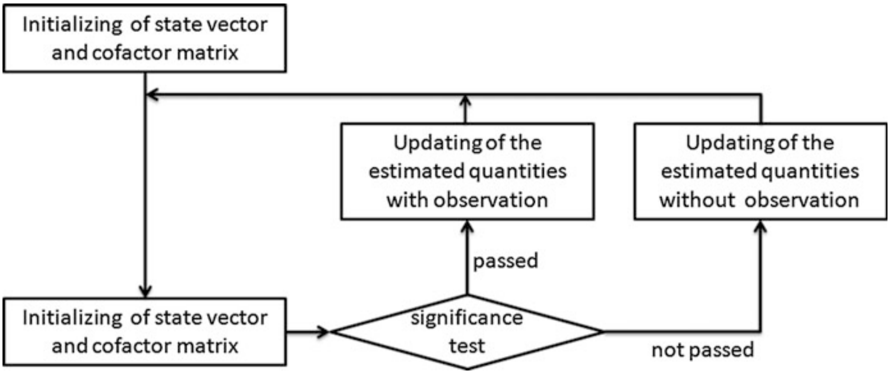


Fig. 16 Kalman filtering principle

Software Kalman filter tracking + classification

The filter has a so-called “predictor-corrector” structure. First it is predicted based on the system input data and the most likely new position and speed and then compared with the actual measurement data. The difference of the two values is weighted and used to correct the current state.

In simple terms, the distance measuring system is described linear and is based on a state space model with equation of state (Berges et al. 2002):

$$x(k) = Ax(k-1) + Bu(k) + Gv(k-1) \tag{11}$$

Prediction observation equation (Luo and Kay 1998):

$$z(k) = Hx(k) + w(k) \tag{12}$$

At a constant speed ($a_{rel} = 0$) can be described by the following state vector:

$$x = \begin{bmatrix} d \\ v_{rel} \\ a_{rel} \end{bmatrix}$$

and the system matrix:

$$A = \begin{bmatrix} 1 & T_k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

When multiple objects are tracked simultaneously, the objects of a measurement step must be assigned to the correct object path. For this a Kalman filter has an association gate added. A common solution is the method of nearest neighbors. Taking into account the uncertainty of the measurements and deviations from the assumption of constant velocity results in a search area of the object in the new measurement. By estimating the new measured value with the Kalman filter, the measurement window can be specified. A track is assigned to the object with the least difference between prediction and measurement. The measured data are then used for the innovative step of the Kalman filter.

Readings outside the measuring range are discarded directly, as well as objects only once found to be excluded as a measurement error of tracking. On the other hand, a new path is initialized when the first time the detected object in the next measurement step, a measured value can be assigned. If no object is assigned to an existing track, the predication is done via additional measuring steps. The track will be finished if prospectively no further measurements can be assigned. Several closely spaced objects with similar relative velocity can be summarized (clustering). Misinterpretation can, however, only exclude retroactive, wherefore two measured values are initially kept separate.

2 Application in Vehicle

2.1 Laser Safety

Basically, it is only allowed to install in a vehicle LIDAR sensors which are certified according to laser protection class 1. Relevant for the determination of the laser class is the ICE 60 825-1, Amendment 2: 2001 (Wikipedia 2015b).

The explanation of details of the laser standard is certainly not the intention of this paper. The basic input is the emitted energy of the sensor and thus the energy balance on the retina of the human eye. As the frequency spectrum used for LIDAR is close to the visible range for humans, the eye acts as a focusing lens with his magnifying glass. However, since the laser light is not visible to humans, natural protection mechanisms of the eye such as closing of the pupil will not work. The energy transferred from the LIDAR laser into the eye heats the retina and leads in the worst case to a combustion of the visual cells (thermal retinal damage).

The calculation of the maximum radiated energy takes into account the ability of the energy to be forwarded into the tissue (heat dissipation). Thus, there are criteria that take into account the average heating, as well as the short term, caused by single pulse heating.

The following technical constraints are criteria that affect the eye safety:

- Wavelength (typical for automotive use is 850 nm to about 1 μm)
- Output pulse peak power (typically between 10 and 75 W)
- Output average power (typically between 2 and 5 mW)
- Duty cycle (pulse/pause ratio)
- Exit surface which can be refocused (fractions of mm^2 are typical when using optical fibers; the laser itself is just a few μm^2)

Details are strongly dependent on the particular design of the sensor. In particular, the duty cycles, output pulse power, and the exit surface which can be refocused are very different for the various established products on the market.

2.2 Integration for Forward-Oriented Sensors (e.g., ACC)

In general the integration into the vehicle with respect to the position is not that difficult. Basically, a LIDAR can be placed anywhere in the front. However, preferred positions are in the horizontal plane between the headlights and vertically between the upper edge of the roof and the bumper (see Fig. 17).

It does not matter whether the sensor is placed outside of or behind the windshield. Due to the detection of pollution, cleaning measures can be initiated, or if necessary the driver is informed. In contrast to a camera, however, “only” energy is transferred and thus a “clear” picture is not important. This reduces the requirements for a clean sensor surface considerably. Depending on the installation and

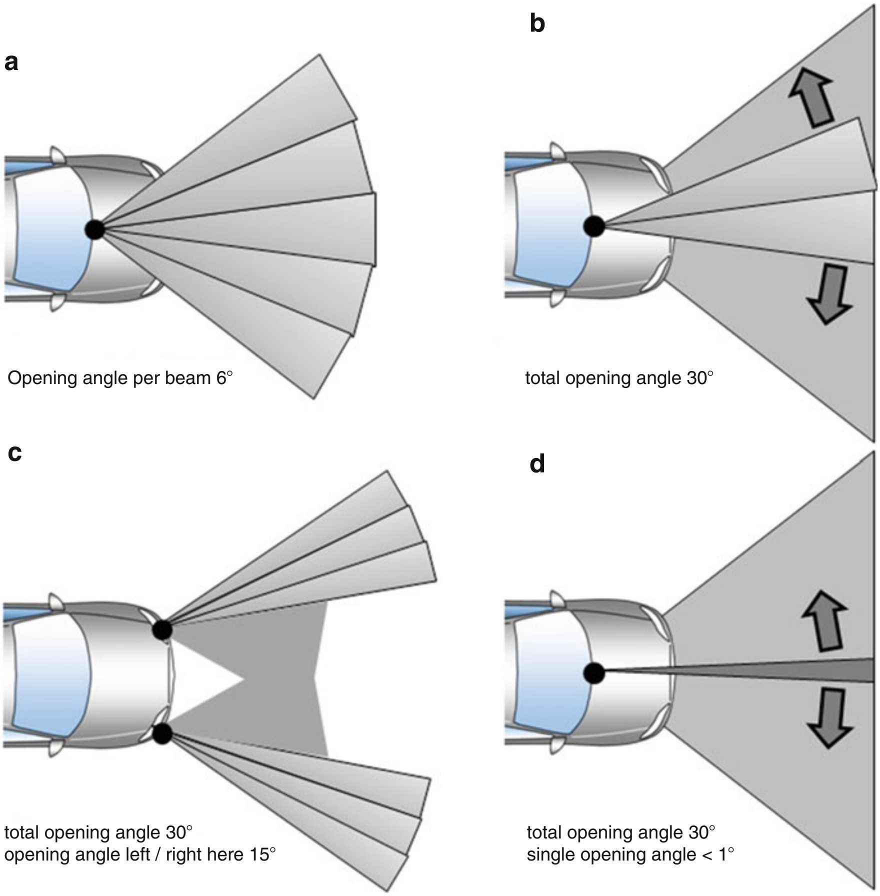


Fig. 17 Examples of different beam sensors: (**a**) multibeam rigid, (**b**) multibeam SWEEP, (**c**) multibeam spread, (**d**) single beam scan

integration considering aerodynamic requirements, a minimal negative impact on the sensor performance is caused by dark impurities (e.g., insects) on the sensor.

3 Additional Functions

With the LIDAR, it is possible to realize further very useful sensor functions.

3.1 Visibility Measurement

So with the abovementioned “soft target” detection, it is relatively easy to use it for computing the visibility by analyzing absorption. It can be used, e.g., for speed recommendation. Due to the wavelength which is close to the visible spectrum of human people, the measured reflection and absorption in the atmosphere is comparable to human visual obstruction.

3.2 Day/Night Detection

The measurable background illumination in the receiver changes significantly between day and night, when the sun gives off infrared beams many times higher than those emitted by the LIDAR.

This signal, adapted accordingly, can also be used as an additional control of driving lights (cf. the day/night or tunnel-dependent control of driving lights).

3.3 Pollution Detection

The basic functions of the self-diagnosis of a distance sensor include the detection of the degree of contamination of the sensor at the transmitter and receiver. This signal results in most of the cases not to a request to clean with the sensor, but the signal can easily be used to trigger an automatic cleaning of the headlight or windshield.

3.4 Speed Estimation

Today's LIDAR sensors have a sophisticated tracking and track up to 20 or more objects on and next to the road. The measurement of the distance and relative speed of objects next to the roadway such as delineator allow the determination of the intrinsic speed of the vehicle by means of distance sensor.

3.5 Driver Behavior/State

If the LIDAR is in passive mode, i.e., no active control system, the information on distance behavior, combined with the steering behavior, allows conclusions on driver condition and may then be communicated to the driver suitable (fatigue, inattention, etc.).

3.6 Object Expansion/Recognition

A scanning LIDAR system with high angular resolution can provide data which can be processed by further mathematical analysis, and as a result partly spatial extension of real obstacles as well as a recognition of the nature of the object can be obtained.

4 Current Series Examples

The sensors shown above satisfy all the required demands of modern ACC, FSRA, or even precrash systems. However, the realization of the optical properties is fundamentally different (Fig. 18, Table 2).

Hella relies on a multibeam principle. This is represented by multiple independent transmit and receive channels. In this case, an array of laser diodes is driven in multiplex procedure. Via the receiving optics, the information about a PIN diode array can be detected. The angular resolution corresponds more or less to the beam width of the individual transmit/receive channels. Up to 16 of these pairings are used to generate the corresponding lateral angle (see Fig. 19) (Höver et al. 2006).

Another method used in practice is the "sweeping" of "beams" as realized by **OMRON** in their second generation. Depending on the course of the road, five independent transmit/receive channels are pivoted laterally via a movable optic. The five transmit/receive channels are modeled by means of optical fibers. Different opening angle can be generated in the lateral and horizontal position depending on the channel. The "line of sight" of the beam light is tracked laterally as a function of the estimated road shape/curve radius. The advantage can be seen in less laser diodes and few moving parts. The disadvantage is that the detection depends on the quality of the estimation of the trajectory.

OMRON's third generation tries to improve the disadvantages of their second generation. The detection range is expanded to 30 × 10°. Instead of "sweep,"

Fig. 18 (**a**) AIS200 – Continental. (**b**) gen2 – Omron. (**c**) gen3 – Omron

Table 2 Examples from series production I

	AIS200 – Continental	gen2 – OMRON	gen3 – OMRON
Wavelength	905 mm	905 mm	905 mm
Eye safety	Class 1 (IEC825)	Class 1 (IEC825)	Class 1 (IEC825)
Radiated power	40 W (peak) 3.5 mW (average)	12 W (peak) 5 mW (average)	12 W (peak) 5 mW (average)
Detection area	±15° (azimuth/hor.)	±11° (azimuth/hor.)	±15° (azimuth/hor.)
	6.5° (elevation/vert.)	3° (elevation/vert.)	10° (elevation/vert.)
	Scan	Sweep	3D scan
Auto-alignment	Vertical		
Number of beams	15 . . . 30	5	1
Min. curve radius	100 m	300 m	100 m
Detection range	1 . . . 180 m	1 . . . 180 m	1 . . . 150 m
Size L × W × H	88 × 72 × 57 mm^3	180 × 89 × 60 mm^3	140 × 68 × 60 mm^3
Accuracy speed	1 kp/h	1 kp/h	1 kp/h
Specifics	Angular resolution 0.01°		

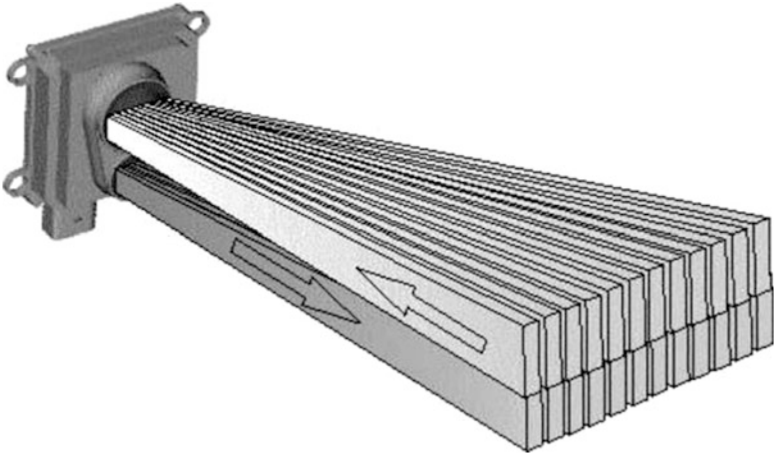


Fig. 19 Hella IDIS® – 12 Kanal laser

“scanning” is used. In doing so the entire lateral detection range is always detected and thus supposedly not captured interesting image sections. A unique feature is also the possibility to detect two other planes in the horizontal direction. This enables to use the sensor also in medium and compact class vehicles which do not have a level control. The mechanism is as robust as easy. Similarly, the swinging shaving head of a shaving apparatus thereby stimulates only the optics of the transmit and receive channels (Arita et al. 2007).

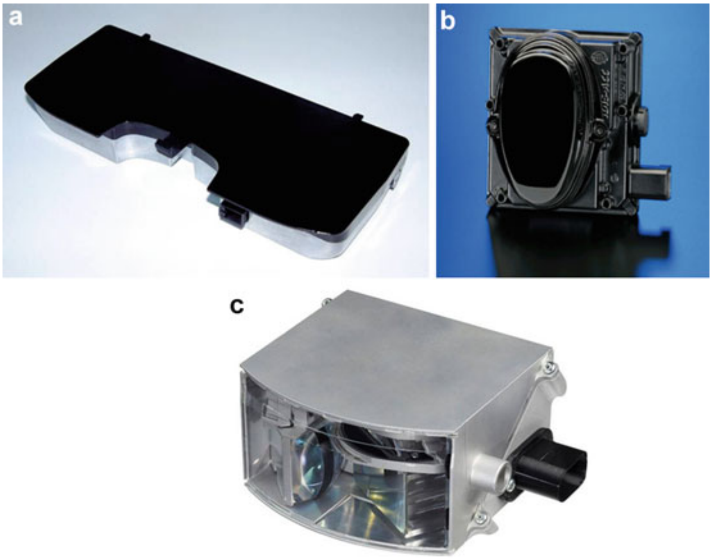


Fig. 20 (**a**) Siemens VDO. (**b**) Hella IDIS®. (**c**) ScaLa Valeo

Continental's latest series laser development, introduced by the takeover of the company **Siemens VDO**, also uses the already mentioned "sweeping" (see gen2 OMRON) of the total beam combination (five beams). Exceptional is the "sweep" field superimposed "Microscan," which makes an exact determination of the vehicle edges possible. Information of cut in of vehicles which can be earlier detected as "relevant" for the ACC system. Mirror optics allow the flat design of the sensor directly, like a rain sensor, mounted on the windshield. There is no unused optical free space, such as view funnel in front of the transmission range, and this can be integrated to save space in the rearview mirror area. This installation is located in the area of the windscreen wiper and is therefore always protected from contamination. In contrast, laser sensors mounted in the exterior have to deal with the strong salinity (winter) or during rainy conditions by the water droplets which causes significant attenuation. Different ranges, depending on weather conditions, are the result. The ACC system may work noticeably different (Mehr 2008).

Currently, a new sensor module is developed, which is provided for a series starting in 2015, "SRL-CAM400." Here, a CMOS camera with a LIDAR is integrated in a compact unit, which is housed in the mirror. The design is scalable provided, depending on vehicle class and performance requirements (Continental 2015) (Fig. 20, Table 3).

In 2010 **Valeo** has started a cooperation with company Ibeo. The aim is to develop Ibeo laser scanner technology (LUX 2010) in such a way so that it meets automotive mass market requirements. The Valeo LIDAR ScaLa takes into account the requirements for autonomous driving, i.e., broad coverage with a horizontal opening angle of 145°, an angular resolution of 0.25°, and a maximum distance of up to 150 m.

Table 3 Examples from series production II

	Siemens VDO	Hella IDIS®	ScaLa Valeo
Wavelength	905 mm	905 mm	905 mm
Eye safety	Class 1 (IEC825)	Class 1 (IEC825)	Class 1 (IEC825)
Radiated power		50 W (peak)	70 W
Total power consumption			<5.5 W (over the full range of temperature and the supply voltage)
Detection area	±15° (azimuth/hor.)	±11° (azimuth/hor.)	±15° (azimuth/hor.)
	6.5° (elevation/vert.)	3° (elevation/vert.)	10° (elevation/vert.)
	Scan	Sweep	3D scan
Auto-alignment	Vertical		
Number of beams	15 ... 30	5	1
Min. curve radius	100 m	300 m	100 m
Detection range	1 ... 180 m	1 ... 180 m	1 ... 150 m
Size L × W × H	$88 \times 72 \times 57$ mm^3	$180 \times 89 \times 60$ mm^3	$140 \times 68 \times 60$ mm^3
Accuracy speed	1 kp/h	1 kp/h	1 kp/h
Specifics	Angular resolution 0.01°		

TOYOTA Research has focused on the development of a new LIDAR sensor system and the performance shown by means of a proof-of-concepts prototype. It has a range of 100 m with 10 frames per second and a resolution of 340 × 96 pixels (IEEE 2013).

5 Outlook

In recent years the following tendencies regarding LIDAR technology can be seen:

(a) The number of companies which develop LIDAR sensors for the automotive sector has decreased.
(b) As common in the automotive industry, better performance for smaller, lighter, cheaper LIDAR is required and achieved, especially in the upper class. On the

other hand, cost-optimized sensors due to pending regulations as NCAP and EUR-NCAP are developed which optimize to these partly reduced requirements.

(c) While in the past driver assistance systems such as ACC, precrash, etc., are reserved for luxury vehicles, there is a clear trend that even middle and compact classes are equipped with such kind of features, finally accompanied by increased media interest in the relevant journals. This trend asks for the more cost-effective sensors for the realization of comfort and safety functions. Thus, interest in LIDAR sensors increases.

(d) For developments that deal with the topic of "autonomous driving" (Wikipedia 2015c) (see ▶ Chap. 62, "Autonomous Driving"), which became known through the "DARPA Urban Challenge," the main sensor was a laser scanner, for example, Velodyne LIDAR (Velodyne 2015). As autonomous driving can be found in almost any roadmap of vehicle manufacturers, the corresponding activities have been started to develop LIDAR sensors that meet the automotive requirements like cost and reliability and show a performance similar to the abovementioned laser scanner system.

(e) Sensors to record reference data known as "ground truth data" for system validation of series or pre-series development applications in the area of "part/fully autonomous driving" increasingly use laser scanner, for example, Velodyne LIDAR.

References

Arita S, Goff D, Miyazaki H, Ishio W (2007) Wide Field of View (FOV) and high-resolution Lidar for advanced driver assistance systems. SAE 07AE-238

Berges A, Cathala T, Mametsa HJ, Rouas F, Lamiscarre B (2002) Apport de la simulation aux études de radar pour applications en vision renforcée. Revue de l„électricité et de l „électronique: REE. Revue de la Société des Électriciens et des Électroniciens 4:35–38

Continental (2015) Continental integrates camera and infrared functions in one compact unit (press release). http://www.continental-corporation.com/www/presseportal_com_de/themen/pressemitteil-ungen/3_automotive_group/chassis_safety/press_releases/pr_2012_10_17_srl_cam_de.html. Acce-ssed 11 June 2015

Filos J, Karseras E, Dai W, Yan S (2015) Tracking dynamic sparse signals with hierarchical Kalman filters: a case study. http://www.commsp.ee.ic.ac.uk/~jf203/dsp2013.pdf. Accessed 11 June 2015

Höver N, Lichte B, Lietaert S (2006) Multi-beam Lidar sensor for active safety applications. SAE 06AE-138

IEEE (2013) IEEE J Solid-State Circ 48, No.2. Feb 2013

Jiang QN, Ka LH, Kai WT (2015) Model-based multirate Kalman filtering approach for optimal two-dimensional signal reconstruction from noisy subband systems. http://hub.hku.hk/bitstream/10722/42383/1/44479.pdf?accept=1. Accessed 11 June 2015

Laser Components (2015a) Pulsed laser diodes. http://www.lasercomponents.com/de-en/product/pulsed-laser-diodes-at-905-nm/. Accessed 11 June 2015

Laser Components (2015b) Silicon APDs for photon counting information about the detection of the smallest amounts of light with avalanche photodiodes. http://www.lasercomponents.com/de-en/product/silicon-apds-for-photon-counting/

Luo RC, Kay MG (1998) Multisensor integration and fusion in intelligent systems. IEEE Trans Syst Man Cybern 19:901–931

Mehr W (2008) Continental – segment advanced driver assistance systems. 2 Aug 2008

OSRAM (2015) OSRAM SPL LL90 – hybrid pulsed laser diode with integrated driver stage – internet investigation/product information OSRAM. http://datasheet4u.com/datasheet/S/P/L/SPLLL85_OSRAM.pdf.html. Accessed 11 June 2015

Tischler K (2006) Sensor data fusion for cooperative perception with multiple vehicles. In: Proceedings of the SEE conference cognitive systems with interactive sensors, Paris, Mar 2006

Velodyne (2015) http://velodynelidar.com/lidar/lidar.aspx. Accessed 11 June 2015

Wikipedia (2015a) Kalman filter. https://en.wikipedia.org/wiki/Kalman_filter. Accessed 11 June 2015

Wikipedia (2015b) Laser safety. http://en.wikipedia.org/wiki/Laser_safety. Accessed 11 June 2015

Wikipedia (2015c) Google driverless car. https://en.wikipedia.org/wiki/Google_driverless_car. Accessed 11 June 2015

Automotive Camera (Hardware)

19

Martin Punke, Stefan Menzel, Boris Werthessen, Nicolaj Stache, and Maximilian Höpfl

Contents

M. Punke (✉) • S. Menzel • B. Werthessen
Continental AG/ADC Automotive Distance Control Systems GmbH, Lindau, Germany
e-mail: martin.punke@continental-corporation.com; stefan.menzel@continental-corporation.com; boris.werthessen@continental-corporation.com

N. Stache
Continental AG/Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: nicolaj.stache@continental-corporation.com

M. Höpfl
Continental AG/Continental Automotive GmbH, Frankfurt am Main, Germany
e-mail: maximilian.hoepfl@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_20

Abstract

Today's traffic environment, such as traffic and information signs, road markings, and vehicles, is designed for human visual perception (even if first approaches for automatic evaluation by electronic sensor systems in the vehicle exist – see ▶ Chap. 50, "Intersection Assistance"). This is done by different shapes, colors, or a temporal change of the signals.

It is therefore a good choice to use a system similar to the human eye for machine perception of the environment. Camera systems are ideal candidates as they offer a comparable spectral, spatial, and temporal resolution. In addition to the "replica" of human vision, specific camera systems can provide other functions, including imaging in infrared spectral regions for night vision or a direct distance measurement.

This chapter covers details on specific applications of camera-based driver assistance systems and the resulting technical needs for the camera system. Use cases covering the outside and inside of the vehicle are shown. The basis of every camera system is the camera module with its main parts – the lens system and the image sensor. The underlying technology is described, and the formation of the camera image is discussed. Moving to the system level, basic camera architectures including mono and stereo systems are analyzed. The chapter is completed with a discussion of the calibration of camera systems.

1 Applications

Due to their versatility, camera systems in automobiles are used both for surveillance of the interior of the car as well as the surroundings (Loce et al. 2013). The following section discusses these applications and explains the specifics of using camera systems for these.

A first driver assistance system using a camera system was the so-called rear view camera. The driver is assisted by the display of a live video stream on a monitor system. Advanced functions using computer vision are, for example, used in the high-beam assist function. In these systems, the video image is not displayed, but a specific function is directly derived from the camera image.

In addition, cameras are used in the interior of the vehicle. Here, two main functions are of importance. First, the driver monitoring for detecting the state of the driver and the driver's intention and, secondly, the use of camera systems in the context of an advanced man–machine interface for controlling functions, for example, by means of gesture and gaze control.

1.1 Driver and Interior Detection

The application of a camera in the interior requires a different camera design compared to the exterior. For driver monitoring applications, the object distance

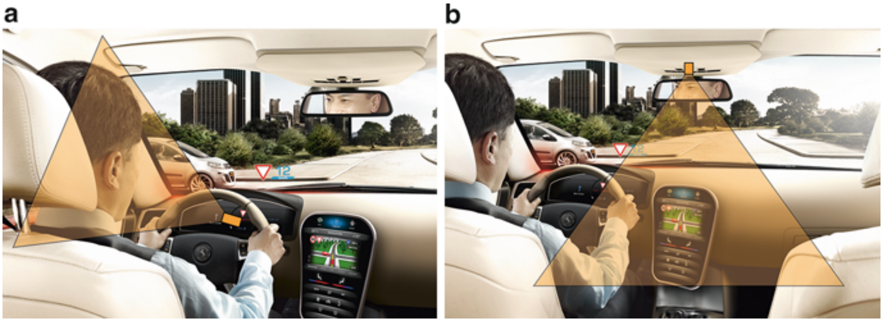


Fig. 1 Field of view of an interior camera for (**a**) driver monitoring and (**b**) hand gesture recognition

is smaller, and therefore, the optical parameters, e.g., depth of focus, differ significantly. Imaging in the near-infrared spectral region with artificial lighting is selected, because it is invisible to the driver and can ensure a high image quality at nighttime and in quickly changing light conditions.

1.1.1 Driver Monitoring and Gaze Control

New driver assist systems support the driver and at the same time introduce a new role to the driver: the driver turns from an acting role to an observing role. The driver is more of a moderator.

The different assist functions provide information to the driver, who can start and adopt the functions to the traffic situation or can completely take over the driving task again.

The information about the driver, i.e., state and intention, will be part of the interaction concept in the car enabling a holistic HMI (human–machine interface).

Initially, the interior camera applications are focused on drowsiness detection. Additional functions like driver identification are used to apply personal preferences, adaptive warning depending on the driver's head position and gaze direction, as well as an augmented HMI (augmentation: overlay of HMI information with objects in the car's environment), e.g., in the augmented reality head-up display, and are also enabled by interior camera technology.

A driver monitoring system can include up to four cameras. The number of cameras in the system depends on the application of the system. The more cameras, the higher the precision for the gaze detection and the wider the detection area. In Fig. 1a, a possible field of view for a mono camera driver monitoring system is shown.

1.1.2 HMI: Hand Gesture Recognition

Today, multimodal HMI concepts allow the driver to interact with the system in many different ways, including traditional knobs, push/turn control knobs, and modern touch surfaces. The recognition of hand gestures in space and the

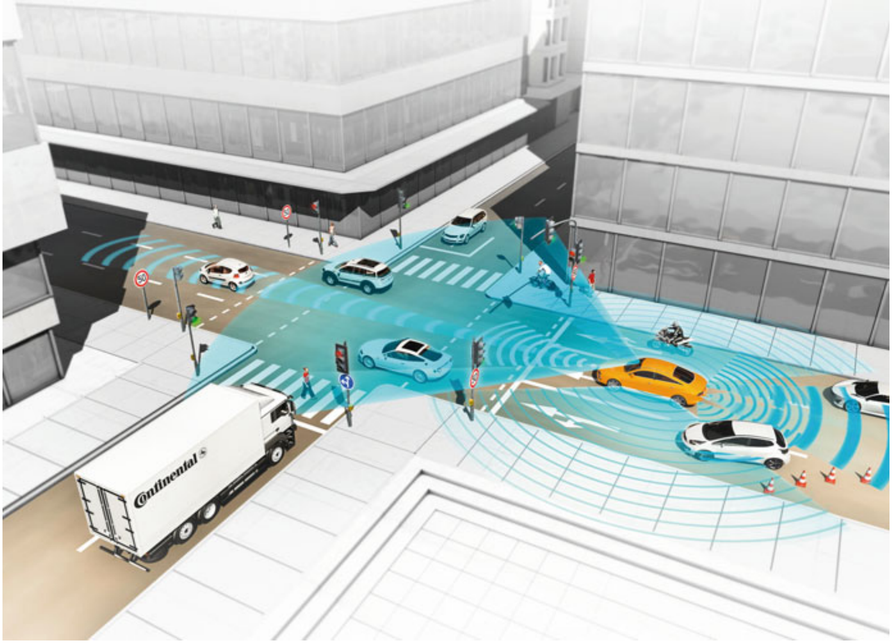


Fig. 2 Environmental detection by different sensor systems

approximation of the driver's hand to the central touch display can be detected by an interior camera.

Two technologies can be applied for hand gesture recognition. On the one hand, a conventional 2D image is used and, on the other hand, a 3D depth map, which is generated by a *time-of-flight- (TOF)* sensor. Both use similar optics, but differ in the sensor and the near-infrared illumination unit. Time-of-flight cameras have a smaller resolution, but can provide depth information, which makes the object detection and object tracking work easier. Fig. 1b shows an interior camera for hand gesture recognition and its field of view.

1.2 Environmental Detection

The aim of the environmental detection is the full recognition of all relevant road users, road scenery, and road signs in order to be able to react accordingly. Here, a variety of different sensor technologies can be considered thus ensuring both the correct recognition as well as the best possible availability.

Figure 2 shows such a situation. It depicts the field of views of various sensors on the car. The sensors are designed such that the fields of view overlap, and different detection ranges can be achieved. In this example, the detection by camera systems as well as short and long range RADAR sensors is shown.

Fig. 3 Combination of a camera system with a LIDAR sensor (Continental SRLCam)

Today, different types of sensors can be integrated in a single housing. An example is the SRLCam (Fig. 3) of Continental, where a short-range LIDAR (SRL) is combined with a multifunction camera. This enables a compact and inexpensive sensor. The performance of the emergency braking function is increased since the systems are working with a sensor fusion (see also ▶ Chap. 23, “Data Fusion of Environment-Perception Sensors for ADAS”). Another example of sensor combination is RADAR and camera system (RACAM) from Delphi (Delphi 2014).

1.2.1 Front View Cameras

Front view cameras are usually placed behind the windshield of the automobile and close to the rearview mirror (or even integrated in the rearview mirror (Gentex 2014)). This position has the great advantage in the wide field of view and the protection through the windshield. In addition, the windshield in front of the camera is swept by the wipers thus ensuring a clean view. Exceptions to this placement position are camera systems working in the far-infrared range. These systems are installed in the headlights or radiator area due to the low transmission of the windshield in this spectral region (Källhammer 2006).

Cameras in the Visible Spectral Range

The majority of the systems currently used are operating in the visible spectral region and thus similar to the human eye. As previously explained, this enables the detection of all relevant traffic signs, etc. by the camera systems that are also relevant for humans.

The *high-beam assist* function is used to automatically turn on and off the high beams of the car. More advanced systems also have a variable range control and independently dim certain areas. Such systems are enabled by segmented LED (light-emitting diodes) headlamps. The various light functions are controlled by the camera that is analyzing the oncoming traffic. Important for this camera function is the ability to distinguish at least the color of the tail and front lights of a vehicle.

Therefore, color-sensitive camera systems are used (details in Sect. 3.3.4). Another requirement for this function is the need for a high dynamic range of the camera system, arising from the large differences of light intensities that occur at night (see also Sect. 2.1.4).

Traffic sign recognition detects all relevant traffic signs (e.g., speed limits, one-way street labels) and the information is made available to the driver. The traffic sign recognition feature requires a high-performance camera system (Stein et al. 2008). The traffic signs must be recorded with a high resolution (>15 pixels per degree) so that the character recognition functions correctly. Since most traffic signs are placed at the edge of the road and the automobile is moving at high speed, a short (<30 ms) exposure time is necessary to avoid strong motion blur.

For safety reasons, many vehicles are equipped with a *lane detection* feature (see also ▶ Chap. 49, "Lane Change Assistance"). Important for this function is a very high recognition rate even at nighttime and with poor road conditions. It is advantageous if the camera system can distinguish colors, since a detection of different colored markings on a street is enabled, for example, at construction sites (see also Sect. 3.3.4).

In order to respond to other road users (e.g., vehicles, pedestrians, cyclists), a robust *object recognition* is necessary. Here, a variety of aspects are relevant for the camera system (Raphael et al. 2011). For a large detection range, e.g., for vehicle detection on highways, a high resolution is necessary. Pedestrian detection benefits from a large field of view. In general, a high sensitivity of the camera system is very important.

In particular, for object recognition, stereo camera systems have many advantages. By generating a depth map, various objects can be detected, and the distance to the vehicle is measured directly. In addition, the so-called free space detection is possible, which identifies areas on the street that are possible to use. Another function which can be implemented using a stereo camera system is the *road condition recognition*. In this way, the vehicle can adjust to bad road conditions, etc. (Daimler 2014).

Cameras in the Infrared Spectral Range

A disadvantage of cameras operating in the visible spectral range is the insufficient sensitivity at very low light conditions. A possible alternative or add-on is the usage of cameras operating in the infrared spectral range (see also ▶ Chap. 43, "Visibility Improvement Systems for Passenger Cars"). Here, mainly two approaches are used. In the first approach, the traffic scene is illuminated using special infrared headlights (LED or halogen). The camera system is equipped with spectral filters, so that the camera sensor is only sensitive to the IR wavelength. Another possibility is the use of special cameras that are sensitive in the far infrared (FIR). These cameras can directly detect heat radiation of pedestrians and animals and thus triggering an assistance function. A major drawback is the high system cost since these camera systems are not based on conventional image sensors (Källhammer 2006).

1.2.2 Cameras for Detection of the Surroundings of the Car

Compared to the front view cameras, the class of camera systems that control the surroundings of the car has other objectives. These camera systems often cover a wide field of view and also provide a video stream to the driver.

Rearview Cameras

The camera module of rear view cameras is usually integrated into the tailgate of the car (e.g., close to the license plate). The video stream is then displayed on a monitor on the instrument panel. Firstly, accidents by running over people behind the car can be avoided. On the other hand, advanced systems support the driver during parking by displaying a graphical overlay (see ▶ Chap. 44, "Parking Assistance").

Surround View Cameras

Surround view systems are equipped with four or more cameras around the vehicle. The video information of the cameras are transmitted to a central processing unit. Camera modules for such systems are usually equipped with the so-called fish-eye lenses, which allow a horizontal field of view of more than 180°. From the camera images, a 360° view of the environment is generated and provided to the driver as a parking aid on a monitor. In future, not only a parking support will be possible, but the camera images are also used for object detection and general environmental detection in addition to the front view camera.

Mirror Replacement

An approach that is likely to play a significant role in future is the replacement of normal exterior mirrors with camera systems. This is beneficial for fuel consumption (less drag) and opens up completely new design possibilities. Similar to surround view systems, a high dynamic range and a good color reproduction of the camera image are necessary for a good quality of the displayed video. Such camera systems are addressed in the international standard ISO/DIS16505 (ISO/DIS 16505).

2 Cameras for Driver Assistance Systems

As shown in the previous sections, the application range of camera systems in vehicles is very diverse. Therefore, many different variants of the camera systems exist. The following section further describes some aspects of the design.

A basic camera architecture is shown in Fig. 4. An object or a scene is projected through an imaging lens onto the image sensor. The pixels of the image sensor are converting the photons into an electronic output signal, which is analyzed by a processor unit. In the case of a direct presentation to the user, the output is displayed on a screen.

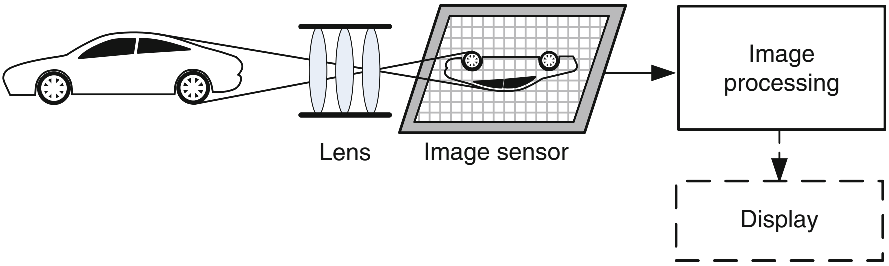


Fig. 4 Diagram of a basic camera architecture

2.1 Criteria for the Design

For the design of a camera system, both the analysis of the individual parts and the complete system is necessary. Many performance parameters are influenced by several parts of the system.

The lens of the camera system is very important for a good overall performance. Among other components, it affects the possible resolution, field of view, depth of field, color reproduction, and also the sensitivity of the system. Since an optical system never produces a perfect image (see Sect. 3.2), potential errors, such as a distortion, must be corrected.

The optical image is converted by the image sensor into digital values. Therefore, the design and adaptation of the optical system to the image sensor are crucial to image quality. The sensor mainly influences the resolution (number of pixels), field of view (number and arrangement of pixels), dynamic range, color reproduction, and especially the sensitivity (see Sect. 3.3).

In the next step of the processing chain, the image quality is affected by the image processing steps in the processor. In addition, the performance of the computer vision algorithms crucially depends on the performance of the processing unit.

2.1.1 Field of View

The *field of view (FOV)* of a camera plays an important role in the application and is essentially defined by the lens and the image sensor. One distinguishes the field of view in horizontal and vertical direction (HFOV, VFOV).

In front view camera systems, usually the horizontal field of view is the most important. However, the different assistance functions require differently wide horizontal fields of view. Relatively large HFOV values are needed in case of lane detection (in tight curve scenarios) and object detection (e.g., for the detection of crossing vehicles or pedestrians that are running onto the road). Values of more than 40° are meaningful for these applications (see Fig. 5a).

Another aspect in the choice of field of view values is the influence of motion blur in the image. For large field angles, the position of the object in the image

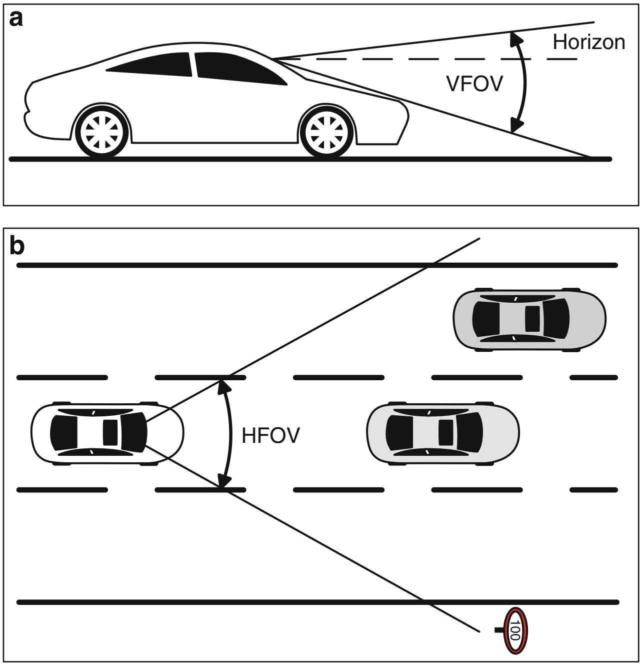


Fig. 5 Vertical (**a**) and horizontal (**b**) field of view of a front view camera

during the exposure time varies greatly. This manifests itself in a blurred image in the border areas for longer exposure times. Therefore, the maximum usable angle is also limited by this effect.

The vertical field of view is determined primarily by the mounting height and the minimum detection distance at close range. An exemplary calculation for a passenger car results in an angle α of 18° below the horizon at a height h of 1.3 m and a distance d of 4 m (see Fig. 5b).

$$\alpha = \tan^{-1}(h/d)$$

In surround view systems, almost all camera modules are using a horizontal field of view of more than 180°. This is due to the desired 360° representation of the vehicle environment. In order to calculate a single picture from the single frames, an overlap between the FOVs of the different cameras is required.

In the area of driver monitoring, the imaging of the drivers head is important. Different anatomical requirements and installation situations result in a field of view of about 40–50°. For gesture recognition, for example, using a camera module in the functional unit in the roof, larger FOVs (>50°) are usually selected. This provides the driver with more freedom in the gesture operation.

Fig. 6 Effect of decreasing resolution using the example of a traffic sign (480 × 650/72 × 96/36 × 48/24 × 32/18 × 24/12 × 16)

2.1.2 Camera Resolution

The possible resolution of a camera system is a complex interplay of the resolution of the lens and the image sensor as well as the image processing. For the design of a camera system, a purely theoretical analysis is necessary as a first step. Here, the object to be resolved (e.g., a vehicle 100 m away) is analyzed and the necessary resolution for the image processing defined (e.g., 10 pixels per vehicle width). Then, using the geometric relations, a necessary resolution in pixels per degree is calculated. Usual values are >15 pixels per degree in the case of driver assistance functions in the field of environmental detection. In particular, traffic sign recognition has high demands toward the overall resolution (e.g., to recognize additional characters). In Fig. 6, the effect of different resolutions is shown. While the form and the warning symbol can be extracted from the right images, this will not work with pictogram and text in the supplementary sign. In the area of driver monitoring, higher resolutions may be necessary, e.g., to realize an *eye tracking*. An important aspect in choosing the optimal resolution is – in addition to the geometric requirements – the available computing power for image processing.

2.1.3 Color Reproduction

In the area of advanced driver assistance systems (ADAS), the widely used CMOS image sensors are predominantly sensitive in the visible (VIS) and the near-infrared (NIR) spectral region. A separation in different color channels is realized via color filters on the image sensor (see Sect. 3.3).

How Sect. 1 describes, the color reproduction is of great advantage for applications in the field of front view and surround view cameras. While in surround view systems, a realistic representation of the camera image on the monitor is very important; in the front view applications, the ability to distinguish between individual color channels is critical.

An example of the importance of color separation is shown in Fig. 7. While a camera system with color information (a) is clearly able to distinguish between white and yellow lane markings, in the case of a monochrome image (b), this is no longer possible.

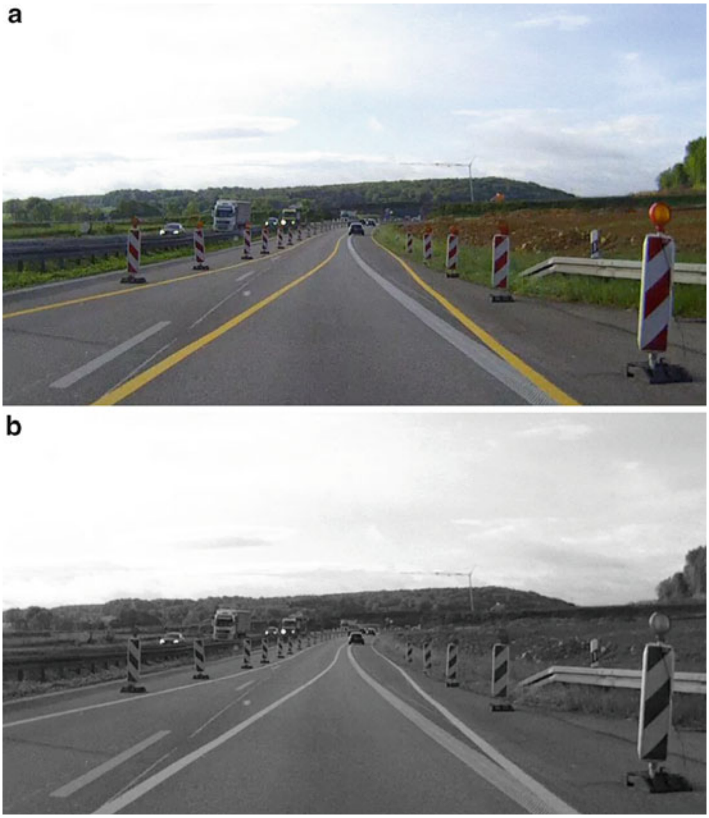


Fig. 7 Importance of color separation for the detection of lane markers

2.1.4 Dynamic Range

The dynamic range (DR) of a camera system describes the ability to record both dark and bright areas in the image. In dark areas, the dynamic range is limited by the noise limit of the image sensor and in bright areas by the saturation limit of the image sensor. In addition to the image sensor, the dynamic range of the camera lens and the optical path defines the overall system dynamic range. The dynamic range of the lens is adversely affected by stray light. In addition, effects such as ghosting and flare can occur, which reduces the image quality especially in strong front light situations. Elements in the optical path such as the windshield can limit the total dynamic range of the system.

Traffic situations in the field of driver assistance systems are showing large differences in brightness levels. Examples are scenes with low-standing sun (see Fig. 8), entries/exits of tunnels and parking garages, or oncoming vehicles at night. A pavement marking at night can exhibit a luminance L of < 10 cd/m^2, while the headlights of a vehicle in the same scene can have a luminance of up to 100,000 cd/m^2 (Hertel 2010). Due to these issues, a very high dynamic range of the camera

Fig. 8 Illustration of a traffic scene with a high dynamic range

system is necessary. With an appropriate design (see also Sect. 3.3), one can achieve more than 120 dB dynamic range within imaging systems. The dynamic range is defined as follows:

$$DR(dB) = 20 \cdot \log_{10}(L_{\mathrm{MAX}}/L_{\mathrm{MIN}})$$

3 Camera Module

Camera modules can be very different in design. A camera module is defined here as the combination of lens, image sensor, electronics, and packaging technology. It is of course possible to accommodate more components within the camera module assembly, such as image processors.

3.1 Construction of a Camera Module

The main elements of the camera module are the lens and the image sensor which are held by a suitable mechanical structure. Additionally, electronic components and a connection to the image processor are necessary. As shown in Fig. 9, the single elements are assembled into a compact module using different packaging technologies.

The camera lens consists of multiple single lens elements that are built into a complete lens system. Part of the optical system is often an *infrared cut-off filter (IRCF)*, which only transmits the non-infrared parts of the light spectrum. The lower part of the camera module exhibits the image sensor, the *printed circuit board (PCB)*, and the electronic components.

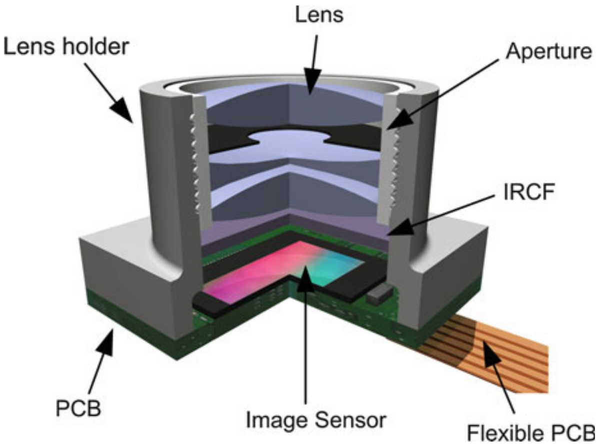


Fig. 9 Schematic structure of a camera module

Of course the design parameters such as resolution and dynamic range are critical for the basic design of a module. For a robust design, in particular the environmental impacts during the lifetime of the system – such as changes in temperature and humidity – are critical. The different applications lead to different requirements for the module design, since a camera in the area of surround view has direct contact with the external environment, while modules for front view and driver monitoring are only used inside the passenger compartment.

3.1.1 Electronics

Image sensors have a wide variety of analog and digital inputs and outputs. The most important are the power supply for the analog and digital parts of the image sensor, external timing signal (clock), digital control inputs and outputs, and interfaces for the configuration and image data transmission to the computing unit.

The transfer of settings (e.g., integration time) is realized via the configuration bus system with a low bandwidth. The transmission of the image data can be achieved via a parallel or serial interface. As data rates increase significantly with higher pixel counts and refresh rates, a trend is the usage of high-speed serial interfaces, such as the Camera Serial Interface (CSI) (MIPI 2014).

3.1.2 Packaging Technology

To ensure compliance with the requirements of the optical parameters over the lifetime of the camera module, the choice of suitable packaging technologies is very important. Firstly, the attachment of the image sensor to the PCB is of importance; on the other hand, the alignment of the lens system to the image sensor is critical.

The image sensor and other electronic components can be assembled to rigid printed circuit boards made of organic material, flexible printed circuit boards, or ceramic carriers. Image sensors are mounted onto the PCB in packaged or unpackaged form.

The alignment of the camera lens to the image sensor can be achieved using several mechanisms. Common are approaches that actively align the image sensor to the optical system via multi-axis adjustment and fastening with an adhesive bond. The biggest advantage is the precise adjustment. Alternatively, an adjustment only in the direction of the optical axis is possible, which is realized by means of a screw thread. However, a possible tilt of the optical axis to the image sensor plane will not be corrected with the latter method.

3.2 Optics

The camera lens typically comprises a stack of single lens elements, optical filter elements, and the overall lens housing. The applied lens designs in the automotive industry are a trade-off between optical performance, costs, and durability. This trade-off strongly influences the choice of materials of the single lens elements, as well as their number, and the choice of lens housing materials. An example for durability and robustness requirements, respectively, is the focal point stability over temperatures ranging up to 100 °C.

3.2.1 Lens Design for Automotive Applications

Optical materials of choice are standard crown glasses or flint glasses for their moderate costs. Additional single plastic lens elements might also be applied, but the high use-case temperatures in conjunction with the high robustness requirements limit the number of plastic single lens elements.

Single glass lens elements are fabricated by conventional polishing methods. Alternative techniques, e.g., lens pressing for aspherical glass elements, are not preferred for their higher costs. Plastic lens elements are fabricated by injection molding and are available in both spherical and aspherical shapes (Fischer 2008).

A camera lens for driver assistance systems is typically delivered in a water proof and humidity proof housing, either made of plastic or metal, to assure use-case robustness. The surfaces of the housing material are preferably black by choice of material or are covered by a black coating to minimize stray light in the optical path. An additional optical filter to suppress UV radiation and infrared radiation is inserted into the optical path if only the visible wavelength range is of interest. Cameras for applications in the near-infrared wavelength range utilize a band–pass filter to suppress all light except for the near-infrared light. All optically active surfaces are also coated with antireflection layers to avoid parasitic multiple reflections resulting in lens flares and other unwanted image artifacts.

3.2.2 Optical Requirements for Driver Assistance Applications

The optical lens design depends on the specific application. Important parameters are the field of view, the light sensitivity, the resulting image distortion, and the image sharpness. In a first approximation, the field of view is given by the focal length, which is typically constant in camera lenses for automotive applications because of robustness and cost considerations.

Lens Design for Front View Cameras

Low light conditions and the demand for sufficient frame rates require a short exposure time in a front view camera, which is compensated for in terms of light sensitivity by a low f-number, i.e., a large aperture and/or a small focal length. F-numbers < 2 are in use. Large apertures introduce image aberrations to sharpness and color, which need to be removed by additional single lens elements, resulting in multiple single lens elements in an automotive camera lens.

The ability of a camera lens to produce a sharp image is expressed by the *modulation transfer function (MTF)*. The MTF describes how well real-world contrasts at different spatial frequencies are reproduced at the image plane.

Camera lenses become more expensive with better performance, because more single lens elements are required. Therefore, an overshoot in performance has to be avoided due to cost considerations, and thus, it is important to know the minimum required image quality in order for the driver assistance to run properly. Image quality is degraded by the usual optical aberrations, which are accumulated and expressed in the so-called point–spread function (PSF). It describes how a real-world point source (or point in the scene) is reproduced on the image plane, where it is typically a spatial intensity distribution and not a point anymore (Sinha 2012). A practical example is a head light at night in the far distance. Another image-forming influence is the sampling nature of the image sensor and the discrete spatial distribution of the pixels. A general design rule for lenses is to keep the PSF close to the size of a pixel, because PSFs smaller than the pixel size (provided by better and more expensive lenses) cannot be resolved at all.

Image sharpness varies along the optical axis, i.e., the position of the image sensor is important. This so-called depth of focus provides a narrow region of sufficient sharpness and is smaller at lower f-numbers. By means of focusing close to the hyperfocal distance, it is nevertheless possible to image large portions of the real-world scene with high sharpness and leave margins for temperature-induced focal point shifts.

Image aberrations typically found in automotive camera lens designs are spherical aberration, chromatic aberration, coma, astigmatism, and field curvature and act on the intensity distribution of the PSF (Hecht 1998). Chromatic aberration, for example, results in wrong reproduction of colors. All image aberrations can be reduced to negligible amounts by using more single lens elements, aspherical lens elements, or expensive lens materials (Sinha 2012). Other image artifacts, for example, lens flare and ghost images, can appear due to multiple internal reflections and scattering on optical and mechanical surfaces (Reinhard et al. 2008).

Image distortions of a few percent are negligible in lens design, because they can be corrected by software algorithms. Therefore, designing an expensive and distortion-free camera lens is not required. Especially for correspondence in stereo camera images, the distortion correction has to work accurately, and the lens distortion is therefore measured during fabrication (see Sect. 5).

Lenses for Surround View Cameras

Surround view cameras typically require a large field of view. As a result for moderately priced lenses, a large distortion is observed toward the edges of the image. The high distortion needs to be corrected very accurately, because images from more than one camera are stitched together to generate the surround view, and finding correspondences in single images is the basis of image stitching. Also, high brightness reduction toward the edges (optical vignetting) needs to be considered and corrected. Additionally, the first lens is usually exposed to the environment and has to be designed in a robust way (e.g., glass instead of plastic).

Lenses for Interior Cameras

Lenses for interior cameras need to monitor the driver in a range from 40 to 100 cm. Typical f-numbers for such lenses are $f > 2$, because a relatively large depth of focus is required matching the abovementioned range.

3.3 Image Sensor

There are two fundamental architectures of digital image sensors – *CMOS (complementary metal oxide semiconductor)* and *CCD (charge-coupled device)*. CCD sensors had great advantages over CMOS sensors until the end of the last century in terms of noise performance, but are hardly used anymore in driver assistance systems. Today, the advantages of CMOS sensors predominate, and disadvantages in noise performance are compensated (Miller et al. 2004; El Gamal and Eltoukhy 2005).

Active pixel sensors (APS) are based on CMOS technology. The terms APS and CMOS sensors are used interchangeably herein. The following sections briefly describe the image sensor attributes sensitivity and noise, resolution, concepts to increase the dynamic range, the reproduction of color by use of color filters, and shutter concepts of APS sensors.

In CCD sensors and *PPS (Passive Pixel Sensor)*, the charges generated by the pixel need to be transferred to a common conversion node where they are converted to a voltage. Contrary, APS have active pixels in the meaning of each pixel individually converting charge to voltage and having integrated analog-to-digital converter, which converts the voltage to a digital signal.

Since the image information of the sensor is already available in digital form, it can directly be used externally as well as internally.

Examples for internal use of digital data are the creation of a histogram of the recorded scene, the realization of automatic exposure control, and the generation of high dynamic images or to guarantee functional safety.

Good guidelines for the characterization of image sensors are standards such as EMVA1288 and ISO standards such as 12232, 12233, 14524, 15739, and 16067.

3.3.1 Sensitivity and Noise

The *sensitivity* of the image sensor is substantially influenced by its noise performance. Therefore, noise performance over the required temperature range is an important criterion when selecting an image sensor.

The noise of an image sensor mainly consists of the temporary, photon shot, and spatial noise components (Holst and Lomheim 2011; Yadid-Pecht and Etienne-Cummings 2004; Fiete 2010).

At very low signal levels, *temporal noise* is the dominant source of noise, which among other things is composed of reset, thermal noise, and quantization noise.

Reset noise describes differences of the remaining charges after reset at the start of the integration and can be suppressed by the use of *correlated double sampling (CDS)*. An additional readout is performed just before the integration starts to gather the current reset level, and its result is subtracted from the total signal after integration.

Dark current noise is generated by charges caused by thermal energy. With increasing temperature, this noise component increases nonlinearly.

At medium and high signal levels, the *photon shot noise* dominates. It describes the statistical intensity distribution of the number of incident photons. As it has its source outside the sensor, it can't be compensated by the image sensor. Photon shot noise is calculated as the square root of the generated signal and thus also determines the maximum signal-to-noise ratio (SNR).

Quantization noise is caused by the inaccuracy in the conversion of the electrical signal into a discrete digital signal and can be minimized by a higher bit depth of the analog-to-digital converter.

Spatial noise describes relative static differences in the offset and the gain of individual pixels (also known as fixed pattern noise – FPN) caused by variations in the current-to-voltage conversion of the active pixels or as a column FPN in the amplifier and A/D converter circuits of the respective columns. Similar to dark current noise, there is a nonlinear dependency on the temperature of the sensor (Yadid-Pecht and Etienne-Cummings 2004; Fiete 2010; Theuwissen 2008).

3.3.2 Resolution

The quality of digital video is not only influenced by the spatial resolution, but also by contrast and temporal resolution. These influences describe the number of pixels on which an object is imaged, the number of gray levels a scene can be resolved, and the time interval between two images (Fiete 2010).

Spatial Resolution

To reconstruct a structure by a digital image sensor, it must be mapped to multiple pixels. The required number of pixels per angle is therefore determined by the smallest structure one wants to resolve. The total number of pixels is determined by the FOV and the desired resolution in pixels per angle (see Sect. 2.1.2).

To realize a high number of pixels on an image sensor, there are two options. The first is to keep the pixel pitch (distance between the centers of neighbored pixels) and increase the die size. The second approach is to lower the pixel pitch (and thus the pixel size) while keeping the die size. For cost reasons, the latter approach is chosen in most cases. From the perspective of the signal-to-noise ratio, keeping the pixel pitch is preferable, because with smaller pixels, also the filling factor further decreases. Furthermore, less temporal noise will be generated when pixel volume decreases, but it will not go down proportionally (Fiete 2010; Theuwissen 2008). Care must be taken to ensure that the disadvantages of decreasing the pixel size can be compensated by, e.g., new pixel design and improved manufacturing processes.

Contrast Resolution

For the detection of objects, a high differentiation of object brightness is advantageous to also detect, for example, a dark-clothed person at night. This is achieved by an A/D conversion of the signal with a bit depth of 8–10 bit in simple systems and by 12 bit and more at more sophisticated systems. *HDR sensors (high dynamic range)* internally often work with much higher bit depths, which are then compressed for easier data transfer to a bit depth of typically 10–14 bit.

Temporal Resolution

The refresh rate is the time interval between two recordings. A low refresh rate involves the risk that the system’s reaction on an event occurs too late or the event is totally missed and furthermore complicates the tracking of objects. A high frame rate increases the demands on the interface and further image processing. Typical values are at about 30 frames per second.

3.3.3 Dynamic Range

The usable dynamic of an image sensor is defined by the range of light intensities that can be digitally resolved, limited by the clear distinction between signal and noise and the pixel’s saturation level. In the real world, dynamics of about 120 dB, corresponding to a contrast ratio of 1: 1.000.000, must be expected (Darmont 2012).

Linear sensors have a maximum dynamic range of 60–70 dB, so they can’t represent the entire dynamic range of the scene. HDR sensors can achieve dynamic ranges of about 120 dB.

In the following, two time-based (lateral overflow and multi-exposure) and one spatial (split pixels) HDR concepts are discussed. In time-based method, every single pixel performs several partial integrations or several individual integrations in succession, while spatial concepts integrate spatially separate sub-pixels simultaneously (Yadid-Pecht and Etienne-Cummings 2004; Fiete 2010; Darmont 2012).

Lateral Overflow

This concept is based on analog resets which do not fully reset the pixel but allow a certain level of charges to remain (partial saturation). Several such partial resets are performed during the integration time, each allowing an increasing amount of charges to remain while the time intervals between the partial resets are getting shorter (Darmont 2012). Capturing moving objects may result in some motion artifacts if partial saturations are reached due to the repeated partial integration and superposition into a single image.

An advantage of the method is that the high dynamic information is directly available, and hence no information needs to be cached, which makes the process very applicable for global shutter sensors (see Sect. 3.3.5).

Multi-exposure

In this concept, successively several individual captures with different sensitivities (by, e.g., integration time or gain) are taken, and this information is combined to a high dynamic range image. Moving objects are captured at several points in time and thus at different positions within the frame, causing noticeable motion artifacts. Therefore, a small time gap between two single integrations is very important and also an appropriate combination scheme for the transform into an HDR image (Solhusvik et al. 2009).

The benefits of the multi-exposure method are the SNR performance as CDS and that other corrective measures can be performed individually for each of the single integrations.

Split Pixels

In split pixel concept, every pixel is divided into two or more sub-pixels. Different sensitivities of the sub-pixels are achieved by different-sized photosensitive areas (typical ratios of about 1: 4–1: 8), different gain stages, or choice of different integration times.

The great advantage of this approach is the time-parallel capturing of the scene in different dynamic ranges, resulting in lower motion artifacts than is the case with *lateral overflow* or *multi-exposure* process.

The disadvantage is that the dynamic range is generated by only two single integrations, which results either in a lower overall dynamic or a strong compression of the dynamic range (Solhusvik et al. 2009; Solhusvik et al. 2013). Furthermore, there is a somewhat poorer fill factor, since two photodiodes and their circuits must be accommodated in the pixel pitch.

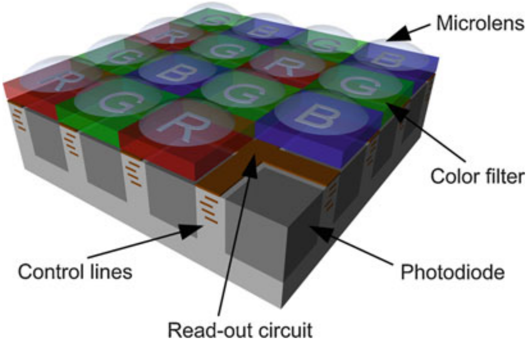
3.3.4 Color Reproduction

A photosensitive element can only provide the information that electrons are generated by incident light, but is not able to provide information about the wavelength of the incident light. Most digital image sensors are therefore monochrome sensors and can only supply gray values of the captured scene.

To assign color information, a color filter needs to be inserted in the pixel's optical path, so that only certain wavelength ranges reach the pixel. Typically, *color filter arrays (CFAs)* with red (R), green (G), and blue (B) color filters are used. To create RGB color information for each pixel, an interpolation with surrounding pixels having different color filters is performed. It needs to be mentioned that a large amount of the incident light power is absorbed by the filters, which leads to a lower effective sensitivity. Depending on whether sensitivity, sharpness, or color fidelity is the focus of development, different approaches might be beneficial.

The *Bayer CFA (RGGB)* is a classic color camera concept with sophisticated approaches to color reconstruction with minimal reduction in the sharpness by color

Fig. 10 Scheme of a CMOS image sensor with RGB color filter



interpolation (Brainard 1994). A Bayer CFA unit cell is composed of two diametrically opposed green filters and one single red and blue filter (see Fig. 10).

Also, very common in front view applications is the *Red-Monochrome-CFA (RCCC)*. Three pixels without color filters (clear – C) are complemented by a red pixel. This preserves the sensitivity of the pixels and produces a high-resolution gray-scale image. The teh reproduction of color is only possible in the sense of "non-red," by comparing the red pixel output to that of the clear ones. The lower the ratio, the more the object is rated as blue (non-red). In the automotive practice, this is sufficient to distinguish between front lights (white) and tail lights (red) (Stein et al. 2008).

3.3.5 Electronic Shutter

In digital cameras, there is usually no mechanical shutter which determines the integration time. Instead, an electronic "shutter" is used, where the pixel is reset just before the integration begins to set it back to its origin state and read out at the end of the integration time to gather the signal actually generated by the incident light. Either global shutter or rolling shutter can be used, which differ in the timing (Yadid-Pecht and Etienne-Cummings 2004; Fiete 2010).

In a *global shutter (GS)* imager, the integration of all pixels of a pixel array is started and stopped simultaneously. The advantage of this technology is that a moving object retains, apart from motion blur, its shape in the captured image (see Fig. 11, lower illustration). Also randomly pulsed light sources such as LED vehicle lighting or active traffic signs produce a homogeneous result, as long as the pulses occur during the integration time.

Disadvantage to the global shutter is the need to store the image information until it is read out. This results in the need for additional transistors per pixel and analog memory cell arrays (sample and hold circuit). This leads to parasitic effects, and also additional chip area is required. The resulting poorer fill factor results in a higher noise level compared to the electronic rolling shutter, where the information is output directly (Yadid-Pecht and Etienne-Cummings 2004).

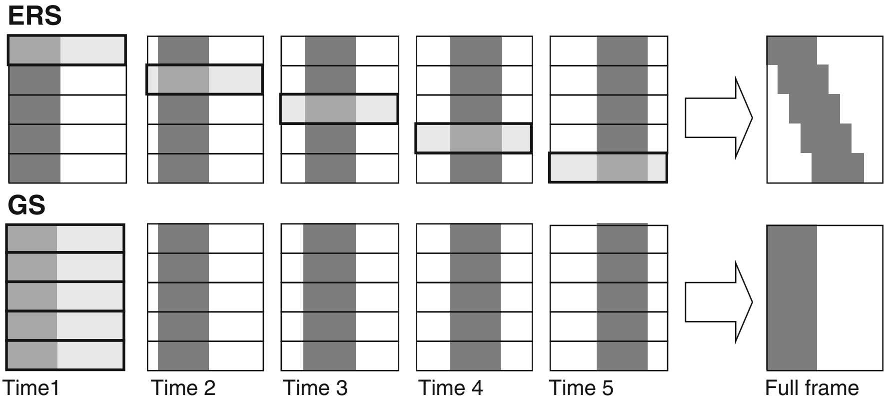


Fig. 11 Difference of *electronic rolling shutter (ERS)* and *global shutter (GS)* using the example of a cube with the direction of movement to the right. ERS records moving objects at different rows at different points in time, GS all the rows at the same time

In an *electronic rolling shutter (ERS)* imager, the integration of each pixel is, with an interval of one clock cycle, individually started and also individually stopped after the integration time (Yadid-Pecht and Etienne-Cummings 2004; Baxter 2013). The shutter "rolls" across the single pixels over the entire active pixel array. At each clock cycle, the information of that single pixel, which just stopped its integration, is available and is read out directly. This eliminates the need of the sample and hold circuit. That is why ERS has great advantages in the SNR compared to the GS. Furthermore, CDS is easier to integrate with ERS due to lower number of pixel transistors (Yadid-Pecht and Etienne-Cummings 2004).

Disadvantage of sequentially integrating the pixel is the time gap of the integration between the image lines. So the upper part of an object that moves horizontally through the image is taken at an earlier point in time than the lower part, which leads to the well-known effect of distorted or "leaning" objects. Recording pulsed light sources (active variable traffic signs) results in the effect that short single pulses are seen only by those few lines which are integrating photons at that time.

Figure 11 shows the position of a moving cube and the line (light gray), whose integration was started at different times while the ERS image (top) was captured. When the lines are set together, this will result in a distorted representation of the cube. With GS (below), the integration of all pixels happens at the same time, resulting in an undistorted representation of the cube.

4 System Architecture

To meet the demands of all required functions, the system architecture requires the correct design of hardware and software components as well as the image processing algorithms. In addition, the mechanical design of the system and the mechanical and electronic connections with the vehicle are of importance.

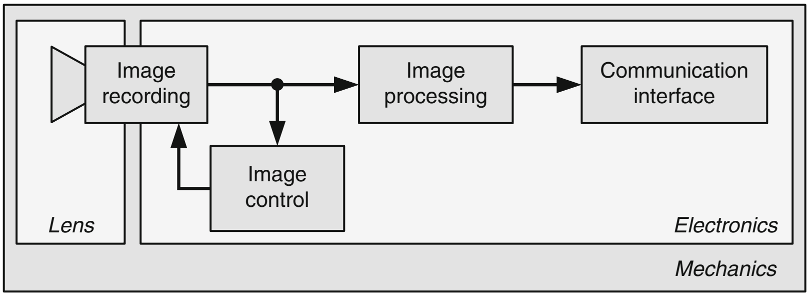


Fig. 12 System components of a driver assistance camera system

Since camera systems for driver assistance functions are safety-related vehicle components (e.g., by braking intervention), the system must comply with the ISO norm 26262 ("*Road vehicles – Functional safety* ") (ISO 26262). Here, depending on the function, different *automotive safety integrity levels (ASIL)* are necessary to be implemented.

4.1 System Overview

A camera system consists of the components for image acquisition, image acquisition control, image processing, and the communication to the vehicle. This is schematically illustrated in Fig. 12. Camera systems can be designed as either a single unit containing all the components or as a system that uses separate components (such as a camera module with an external image processing unit).

4.1.1 Image Acquisition

Images are captured by one or more camera modules in the vehicle. The image data of the camera module is influenced by the image acquisition control and then passed to the image processing unit. Also, additional image (pre-) processing functions can be integrated directly onto the image sensor. This is called a *system on chip (SOC)* sensor. However, it must be analyzed whether the increased demand for power and associated heat generation (and thus increased image sensor noise) make a high integration meaningful.

In the case of preprocessing on the image sensor or in the case of camera modules with an external processing unit, the image data is transferred via special interfaces based on the *low-voltage differential signaling (LVDS)* standard or via Ethernet in compressed form (e.g., as Mjpeg or H.264 video).

4.1.2 Image Acquisition Control

The control of the camera image sensor is necessary to set the optimal image parameters in all environmental situations. The two main control systems are the exposure control and the white balance.

To adapt to different lighting situations, the exposure control is designed so that in dark areas, structures can be detected and that bright areas are not saturated. Of course this requires large dynamic range of the camera module.

The control of the white balance has the goal to achieve consistent colors even with different color temperatures of the light source of the scene illumination. For example, white pavement markings must be recognized as white both in daylight and in tunnels with, e.g., yellowish light.

4.1.3 Image Processing

In the image preprocessing, the preparation of the camera image and the first processing steps take place. When using an RGB image sensor, a color image is reconstructed in a so-called *demosaicing* process. As a further step, a gamma correction takes places, i.e., input values in the image are transferred through a transformation step in different output values. Background of the operation is either an adaption to a particular display system or the improved image processing. In the case of a representation on a display, a noise reduction, an edge enhancement, and color correction are typically performed (Reinhard et al. 2008; Nakamura 2006).

If the image data is used for machine vision tasks, often a distortion correction, a calculation of the optical flow, and, in the case of a stereo camera, an image rectification and a creation of the disparity map are performed. From the preprocessed images, the desired information is extracted in the main image processing step (see also ► Chaps. 20, “Fundamentals of Machine Vision,” and ► 21, “Stereovision for ADAS”).

4.1.4 Communication

The data exchange with other controllers in the vehicle is realized via the communication interface of the camera system. Common vehicle bus systems are the *controller area network (CAN) bus*, the *FlexRay* bus, and the *Ethernet* standard. While CAN and FlexRay bus systems are only used to control the camera system and to transfer the output data in the form of, e.g., object lists, a transmission of raw image data is possible when using Ethernet because of the higher data rates.

4.1.5 Electronics

The design of the electronics follows the high standards in the automotive industry, among others, with regard to durability and electromagnetic immunity and compatibility. The large amounts of data that must be processed in real time using complex algorithms lead to a system design with several processors or multi-core processors (Stein et al. 2008). The resulting amount of heat to be dissipated is a challenge in the vehicle installation space and has to be considered already in the electronics and housing design.

4.1.6 Mechanics

The housing of the camera system is the interface between the electronics and the camera module to the vehicle. It must be thermally stable and easy to assemble. Moreover, the housing usually forms the shielding of the electronics for better

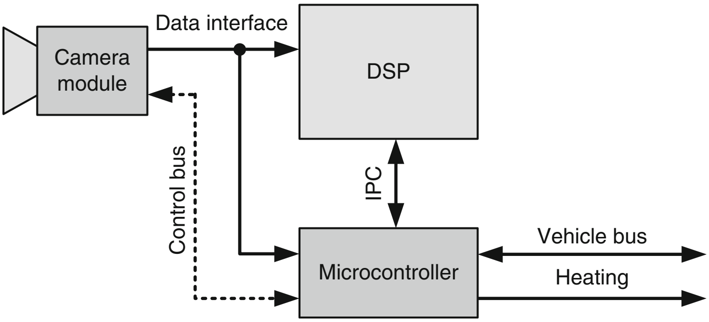


Fig. 13 Architecture of a mono camera system

electromagnetic compatibility. In the case of the front view camera, the housing is located behind the windshield. To avoid reflections between the windshield and the housing, a so-called stray light cover is often used. In camera systems where the camera modules are in direct contact with the environment (e.g., surround view), the housing must also be sealed against humidity.

4.2 Mono Camera Architecture

A typical mono camera architecture of a front view camera system is shown in Fig. 13. The camera module is controlled via a communication bus, and the image data is transferred via a parallel interface to the image processing unit. In this case, a *digital signal processor (DSP)* is implemented, which can perform video processing in real time. In other systems, *field programmable gate arrays* (FPGA) or dedicated *application-specific-integrated circuits (ASIC)* are used (Nakamura 2006). The image processing unit is supported by fast memory modules that are used for the temporary storage of processed data and in addition for storing multiple pictures in the case of the application of tracking algorithms.

The microcontroller handles the exposure control, the control of the windshield heater, the communication on the vehicle bus, and other control and monitoring functions. Communication between the microcontroller and DSP takes place through an *interprocess communication (IPC)* interface.

4.3 Stereo Camera Architecture

The main differences between a stereo camera system and a mono camera system are an additional camera module and further processing units. The complex stereo functions lead to a significantly higher requirement on processing power.

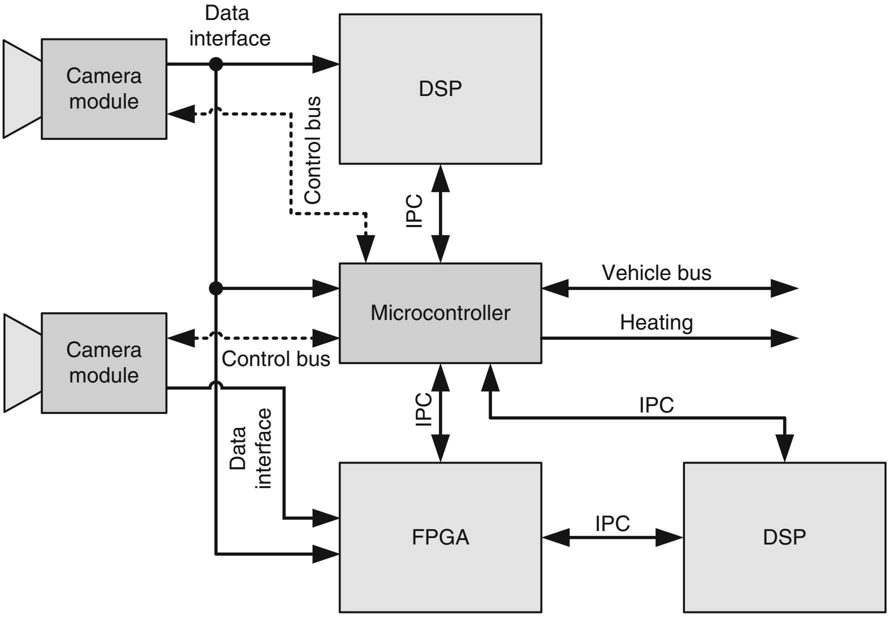


Fig. 14 Architecture of a stereo camera system

4.3.1 Structure of a Stereo Camera

The basic structure of a stereo camera is shown in Fig. 14. The two image sensors are controlled by a single microcontroller. The image signals are transferred for the preprocessing steps to a DSP and an FPGA. The FPGA is used for the rectification, calculation of the disparity map, and calculation of the optical flow. The object formation is done on the DSP, i.e., identification of pedestrians, vehicles, etc.

A second DSP is used for ADAS functions like lane departure warning and traffic sign recognition. Communication with the vehicle is realized by means of a microcontroller. The image processing units are each connected to fast memory chips.

4.3.2 Differences to Mono Camera Systems

In addition to differences in the image processing, stereo camera systems have additional requirements to the system architecture. The image recording of both camera modules must be synchronized to avoid an influence by a temporal and spatial shift of the frames.

Another difference is the increased demand with respect to the assembly and calibration of the camera system. The two camera modules have to be very precisely aligned to each other in all spatial axes (pitch, yaw, and roll angles) and then mounted in the housing. Any change in the optical axis of the cameras can lead to a de-calibration of the system, resulting in failure of the function. Therefore, the housing is designed to be stable even in varying environmental conditions, and different calibration procedures are carried out during assembly (see Sect. 5).

4.3.3 Design

An important criterion for the use of a stereo camera system is the accuracy of the depth estimation. Only with precise depth estimation, functions such as a camera-based emergency brake assist or an automatic distance control system can be realized reliably. These functions rely on precise distance information from the vehicle to the object.

Important parameters in a stereo system are the paraxial focal length of the cameras f and the distance between the cameras (base width b). The so-called *disparity* d is derived from the distance of the pixels defined by the projection of lines from the image point P incident on the image sensors.

With the size of a pixel on the image sensor s_P, a distance to the object z_C can be calculated:

$$z_C = \frac{b \cdot f}{d \cdot s_P}$$

After derivation, the absolute distance error is

$$\Delta z = \frac{\Delta d \cdot s_P \cdot z_C^2}{b \cdot f}$$

Accuracies of Δd of less than one pixel can be achieved to minimize the error of the disparity estimation.

An example calculation of a camera system with a focal length of 5 mm, a pixel size of 3.75 μm, a base width of 200 mm, and a disparity error of 0.25 pixels results in an error of the distance measurement of 2.3 m (4.7 %) at an object distance of 50 m.

Improvements are possible by increasing the resolution of the image sensor using a smaller pixel size, a larger base width, or a larger focal length. Changing these parameters requires, however, a change in the system design. The field of view of the camera would be limited by a longer focal length. The base width should be kept as small as possible because of the requirements of the vehicle design. A smaller pixel size can lead to a lower sensitivity of the system. In addition, the required computing power increases significantly with higher resolutions.

5 Calibration

To ensure that driver assistance functions like sign recognition, lane keeping assist, or head light control work in a reliable and correct manner, it is important to interpret images of the used camera system correctly. To this end, additional information about the images is necessary which can be determined by calibration algorithms. Such information, which we denote as calibration parameters, cannot only be used for better interpretation of the images but also for a compensation of deviations from a defined norm. One example might be the linearization of an image sensor's response curve.

This section provides a discussion about the calibration parameters that are typically determined for driver assistance systems. Furthermore, the section describes where a calibration normally takes place and how it can be conducted.

5.1 Calibration Parameters

Calibration parameters are variables of a model which is used to precisely describe the camera-based image acquisition system. For a specific system, the values of these parameters are found by using a calibration algorithm. One can distinguish between different classes of parameters, depending on the characteristics of the system to be modeled. The following list gives an overview on the most relevant parameters in the automotive field and their grouping:

5.1.1 Characterization of the Camera Module

- OECF – *opto-electronic conversion function*, image sensor response curve
- Noise due to dark current, defect pixels

5.1.2 Geometric Camera Calibration

Intrinsic camera parameters:

- Principal point
- Focal length
- Image distortion
- Pixel scale factors

Extrinsic camera parameters:

- Camera position
- Camera orientation

5.2 Calibration Environments and Calibration Procedures

Where and when a calibration is performed highly depends on the calibration parameters to be determined and thus the resulting requirements. In principle, a camera-based driver assistance system is calibrated already during the production process with respect to sensor characterization and intrinsic parameters. For doing this, target-based calibration setups are installed in production environments which enable measurements with a high reproducibility.

5.2.1 Characterization of the Camera Module: Camera Production Line

To determine the OECF, images with different but defined illumination intensities and constant exposure times are needed. Alternatively, the illumination intensity on the sensor remains constant, but the exposure time varies. Both ways have the same effect – the irradiation energy captured by the sensor is modified in a defined way such that the sensor response can be evaluated and plotted against the irradiation energy. With such a response curve, the dynamic behavior of the sensor is known. Potential deviations from the desired behavior can be compensated, and hence, fluctuations in the product quality are balanced out.

5.2.2 Intrinsic Camera Calibration: Camera Production Line

The calibration procedure during camera production can additionally be used to determine intrinsic parameters. Typically, a three-dimensional setup with checkerboard targets is used. With the known setup geometry and the positions of the checkerboard corners in the acquired images, the parameters of the model to describe the camera can be estimated. In most of the cases, a simple pinhole camera model plus a model for the lens distortion of low order is already sufficient. A well-known approach of determining these intrinsic (and additionally extrinsic) parameters is, e.g., the method described by Tsai (1987).

In the case of a stereo camera module, the intrinsic parameters of both cameras are estimated in the production plant. Additionally, the position and orientation of the cameras to each other are computed.

5.2.3 Extrinsic Camera Calibration: Production Line of the Vehicle

During its production, the camera is not mounted at its final position in the vehicle. Hence, a calibration of the extrinsic parameters cannot be done at this point – but it can be done in the production line of the vehicle, as soon as the camera is mounted. With a simple target which is placed at a known position with respect to the camera, the viewing orientation of the camera can be determined. The position of the camera is derived from construction data or it is measured by using external devices.

5.2.4 Extrinsic Camera Calibration: During Vehicle Operation (Online)

The step of extrinsic camera calibration in the production line of the vehicle is not always practical for vehicle manufacturers. In addition, the camera orientation to world is not always constant over the lifetime of the vehicle. Different states of vehicle loading can have influence on the camera orientation; but the extrinsic parameters need to be precise also in these cases. Hence, the orientation of the camera is being determined also during vehicle operation. Depending on the capture range of the used calibration procedure and the mounting tolerances of the camera module, the calibration in the production line of the vehicle can thus be omitted.

To determine such a calibration online, several approaches exist. Of big advantage is to use results of driver assistance functions which are running on the system anyway and to derive calibration parameters from their outputs. An example to calibrate by using such a “structure from motion” approach is described in Civera et al. (2009).

5.2.5 Stereo Camera Calibration: During Vehicle Operation (Oline)

Stereo cameras additionally need a particularly precise calibration of the camera orientation and position to each other. Although these parameters have been determined during camera production, there is a need to adjust them during normal camera operation. Mechanical influences and temperature changes could have a severe negative effect on the measurement accuracy otherwise. In addition to the aforementioned calibration procedures for mono cameras, approaches especially tailored to stereo cameras can be used. These methods tune the matrix used for the transformation from left camera image to right camera image iteratively in a way that feature correspondences which are extracted from the images meet the so-called epipolar constraint (Zhang 1998).

6 Outlook

The technology in the field of camera systems is developing rapidly. One reason is the area of consumer electronics with its demand for ever more powerful and less expensive image sensors, camera lenses, and computing platforms. These advances will be used in future also in the vehicle environment. Greater computing power enables both advances in image processing as well as the usage of larger resolutions of the image sensors. Higher refresh rates and improved sensitivities are additional developments in the field of camera sensors. With these technological advances, camera systems will be used in the future in various applications in the vehicle.

References

Brainard DH (1994) Bayesian method for reconstructing color images from trichromatic samples. In: Proceedings of the IS&T 47th annual meeting, Rochester, NY, pp. 375–380

Baxter D (2013) A line based HDR sensor simulator for motion artifact prediction. Proc SPIE 8653:86530F

Civera J, Bueno DR, Davison AJ, Montiel JMM (2009) Camera self-calibration for sequential bayesian structure from motion. In: Proceedings of the 2009 I.E. international conference on Robotics and Automation, Kobe, Japan

Daimler (2014) Homepage Daimler. http://www.mercedes-benz.com. Accessed 10 Jan 2014

Darmont A (2012) High dynamic range imaging, sensors and architectures. SPIE Press, Washington

Delphi (2014) Homepage Delphi. http://delphi.com/manufacturers/auto/safety/active/racam/. Accessed 10 Jan 2014

El Gamal A, Eltoukhy H (2005) CMOS image sensors. IEEE Circuits Devices Mag 21(3):6

Fiete RD (2010) Modelling the imaging chain of digital cameras. SPIE Press, Bellingham

Fischer RE (2008) Optical system design. McGraw-Hill, New York

Gentex (2014) Homepage Gentex. http://www.gentex.com/automotive/products/forward-driving-assist. Accessed 10 Jan 2014

Hecht E (1998) Optics. Addison Wesley Longman, New York

Hertel D (2010) Extended use of incremental signal-to-noise ratio as reliability criterion for multiple-slope wide-dynamic-range image capture. J Electron Imaging 19(1):011007

Holst GC, Lomheim TS (2011) CMOS/CCD sensors and camera systems. SPIE Press, Washington

ISO 26262: Road vehicles – functional safety

ISO/DIS 16505: Road vehicles – ergonomic and performance aspects of camera-monitor systems – requirements and test procedures

Källhammer J (2006) Night vision: requirements and possible roadmap for FIR and NIR systems. Proc SPIE 6198:61980F

Loce RP, Berna I, Wu W, Bala R (2013) Computer vision in roadway transportation systems: a survey. J Electron Imaging 22(4):041121

Miller JWY, Murphey YL, Khairallah F (2004) Camera performance considerations for automotive applications. Proc SPIE 5265:163

MIPI (2014) Homepage MIPI-Alliance. http://www.mipi.org/specifications/camera-interface. Accessed 10 Jan 2014

Nakamura J (2006) Image sensors and signal processing for digital still cameras. CRC Press, Boca Raton

Raphael E, Kiefer R, Reisman P, Hayon G (2011) Development of a camera-based forward collision alert system. SAE Int J Passenger Cars Mech Syst 4(1):467

Reinhard E, Khan EA, Akyüz AO, Johnson GM (2008) Color imaging. A.K. Peters, Wellesley

Sinha PK (2012) Image acquisition and preprocessing for machine vision systems. SPIE Press, Washington

Solhusvik J, Yaghmai S, Kimmels A, Stephansen C, Storm A, Olsson J, Rosnes A, Martinussen T, Willassen T, Pahr PO, Eikedal S, Shaw S, Bhamra R, Velichko S, Pates D, Datar S, Smith S, Jiang L, Wing D, Chilumula A (2009) A 1280 × 960 3.75um pixel CMOS imager with triple exposure HDR. In: Proceedings of 2009 international image sensor workshop, Utah, USA

Solhusvik J, Kuang J, Lin Z, Manabe S, Lyu J, Rhodes H (2013) A comparison of high dynamic range CIS technologies for automotive applications. In: Proceedings of 2013 international image sensor workshop, Utah, USA

Stein GP, Gat I, Hayon G (2008) Challenges and solutions for bundling multiple DAS applications on a single hardware platform. Israel computer vision day

Theuwissen AJP (2008) Course "digital camera systems" – hand out. CEI.se, Finspong

Tsai RY (1987) A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J Robot Autom 3(4):323

Yadid-Pecht O, Etienne-Cummings R (2004) CMOS imagers: from phototransduction to image processing. Kluwer, Dordrecht

Zhang Z (1998) Determining the epipolar geometry and its uncertainty: a review. Int J Comput Vis 27(2):161

Fundamentals of Machine Vision

20

Christoph Stiller, Alexander Bachmann, and Andreas Geiger

Contents

C. Stiller (✉)
Institut für Mess- und Regelungstechnik, Karlsruher Institut für Technologie (KIT), Karlsruhe, Germany
e-mail: stiller@kit.edu

A. Bachmann
ADC Automotive Distance Control Systems GmbH, Lindau, Germany
e-mail: alexander.bachmann@continental-corporation.com

A. Geiger
MPI for Intelligent Systems, Tübingen, Germany
e-mail: andreas.geiger@tue.mpg.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_21

Abstract
Automobiles may acquire a rich variety of relevant information from image data and its analysis using machine vision techniques. This chapter provides an overview on the principles underlying image formation and image analysis. The perspective projection model is formulated to describe the mapping of the 3D real world onto the 2D image plane with its intrinsic and extrinsic calibration parameters. Image analysis typically begins with the identification of features. These may describe locations of particular local intensity patterns in a single image, such as edges or corners, or may quantify the 2D displacement of corresponding pixels between two images acquired at different time instances or by a multicamera system. Such features can be used to reconstruct the 3D geometry of the real world using stereo vision, motion stereo, or multiview reconstruction. Temporal tracking using Bayesian filters and its variations not only improves accuracy but readily allows for information fusion with data of other sensors. The chapter closes with two application examples. The first addresses object detection and tracking using multiple image features. The second application focuses on intersection understanding illustrating the large potential of high-level scene interpretation through machine vision.

The vast majority of creatures able to navigate through space strongly rely on its visual system for this task. In particular, it is well known that humans perceive about 90 % of the information required for driving visually. This allows the conclusion that sufficient information is available in the visual domain. Furthermore, it can be expected that vision-based driver assistance systems exhibit a fairly transparent behavior, e.g., drivers are prepared for cautious driving at decreased velocity under bad weather conditions. Finally, vision allows the perception of a multitude of different information relevant for vehicle control. Much of this information has been designed for visual perception and is hardly recognizable by any other technology, such as lane markings or traffic signs. Thus, it comes as no surprise that cameras are indispensable sensors in driver assistance and automated driving.

A camera projects the three-dimensional (3D) real world onto a two-dimensional (2D) imager. Hence, image acquisition reduces the available information by an entire dimension. While 2D information suffices in several tasks, such as the classification of objects, the majority of driver assistance functions requires 3D perception of the vehicle's environment. This holds particularly true for safety-critical functions including any kind of longitudinal and lateral control. Hence, the scope of image analysis techniques includes 3D reconstruction of the scene geometry and dynamics. Buoyed by the continuing decline in prices of camera and processing hardware on the one hand and the rich information extractable from image sequences on the other, image sensors are used in a constantly growing number of applications.

Compared to the seamless perception performed by biological vision systems, machine vision is still in its infancy with application limited to narrowly defined domains. Even with prior domain knowledge, machine vision is currently far inferior to human visual performance. This chapter provides an overview of basic methods of image interpretation, as well as the potential and the limitations of image sensors. The theoretical foundations are thereby illustrated by numerous practical examples.

1 Image Formation

1.1 Perspective Projection

The projection of most cameras can be described by the pinhole model shown in Fig. 1. The aperture, i.e., the pinhole, is assumed so small that a sharp image is attained in the plane of the imager. In practice, the pinhole is replaced by a lens, which yields a brighter image.

Mathematically, the projection expresses the mapping of a point in 3D space onto the 2D image plane. The origin of 3D camera coordinates $\mathbf{X} = (X, Y, Z)^T$ is placed in the aperture and is referred to as optical center. The Z-axis, hereinafter also referred to as optical axis, is perpendicular to the image plane whose X- and Y-axis are oriented along the row and column direction of the imager, respectively. Introduction of a virtual image coordinate system oriented parallel to the imager plane and located at a distance 1 before the pinhole yields a mathematically elegant projection model. The image in this plane is different from the real camera image only by a scaling with the negative focal length f, so that the image is no longer rotated by 180°. The image in this virtual plane with image coordinates $\mathbf{x} = (x, y)^T$ is termed as the image of a calibrated camera. The projection equation relating camera to image coordinates can thus be written as

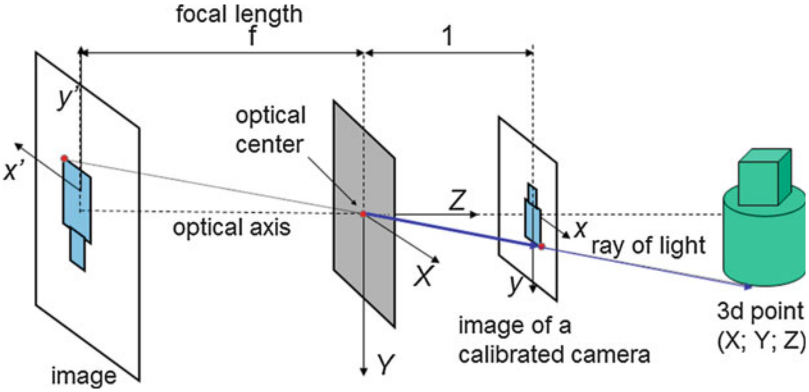


Fig. 1 Pinhole model for perspective projection

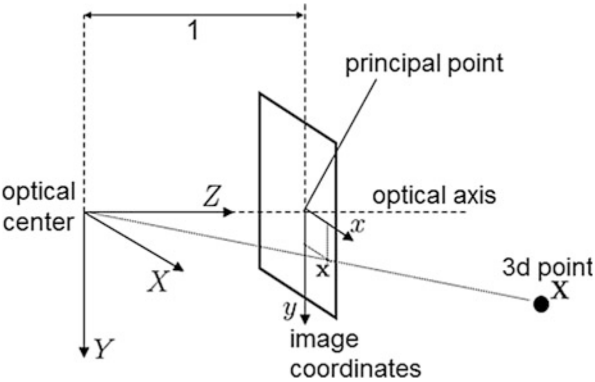


Fig. 2 Geometric camera model with perspective projection

$$\lambda \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad \text{for a} \quad \lambda \in \mathrm{IR}. \tag{1}$$

The most important consequence of this equation is that a camera can determine only the direction of a point in 3D space. Absolute distances can only be recovered when the scale factor λ can be determined. Cameras are therefore said to be scale-blind or bearings-only sensors. In the context of automobiles, knowledge of the installation height of the camera or of the distance between the two optical centers in a stereo arrangement may be employed for scale reconstruction.

Introduction of homogeneous coordinates $\tilde{\mathbf{x}} = (x \quad y \quad 1)^{\mathrm{T}}$ reduces the projection equation to almost deceptive simplicity

$$\tilde{\mathbf{x}} \cong \mathbf{X}, \tag{2}$$

where equality "$\cong$" in homogeneous coordinates means that a nonzero real number λ exists, such that $\lambda\tilde{\mathbf{x}} = \mathbf{X}$. The resulting geometric camera model is shown in Fig. 2.

Image processing commonly uses computer coordinates $\mathbf{x}_R = (x_R, y_R)^{\mathrm{T}}$. The origin of these is placed in the upper left corner of the image, and the scaling is such that the distance between adjacent pixels is 1, i.e., all pixels have integer computer coordinates (Fig. 3). Denoting the pixel spacing in horizontal or vertical direction on the imager with Δx, Δy, computer coordinates and image coordinates are related through a scaling by

$$f_x = \frac{f}{\Delta x}; \; f_y = \frac{f}{\Delta y}; \tag{3}$$

and a shift by $(x_0, y_0)^{\mathrm{T}}$. Such a mapping is linear in homogeneous coordinates, i.e., with $\tilde{\mathbf{x}}_R = (x_R, y_R, 1)^{\mathrm{T}}$ one can write

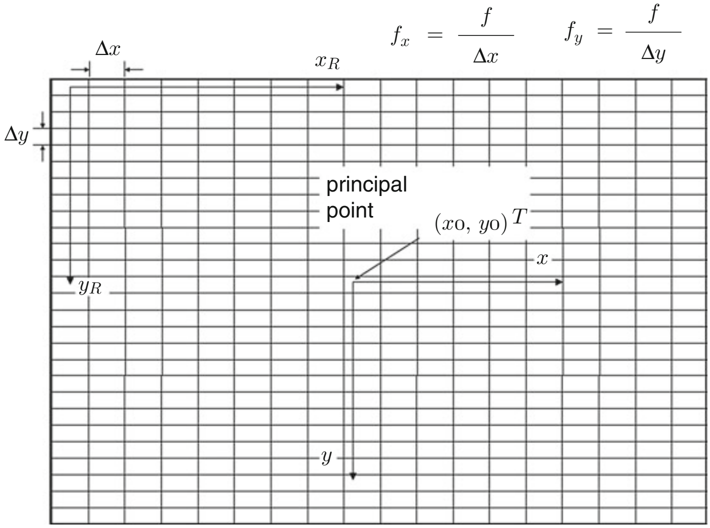


Fig. 3 Intrinsic camera parameters

$$\tilde{\mathbf{x}}_R = \mathbf{C}\tilde{\mathbf{x}} \quad \text{with} \quad \mathbf{C} = \begin{pmatrix} f_x & 0 & x_0 \\ 0 & f_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4}$$

where $\mathbf{C}$ is referred to as calibration matrix incorporating the intrinsic camera parameters, namely, the principal point $(x_0, y_0)^{\mathrm{T}}$ and the focal lengths f_x, f_y.

Finally, the world coordinate system may not be aligned with the camera but be rotated with the rotation matrix $\mathbf{R}$ and shifted by the translation vector $\mathbf{t}$, i.e., $\mathbf{X} = \mathbf{R}\mathbf{X}_W + \mathbf{t}$. In homogeneous coordinates, $\tilde{\mathbf{X}} = (X, Y, Z, 1)^{\mathrm{T}}$, $\tilde{\mathbf{X}}_W = (X_W, Y_W, Z_W, 1)^{\mathrm{T}}$, this equation is also linear

$$\tilde{\mathbf{X}} \cong \tilde{\mathbf{M}}\tilde{\mathbf{X}}_W \quad \text{mit } \tilde{\mathbf{M}} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{pmatrix} \tag{5}$$

wherein the extrinsic calibration matrix $\tilde{\mathbf{M}}$ includes six degrees of freedom of a rigid movement in 3D space. In summary, the projection of a point in 3D world coordinates $\mathbf{X}_W$ onto the 2D computer coordinates $\mathbf{x}_R$ is given as

$$\tilde{\mathbf{x}}_R \cong \mathbf{P}\tilde{\mathbf{X}}_W \quad \text{with} \quad \mathbf{P} = \mathbf{C}\mathbf{M}, \tag{6}$$

wherein $\mathbf{M} = (\mathbf{R} \quad \mathbf{t})$ comprises the upper three rows of the extrinsic calibration matrix $\tilde{\mathbf{M}}$. $\mathbf{P}$ thus represents a 3 × 4 matrix, called projection matrix (Faugeras 1993). The unequal dimension of the matrix in row and column direction illustrates the loss of information caused by perspective projection.

1.2 Image Representation

While the projection described in the previous section produces a signal that is continuous in time, space, and amplitude, images are digitized by sampling and quantization. The spatial discretization is performed by the pixel grid of the imager. Since natural images include unlimited spatial frequencies, the sampling theorem is thereby violated. However, the sensitive surface of the individual pixels, the lens system, and often the A/D converter possess low pass characteristics, so that aliasing effects can be largely suppressed.

In general, gray values are quantized linearly with 8 bits. However, higher-intensity dynamics are desirable in the automotive field. Several cameras have been designed for driver assistance applications with nonlinear characteristics or with 14-bit linear quantization. Some applications would even benefit from further information, e.g., color, associated with each pixel. However, such cameras have not yet been enforced in the automotive sector.

Moore’s law, stating that computing power doubles approximately every 2 years, seems to also apply to the number of pixels on image sensors. Accordingly, the cost ratio between camera and control unit would remain approximately constant. Already in today’s automobiles, cameras produce the highest data stream. Even a moderate monochrome VGA image signal with 640 × 480 pixels, a frame rate of 25 Hz, and 8 bit illumination quantization generates a data rate of about 60 Mbit/sec. The trend is clearly in the direction of megapixel cameras. Due to the lack of any physical limitations, it is reasonable to expect that camera resolution will catch up with the some 120 Mega receptors on the human retina in less than 10 years. Beside programmable processors, this growing amount of data will be evaluated employing digital logic blocks and other highly parallel structures.

2 Image Processing

The term image processing refers to the preparation, analysis, and interpretation of visual information. As the complexity of high-level image processing tasks (such as lane detection and object recognition) grows with the amount of input data, it is common to preprocess the input images using image processing and feature extraction techniques which will be discussed in this section. Preprocessing the input images can reduce noise and defects arising during the image capturing process and prepares the data for the targeted application. The image signal is typically manipulated by a wide variety of filter operations or by means of basic image transformations with the goal of eliminating most of the irrelevant or distracting information. After a successful preprocessing of the image data, relevant features can be extracted and passed to the respective higher-level process. Sect. 2.1 presents a selection of image features used in modern driver assistance systems today.

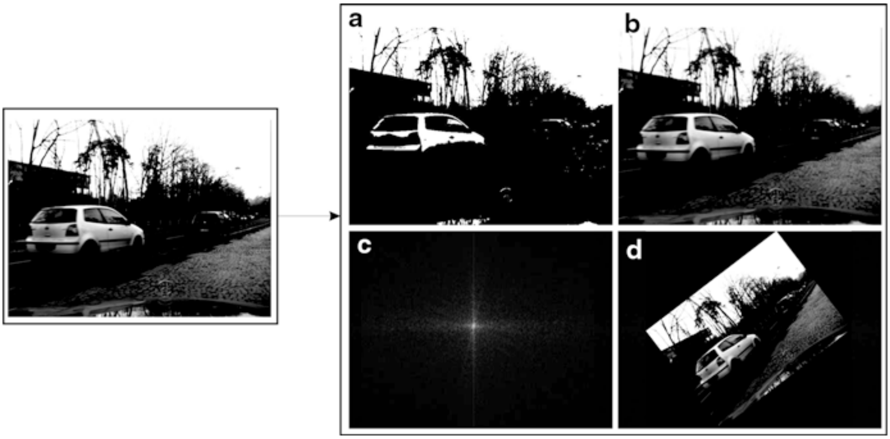


Fig. 4 Examples of image operators. *Left*: Input image; *Right*: (**a**) Binarization (*point operator*); (**b**) Binomial filter (*local Operator*); (**c**) Fourier transform (*global Operator*); (**d**) Rotation (*geometric Operator*)

2.1 Image Enhancement

Image enhancement allows for filtering the relevant information from the recorded image signal in order to pass this information to the subsequent processing units (Fig. 4). Image processing operations can be divided into three classes: point operators, local operators, and global operators, depending on the number of pixels that affect an operation.

Point operators are used to independently manipulate individual gray or color values in an image. Examples include histogram stretching, equalization, or thresholding operators. Point operators act on individual pixels and can be applied independently for each pixel.

Local operators calculate the color or gray value of a pixel based on a local region around the pixel of interest. They are also called proximity operators or FIR filter (finite impulse response filter) and directly operate on the image signal. Examples of local operators include morphological filters, smoothing operators, and gradient filters for extracting texture information from the image. An important local image operator is the smoothing filter. The two most popular smoothing filters are the Gaussian and the Binomial filter. For the lowest order, the latter is specified by the simple convolution kernel $2^{-1}[1 \quad 1]$. Repeatedly convolving the input image with this kernel results in higher-order Binomial filter operations. As an example, the Binomial filter of order seven is specified by $128^{-1}[1 \quad 7 \quad 21 \quad 35 \quad 35 \quad 21 \quad 7 \quad 1]$. The higher the order, the more the Binomial filter shape approximates the Gaussian blur filter and consequently can be used as a substitute. As both the Binomial as well as the Gaussian filter kernels are separable, 2D convolution is not required in order to apply these filters. Instead, the filter operation decomposes into a horizontal filter operation followed by a vertical filter operation with a 1D filter applied to every row/column of the image. A more detailed discussion on filtering with Gaussian and Binomial kernels can be found in Gonzalez and Woods (2006).

Global operators such as the Fourier transformation or the Hough transformation take the whole image into account when transforming the value of each pixel. An important subclass are geometric operators which geometrically manipulate the input image. Typical examples are rotation, scaling, and mirroring operators.

2.2 Feature Extraction

After preprocessing the input image using the image enhancement techniques described above, a variety of image features can be extracted. Features are locally constraint, expressive parts of an image which allow for a symbolic or empiric description of the image properties or the object. Image features can be extracted from a variety of different image primitives. For example, the image gradient can be considered a feature for object contours or boundaries. Object-specific distributions of some measurable image quantity provide evidence for image regions containing the object. The goal of the feature extraction process is to extract structural properties from the vast amount of image information and represent it compactly in terms of a low-dimensional feature vector while suppressing the influence of information irrelevant to the application of interest. This procedure leads to an enormous data reduction, analogous to the human perception of visual information where receptor fields in the retina are responsible for detecting edges, motion, or local maxima in the captured light field.

Features can be roughly categorized into image features and correspondence features: While image features can be directly estimated from the gray value pattern of a single image, correspondence features relate the projection of a single 3D point to several images.

2.2.1 Image Features

The most important image features are edges and corners. Both features are characterized by a significant change of the image signal with respect to the image coordinates. Formally, the change of the image signal is described by the image gradient. The most important algorithms for feature extraction thus often directly operate in the gradient domain. In the following, we will first discuss the extraction of the image gradient by means of local image operators. Further, we will present two of the most commonly used algorithms for edge and corner detection. Corners and edges contain the most relevant image information: it is well known that humans can extract the image content in most cases by looking at image sketches which only reveal the edges of an object.

The gradient of an image $g(x, y)$ is defined by

$$\nabla g(x,y) = \begin{bmatrix} g_x(x,y) \\ g_y(x,y) \end{bmatrix} = \begin{bmatrix} \dfrac{\partial g(x,y)}{\partial x} \\ \dfrac{\partial g(x,y)}{\partial y} \end{bmatrix} \tag{7}$$

As the image is available only in its spatially discretized version, one usually approximates the partial derivatives via FIR filters with a small number of filter coefficients. A common approximation is given by $\frac{\partial g(x,y)}{\partial x} \approx \frac{g(x+1,y)-g(x-1,y)}{2}$ which corresponds to convolving the input image with the filter kernel $2^{-1}[1 \quad 0 \quad -1]$. The Sobel operator makes use of this approximation to calculate derivatives but in addition uses the Gaussian blur kernel $4^{-1}[1 \quad 2 \quad 1]^{\mathrm{T}}$ in direction perpendicular to the gradient in order to reduce the effects of noise in the input signal. Both operations can be conducted in one step using a two-dimensional convolution with the 3 × 3 filter mask

$$\frac{\partial g(x,y)}{\partial x} \approx g(x,y) ** \frac{1}{8} \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{8}$$

The gradient image serves as input to an edge detector which extracts locally coherent line segments corresponding to strong gradient responses. One popular technique is the Canny edge detection (Canny 1986) algorithm, which we briefly describe in the following (Fig. 5). First, the input image is smoothed in order to reduce sporadic noise in the gradient image leading to wrong edge detections. Toward this goal, the input image is convolved using the previously described Binomial filter kernel. The next step consists of calculating the gradient in the smoothed image and extracting the gradient direction α as the direction of a hypothetical edge at every pixel in the image:

$$\alpha = \arctan\left(\frac{g_y(x,y)}{g_x(x,y)}\right) \tag{9}$$

After quantizing the gradient direction into 45° bins, a measure of the local edge strength is calculated. A popular choice is the absolute value of the gradient. Nearby

Fig. 5 *Left*: Original image; *Right*: Result using (**a**) Canny edge detection; (**b**) Harris-corner detection

edge pixels with strong gradient are reduced to the respective maxima using nonmaximum suppression of adjacent pixels orthogonal to the edge direction. This ensures that edges will be localized precisely as edges are "thinned" to a width of 1 pixel using this operation. The edge strength of the remaining pixels is compared against two thresholds. In case the edge strength is larger than the upper threshold, the pixel is marked as edge pixel. For pixels with edge strength lower than the lower threshold, the edge hypothesis is discarded. If the edge strength is in between both thresholds, a pixel is marked as edge if another pixel already marked as edge is located in the direction of the hypothetical edge at that pixel. This process realizes a hysteresis which leads to a more stable detection of continuous edge segments. The extracted edges can serve as input to subsequent processing steps for analyzing the detected edges with respect to their form or other criteria.

Another important image feature is an image corner. In the following, we will describe one of the most popular detectors for corners, the Harris corner detector. First, the gradient of the input image is extracted, followed by the calculation of the local gray value structure tensor

$$\mathbf{S}(x,y)=\sum_u\sum_v w(u,v)\begin{bmatrix}(g_x(x+u,y+v))^2 & g_x(x+u,y+v)g_y(x+u,y+v)\\ g_x(x+u,y+v)g_y(x+u,y+v) & \left(g_y(x+u,y+v)\right)^2\end{bmatrix} \tag{10}$$

where $w(u, v)$ denotes a weight function centered at (0, 0). The eigenvector corresponding to the larger eigenvalue of the structure tensor points in the direction of the local image orientation, and the eigenvalue can be considered a measure of the gradient strength. The eigenvector corresponding to the smaller eigenvalue faces in the orthogonal direction (the direction of the minimal orientation), and consequently the eigenvalue can be considered a measure for the strength of the orientation in this direction. Thus, corners are detected as pixels whose structure tensor exhibits two large eigenvalues. The calculation of the corner strength is typically performed directly on the structure tensor using the formula $E(x,y)=\det(S(x,y))-\kappa\cdot\left(\mathrm{spur}(S(x,y))\right)^2$ where $\kappa\in[0,04;\ 0,15]$ and pixels with a strong corner measure are detected as corners. Alternatively, the eigenvalue decomposition of $\mathbf{S}(x, y)$ can be leveraged, and the smaller eigenvalue can be compared against a threshold. The extracted corners can be further used in order to find correspondences between two images. A more detailed discussion of edge detection and the structure tensor can be found in Gonzalez and Woods (2006) and Harris and Stephens (1988).

2.2.2 Correspondence Features

Knowledge about the projection of a 3D point into different images (captured by one or several cameras) allows to draw conclusions about the position of that point in 3D. Thus, finding correspondences in several images is essential to reconstruct 3D information which has been lost during the imaging process, i.e., the perspective projection onto the image plane. Finding correspondences can be interpreted as a search problem where for an element in one view the corresponding element in the other view is sought (Stiller and Konrad 1999; Trucco and Verri 1998). Search algorithms which apply to this problem can be roughly categorized into the following three classes: descriptor-based approaches, gradient methods, and matching techniques.

Detecting and Describing Salient Feature Points Descriptor-based approaches such as SIFT (Lowe 2004), SURF (Bay et al. 2006), BRIEF (Calonder et al. 2012), DAISY (Tola et al. 2010), and the learned descriptor DIRD (Lategahn et al. 2013; Lategahn and Stiller 2014) typically work in three subsequent stages: First, salient points such as corners (Harris and Stephens 1988) or blobs (Lowe 2004) are detected in the image. Those distinctive points are likely redetected in other images of the same scene. After this discretization, a compact local descriptor is calculated which should be discriminative but also robust against changes in illumination, rotation, translation, change of scale, or other modest distortions which are caused by the perspective effect of the camera. Some variants (such as U-SIFT, U-SURF) exhibit only some of the mentioned invariance properties. In video sequences, for example, rotation invariance is often not required due to the temporal proximity of subsequent frames. As the manual design of feature descriptors is tedious, the DIRD descriptor makes use of a training set to learn the required properties depending on the task at hand. The last step consists of establishing correspondences between the two images using the extracted image features and descriptors. A popular criterion for the quality of a corresponding feature point is a small Euclidean distance (or a large scalar product) of the respective feature vectors. While the simplest approach for establishing correspondences consists of assigning them in decreasing order of the associated similarity measure, globally optimal methods for bipartite graph matching such as the Kuhn-Munkres algorithm (also known as the “Hungarian method“) can be leveraged as well. In the following, we will discuss the SIFT descriptor in more detail. More recent descriptors such as SURF, BRIEF, DAISY, or DIRD follow a similar principle but can be more easily accelerated via efficient calculations and approximation schemes.

The SIFT approach first establishes an image pyramid via smoothing and downsampling of the input image. Next, each scaled image version is independently convolved with a difference-of-Gaussians (DoG) filter, and minima/maxima in the resulting score map are extracted and processed using nonmaximum suppression in order to reduce the number of feature points and enforce a minimal distance between them. To ensure rotation invariance, the gray value gradient histogram in a local window is calculated and the window for describing the feature point is rotated according to the direction of the maximal gradient.

For describing the image region, a local window (centered at the feature point) is divided into equal squares. Each of these regions can be robustly described by a histogram of gradient orientations as illustrated in Fig. 6. The histograms are normalized and concatenated to form a vector with 128 entries. This vector serves as the numerical descriptor of the feature point. In contrast to a direct comparison of

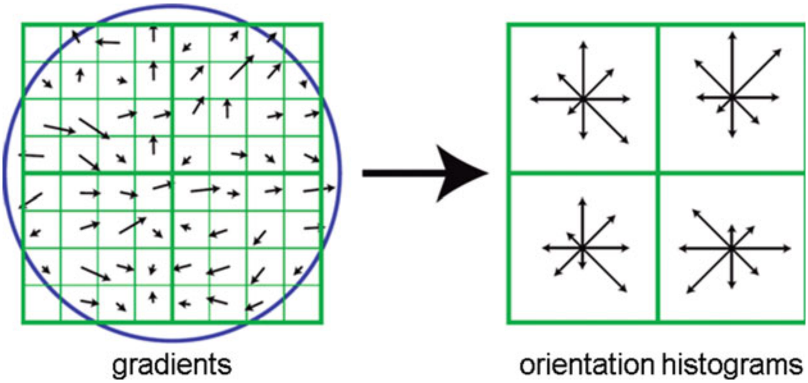


Fig. 6 SIFT descriptor. *Left*: gradients; *Right*: histograms of gradient direction

image intensities or gradients, histograms provide a higher degree of robustness as their form is relatively unaffected when slightly rotating, translating, or distorting the corresponding image region. This property is also exploited by the so-called HoG features (histogram of oriented gradients), which are a fundamental building block in modern object detection and classification systems, for example, in pedestrian detection systems.

Gradient Methods Gradient methods typically approximate the intensity of the gray values in the image, represented as a spatially and temporally varying function $g(x, y, t)$ of the image coordinates (x, y) and time t, using a Taylor series expansion. Lucas and Kanade (1981) presents a gradient-based method which relies on a first-order Taylor approximation of the intensity image in the sense of the optical flow constraint equation. The assumption is that the intensity of a 3D point projected to the image plane remains constant over time. Thus, the optical flow constraint equation can be specified as

$$\nabla g^T \mathbf{v} + g_t = 0 \tag{11}$$

where $\mathbf{v} = (u, v)^T = \left(\frac{dx}{dt}, \frac{dy}{dt}\right)^T$ denotes the optical flow, $\nabla g = \left(\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y}\right)^T$ the spatial gradient, and $g_t = \frac{dg}{dt}$ the partial derivative of the image intensity in temporal direction. Intuitively, this equation represents the fact that the image gradient vanishes in the direction of the motion trajectory. As Eq. 11 is a scalar equation with two unknown optical flow parameters, additional assumptions are required to uniquely identify the flow vector. In the simplest case, one may assume that the optical flow vector $\mathbf{v}$ remains constant for all pixels within a local image region. This assumption in combination with least-square estimation leads to the well-known KLT tracking algorithm (Shi and Tomasi 1994), which yields good results for small displacements, i.e., if the frame rates are high and the 3D motion of the object is relatively small. A disadvantage of the method is that the optical flow constraint Eq. 11 holds only for (infinitesimal) small displacements. Gradient methods are thus often used for motion estimation from image sequences where a small optical flow can be assumed. Another application is to use this technique as a subpixel-accurate postprocessing step after pixel-accurate correspondences have been established using correspondence features or matching techniques as the ones described in the following.

Matching Techniques In matching techniques, correspondences are constrained to the pixel grid of the image. In order to establish correspondences, a small region in the first image is compared against a set of regions (correspondence candidates) in the second image according to some appropriate appearance measure. As the correspondence problem cannot be solved uniquely in homogeneous regions or in the presence of periodic textures, the search space is usually constrained to distinctive image regions (feature-based correspondence matching). Often, corners or line segments are selected as distinctive features in this context. One of the most popular detectors for finding corner features, the Harris corner detector, has been discussed in the previous section.

Often, a symmetric window, such as a rectangular image block, is selected as matching region. Thus, such approaches are often referred to as block matching techniques. In the following, we will denote the set of image pixels in the reference image by $B = \{\mathbf{x}_{R1}, \cdot \cdot \cdot, \mathbf{x}_{Rn}\}$. Let us denote the reference image as $g_1(\mathbf{x})$ and the second image as $g_2(\mathbf{x})$, and let the correspondence be described by the displacement vector $\mathbf{d} = (d_1, d_2)^T$. The SAD (sum of absolute differences) distance measure describes the absolute error of two image blocks:

$$SAD(\mathbf{d}) = \sum_{\mathbf{x}_R \in B} |g_1(\mathbf{x}_R) - g_2(\mathbf{x}_R + \mathbf{d})| \tag{12}$$

The SAD measure thus corresponds to the L_1 norm of the gray value differences. Other distance measures such as the correlation coefficient can be leveraged as well. The correlation coefficient is specified by

$$\rho(\mathbf{d}) = \frac{\sum_{\mathbf{x}_R \in B} \left(g_1(\mathbf{x}_R) - \overline{b}_1\right)\left(g_2(\mathbf{x}_R + \mathbf{d}) - \overline{b}_2\right)}{\sqrt{\sum_{\mathbf{x}_R \in B} \left(g_1(\mathbf{x}_R) - \overline{b}_1\right)^2} \cdot \sqrt{\sum_{\mathbf{x}_R \in B} \left(g_2(\mathbf{x}_R + \mathbf{d}) - \overline{b}_2\right)^2}} \tag{13}$$

where $\overline{b}_i$ denotes the average gray value of the block. It is a similarity measure between two image regions which is invariant with respect to scaling or summation of a constant offset. Correspondences established by maximizing this measure are thus largely unaffected by illumination changes.

3 3D Geometric Scene Reconstruction

Many driver assistance functions require knowledge about the 3D scene geometry in the environment of the vehicle. With monoscopic cameras, 3D reconstruction is limited due to the so-called bearings-only ambiguity. This fundamental property of sensors that only measure angles nut not range is closely related to the scale-blindness of cameras. It states that such a moving sensor may determine the trajectory of another object conducting comparable movement only up to one remaining degree of freedom. Considering the example of vehicles traveling at a constant velocity, this means that the 3D geometry of a static environment can be fully reconstructed from images of a monoscopic camera with known motion. However, position and velocity of other moving vehicles cannot be fully observed with a monoscopic camera, but one indeterminate degree of freedom will remain. Although various heuristics frequently allow to infer this missing information (e.g., planar road assumption and known camera height), these hardly provide the necessary reliability for safety-critical driving functions.

In contrast, a stereo camera system instantaneously supplies the depth information for almost all image points, thus providing information about geometry and texture of a scene with a single sensor system (Greek: *stereós*: hard, bodily, spatial). A formal model of a stereo camera system that allows for a mathematical description simply uses two pinhole cameras as has been introduced in Sect. 1.1. For both cameras, the above-described intrinsic and extrinsic calibration is assumed to be known. Section 3.1 describes the basics of stereoscopy. For static scenes, the binocular images can identically be acquired as two temporally consecutive images of a moving camera. The extension of image pair analysis to multiple-view imaging with a moving camera is presented in Sect. 3.2. A generalization to three cameras, considering triplets of corresponding points, leads to the trifocal tensor in Sect. 3.3.

3.1 Stereoscopy

Stereoscopy is a passive method for 3D scene reconstruction. Two or more images of the same scene are taken from different camera positions. Knowledge of the positions that project the same scene point in two images of a calibrated stereo rig allows to reconstruct the point's 3D position in world coordinates. For the correspondence analysis itself, some methods have been discussed in the previous section already. This section is concerned with the formulation of correspondence conditions that allow an efficient stereo analysis.

Consider a stereo system consisting of two cameras with optical centers $\mathbf{C}_l$ and $\mathbf{C}_r$ that project the same scene point $\mathbf{X}_W$. The indices for left l and right r are used to distinguish the two cameras in the stereo configuration. For practical reasons, the world coordinate system is often chosen so that it coincides with one of the two camera coordinate systems. In the further course, we assume that the world coordinate system corresponds to the camera coordinate system of the right camera, i.e. the extrinsic parameters of the right camera are $\mathbf{R}_r = \mathbf{I}$, $\mathbf{t}_r = \mathbf{0}$, while the left camera's extrinsic parameters are $\mathbf{R}_l = \mathbf{R}, \mathbf{t}_l = \mathbf{t}$. The projection matrices may thus be written as

$$\begin{aligned} \mathbf{P}_r &= \mathbf{C}_r[\mathbf{I}, 0] \\ \mathbf{P}_l &= \mathbf{C}_l[\mathbf{R}, \mathbf{t}] = \mathbf{C}_l\mathbf{M} \end{aligned} \tag{14}$$

expressing a rigid transformation of the 3D world as seen from the two coordinate systems aligned with the cameras (see Fig. 7). Specifically, for any 3D point $\mathbf{X}_W$ this transformation reads

$$\tilde{\mathbf{X}}_l \cong \tilde{\mathbf{M}}\tilde{\mathbf{X}}_r \quad \text{with } \tilde{\mathbf{X}}_r = \tilde{\mathbf{X}}_W = (X_W, Y_W, Z_W, 1)^T. \tag{15}$$

The perspective projection of $\mathbf{X}_W$ to the computer coordinates $\mathbf{x}_{R,l}$, $\mathbf{x}_{R,r}$ of the left and right camera, respectively, is given as

$$\tilde{\mathbf{x}}_{R.r} \cong \mathbf{P}_r\tilde{\mathbf{X}}_W \ \ \tilde{\mathbf{x}}_{R,l} \cong \mathbf{P}_l\tilde{\mathbf{X}}_W. \tag{16}$$

The translation vector is given as the displacement of the optical centers $\mathbf{t} = \mathrm{O}_l - \mathrm{O}_r$. It is often also referred to as a stereo baseline with base width $b = |\mathbf{t}|$. It is selected as a balancing depth resolution that increases with base width and correspondence quality that decreases with base width. The latter is caused by an increasing amount of occlusions and distortion effects for large base widths. The intersections of $\mathbf{t}$ with the two image planes are called epipoles $\mathbf{e}_l$, $\mathbf{e}_r$. As can be seen in Fig. 7, the optical centers O_l, O_r, the observed 3D point $\mathbf{X}_W$, and the image points $\mathbf{x}_l$, $\mathbf{x}_r$ are all coplanar in the so-called epipolar plane. This fundamental finding is referred to as

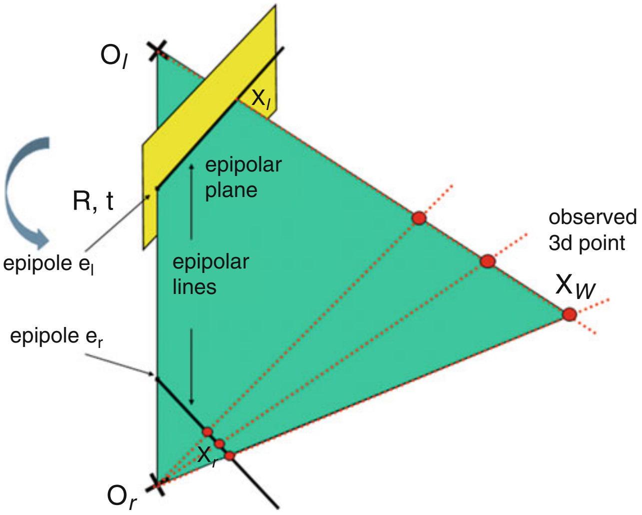


Fig. 7 Epipolar geometry of a binocular camera

epipolar constraint and reduces the search effort for stereo correspondence from the entire image plane to a ray.

The fastest way to derive the epipolar constraint mathematically is to impose that in the coordinates of the left camera $\mathbf{t}$, $\tilde{\mathbf{x}}_l$ and $\mathbf{R}\tilde{\mathbf{x}}_r$ must be coplanar vectors, which can be expressed by the constraint

$$(\mathbf{t} \times \tilde{\mathbf{x}}_l)^T \cdot \mathbf{R}\tilde{\mathbf{x}}_r = 0, \tag{17}$$

i.e., the scalar product of one vector with the cross product of the other two must vanish. Introducing the skew symmetric matrix

$$\mathbf{t}_\times = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}_\times = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}, \tag{18}$$

that satisfies $\mathbf{t} \times \mathbf{X} \; = \mathbf{t}_\times \mathbf{X}$ for any vector $\mathbf{X}$, one yields the epipolar constraint

$$\tilde{\mathbf{x}}_l^T \cdot \mathbf{E} \cdot \tilde{\mathbf{x}}_r = 0 \quad \text{with } \mathbf{E} = \mathbf{t}_\times \mathbf{R}. \tag{19}$$

$\mathbf{E}$ is referred to as essential matrix and is determined from the extrinsic calibration parameters. The essential matrix has originally been introduced by Longuet-Higgins (1981). The epipolar geometry constrains corresponding points in two images of a stereo camera to a single dimension.

In computer coordinates, the epipolar constraint reads

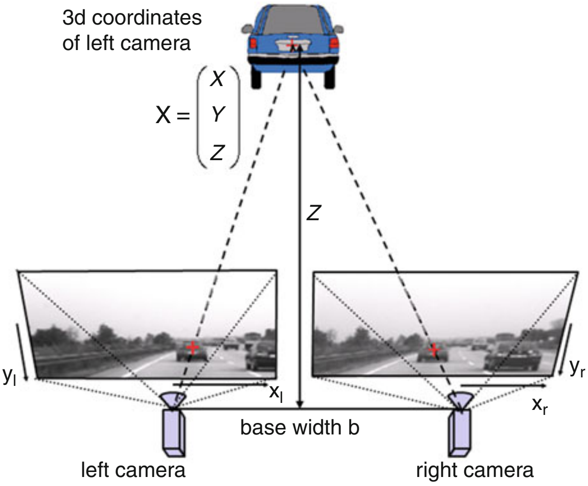
$$\tilde{\mathbf{x}}_{R,l}^T \cdot \mathbf{F} \cdot \tilde{\mathbf{x}}_{R,r} = 0 \quad \text{with } \mathbf{F} = \mathbf{C}_l^{-T} \mathbf{t}_\times \mathbf{R} \mathbf{C}_r^{-1} \tag{20}$$

where the fundamental matrix $\mathbf{F}$ includes the intrinsic and extrinsic parameters of the two cameras.

A wide variety of linear and nonlinear approaches for the estimation of $\mathbf{F}$ or $\mathbf{E}$ from given correspondence pairs has been presented in the literature. The well-known linear 8-point algorithm provides satisfactory results for a set of eight sufficiently precise point correspondences. In practice it turns out, however, that better results can be achieved by using classical nonlinear methods from numerical mathematics. Prominent nonlinear estimation techniques include the Gauss-Newton or the Levenberg-Marquardt algorithms (Faugeras 1993). The essential matrix has five degrees of freedom. These arise from three rotation angles in $\mathbf{R}$ and two components from the orientation of the translation vector $\mathbf{t}$ while its scale is irrelevant and not observable for the homogeneous matrix $\mathbf{E}$. Hence, five point correspondences are minimally required for their determination (Nistér 2004).

Equations 19 and 20 represent a constraint for corresponding pixel pairs, which reduces the search space to one dimension along the epipolar line. However, in general, the slant of the epipolar lines in the image is highly unfavorable for the search. By a homographic transformation, called rectification, the mutually twisted stereo image planes and camera coordinate systems are virtually aligned, i.e., the

Fig. 8 Rectified stereo system



epipolar lines become identical pixel rows in the two transformed images. Hence, the rectified images may be considered as taken by cameras with $\mathbf{R}_r = \mathbf{I}$, $\mathbf{t}_r = \mathbf{0}$, $\mathbf{R}_l = \mathbf{I}$, $\mathbf{t}_l = (b, 0, 0)^T$. Such a configuration is termed an axis-parallel stereo geometry or a rectified stereo system (Fig. 8). A detailed description of common rectification techniques is provided in Trucco and Verri (1998).

Since correspondences in an axis-parallel stereo system lie in the same image line, the different perspective of the two cameras results in a purely horizontal displacement in the image. This can be seen directly in image coordinates by rewriting Eq. 2 for the spatial point $\mathbf{X} = (X, Y, Z)^T$ as follows (strictly speaking, the image of a calibrated camera is considered here whose image plane has the distance f to the optical center. This additional scaling of the image coordinates with the focal length is common to express image coordinates in pixels and metric distances in meters):

$$\begin{pmatrix} x \\ y \end{pmatrix}_l = \frac{f}{Z_l}\begin{pmatrix} X_l \\ Y_l \end{pmatrix} = \frac{f}{Z_r}\begin{pmatrix} X_r + b \\ Y_r \end{pmatrix} \quad \begin{pmatrix} x \\ y \end{pmatrix}_r = \frac{f}{Z_r}\begin{pmatrix} X_r \\ Y_r \end{pmatrix} \tag{21}$$

Since the vertical coordinate y is identical in both images, the displacement is considered in horizontal direction only:

$$\Delta = x_l - x_r = \left(\frac{fX_r}{Z_r} + \frac{fb}{Z_r}\right) - \frac{fX_r}{Z_r} = \frac{fb}{Z_r} \tag{22}$$

The displacement Δ is called disparity and specified in pixels. Depth reconstruction is formed by rearranging Eqs. 21 and 22:

$$\frac{Z}{f} = \frac{b}{\Delta} \Leftrightarrow Z = \frac{bf}{\Delta} \tag{23}$$

Thus, disparity is a measure of spatial depth of a point, and eventually disparity is termed as inverse depth. For points at infinity, in particular, the disparity vanishes. By linear error propagation, the impact of small errors in the disparity estimate $d\Delta$ on the resulting error in the depth estimate can be approximated as

$$dZ = \frac{dZ}{d\Delta} d\Delta = -\frac{Z^2}{bf} d\Delta. \tag{24}$$

This fundamental finding on 3D stereo reconstruction states that the inevitable uncertainty in disparity estimates propagates to uncertainty in depth growing quadratically with depth. Accordingly, 3D reconstruction from disparity is highly accurate at close range, while it becomes unusable for long ranges.

3.2 Motion Stereo

In contrast to the classical stereo geometry in which two cameras laterally offset from one another, the so-called motion stereo method employs a single moving camera. For static environments the 3D position of the corresponding points in space can be unambiguously determined from pixel correspondences. This situation changes completely when the condition of static environments is violated. 3D reconstruction of position and velocity can then only be performed up to an ambiguity in one degree of freedom.

In the following, we consider an object-fixed coordinate system. Due to the movement of the camera, a 3D point $\mathbf{X}(t)$ changes its position over time t. Let the point move in 3D space by $\mathbf{D}(t+1) = \mathbf{X}(t+1) - \mathbf{X}(t)$ between two time instances t and $t+1$. The projected 2D displacement, also termed optical flow, in the image plane can be observed as

$$\mathbf{d}(t+1) = \mathbf{x}(t+1) - \mathbf{x}(t) = \Pi(\mathbf{X}(t+1) - \Pi(\mathbf{X}(t))) \tag{25}$$

where $\Pi(\cdot)$ denotes projection of a spatial point on the image plane, as described in Eq. 6. For a camera system whose movement in space is described by the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$, the trajectory of the point $\mathbf{X}(t)$ in space is given by

$$\mathbf{X}(t+1) = \mathbf{R}(t+1)\mathbf{X}(t) + \mathbf{t}(t+1) \tag{26}$$

The epipolar constraint Eq. 19 limits possible 2D displacement vectors of a rigid object in one dimension. Figure 9 shows the typical special case for automotive applications, in which the rotary motion is negligible, whereby the rotation matrix degenerates to the identity matrix $\mathbf{R}(t+1) = \mathbf{I}$. Furthermore, it is assumed that the camera is moved mainly only along the optical axis.

For rigid objects with unknown motion, the epipolar constraint can only be imposed after estimation of the essential matrix to restrict the search space. This has

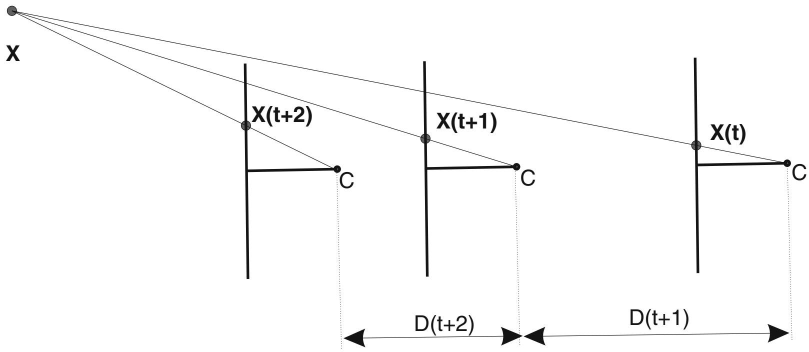


Fig. 9 Motion-stereo configuration for purely translational camera motion along the optical axis $\mathbf{t} = \left(t_x = 0, t_y = 0, t_z \neq 0\right)^T$

the consequence that corresponding points must first be sought in a 2D image area. Beyond the costly computational effort associated with this 2D search, 2D displacement estimation is often ambiguous caused by the aperture effect. Coarsely speaking, in a 2D search a unique displacement vector in a 2D search requires a corner in the image area under consideration, while an edge is sufficient for a 1D search.

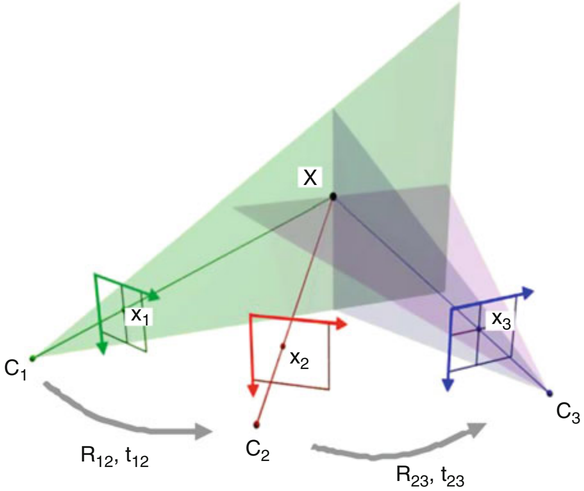
A key advantage of motion stereo compared to standard stereoscopy is that the motion stereo base length is growing with time while the base length of stereo arrays is limited for structural reasons. In motion stereo, the base length accumulates the relative motion of the camera with respect to the object over time. Hence, accuracy and range of the sensor system grows with observation time. In driver assistance applications, methods with highest reliability simultaneously evaluate disparity and optical flow. While instantaneously a 3D reconstruction at close range is possible stereoscopically, the motion stereo method accumulates motion information over time, thus yielding a high range (Dang et al. 2009).

3.3 Trifocal Tensor

The epipolar constraint Eqs.19 and 20 provide a necessary condition for correspondences in two image frames to originate from the same static point in the real world. However, it is not sufficient for multiple frames, i.e., in three or more frames, one can set up image correspondences such that any pair meets the epipolar constraint but still these points may not be projections originating from a single point of the real world.

A sufficient condition is achieved with the trifocal tensor, which describes the spatial arrangement of three cameras. The trilinear constraints composed by the trifocal tensor may be considered to impose the following (Fig. 10):

Fig. 10 Trilinear constraints



- Compliance with the epipolar constraint between the first and second image.
- Compliance with the epipolar constraint between the second and third image.
- The 3D reconstruction of any point **X** does not depend on whether the reconstruction is attained from the first or last two images.

Formally, these conditions are expressed symmetrically for all cameras, resulting in a redundant system of constraints on possible correspondences $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ in three images:

$$\mathbf{f}(\mathbf{T}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) = \mathbf{0} \tag{27}$$

The trifocal tensor **T** comprises 27 elements with 18 degrees of freedom. In practice, multiframe search of correspondences under the trilinear constraints gives superior results as compared to pairwise estimation under the epipolar constraints. Due to the redundancy incorporated in the trilinear constraints and their restriction to only three frames, however, most practical implementations prefer bundle adjustment methods that are applicable to four or more image frames (Dang et al. 2009).

4 Temporal Tracking

For a distinctive set of image features, the information loss that comes along with the projective mapping of the spatial domain into the image domain can be partially compensated by temporal tracking. Starting from the Bayes filter, which states the general probabilistic approach, two very popular approximations will be presented in this chapter, namely, the particle filter and the Kalman filter.

From a mathematical point of view, the objective of temporal tracking is to estimate the unknown quantity $\mathbf{X}_k$ given observations $\mathbf{Y}_k$. Generally, both the observations and the unknown are formulated as vectors. They change with time and are sampled at discrete time steps $k = 0, 1, 2, \cdots$. The relationship between the set of observations $\mathbf{Y}_k$ and the unknown internal state $\mathbf{X}_k$ of a system is modeled in state space. The dynamics of system state $\mathbf{X}_k$ is described by

$$\mathbf{X}_k = \mathbf{f}_k(\mathbf{X}_{k-1}, \mathbf{s}_k) \tag{28}$$

in which $\mathbf{s}_k$ states the process or system noise. The mapping from state space into observed space is made according to the observation equation

$$\mathbf{Y}_k = \mathbf{g}_k(\mathbf{X}_k, \mathbf{v}_k) \tag{29}$$

Here, $\mathbf{v}_k$ is the observation noise. By applying the recursive Bayesian estimation framework, the probability density $p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k)$ for the current system state is determined, taking into account all observations up to time k. From this probability density, an optimal estimate $\hat{\mathbf{X}}_k$ can be derived by, e.g., choosing the realization with maximum probability.

4.1 Bayes Filter

The basic principle of the approach presented here is Bayes' theorem, which is stated mathematically as the conditional probability

$$p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k) = \frac{p(\mathbf{Y}_k|\mathbf{X}_k, \mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1})p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1})}{p(\mathbf{Y}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1})} \tag{30}$$

Assuming that the current observation $\mathbf{Y}_k$ at a given state $\mathbf{X}_k$ is independent of all previous observations, one can write

$$p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k) = c \cdot p(\mathbf{Y}_k|\mathbf{X}_k) \cdot p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) \tag{31}$$

Here, the normalization constant c is the reciprocal of the denominator and thus an independent real number of state $\mathbf{X}_k$. The last factor can be formally described as

$$\begin{aligned} p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) &= \int p(\mathbf{X}_k, \mathbf{X}_{k-1}|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) d\mathbf{X}_{k-1} \\ &= \int p(\mathbf{X}_k|\mathbf{X}_{k-1}, \mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) p(\mathbf{X}_{k-1}|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) d\mathbf{X}_{k-1} \end{aligned} \tag{32}$$

Assuming the state to be a Markov process and the observations for a given state to be independent of previous observations, we obtain the recursive equation of the Bayes filter

$$p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k) = c \cdot p(\mathbf{Y}_k|\mathbf{X}_k) \cdot \int p(\mathbf{X}_k|\mathbf{X}_{k-1}) p(\mathbf{X}_{k-1}|\mathbf{Y}_0, \ldots, \mathbf{Y}_{k-1}) d\mathbf{X}_{k-1} \tag{33}$$

The Bayes filter thus represents a sequential state estimator, which at each time step comprises the following two steps. In the prediction step, the estimate $\hat{\mathbf{X}}_{k-1}$ of the previous time step is predicted to the current time k. For this purpose, the integral in Eq. 27 is evaluated. In the subsequent innovative step, the prediction is improved by the current observation $\mathbf{Y}_k$, wherein the likelihood term on the left-hand side of the integral in the above equation is evaluated.

The sequential evaluation of the observations allows for elegant and efficient implementations. Generally, the probability density $p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k)$ is not representable in a closed-form solution. For the general case, the so-called particle filter is approximating the probability density. Referring further to Gaussian distributed probability densities and a linear state space model, a particularly efficient processing is possible using the Kalman filter. For further reading on the topic, the reader is referred to Barker et al. (1994), Chen (2003), Meinhold and Singpurwalla (1983), and Welch and Bishop (2006).

4.2 Particle Filter

This section provides an introduction to the particle filter. For the evaluation of Eq. 33, this approach approximates the probability density $p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k)$ by a finite sum of Dirac impulses with weights w_k^i: $p(\mathbf{X}_k|\mathbf{Y}_0, \ldots, \mathbf{Y}_k) \approx \sum_i w_k^i \cdot \delta \left(\mathbf{X}_k - \mathbf{X}_k^i\right)$.

The pairs $\mathbf{X}_k^i$, w_k^i of weight w_k^i and associated state $\mathbf{X}_k^i$ are called particles. In the innovation step, the weights are updated incorporating the latest observations

$$w_k^i \propto \frac{p\left(\mathbf{Y}_k|\mathbf{X}_k^i\right) p\left(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i\right)}{q\left(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i, \mathbf{Y}_k\right)} w_{k-1}^i \quad \text{with} \quad \sum_i w_k^i = 1 \tag{34}$$

The importance density must be selected in the design of the particle filter. A common choice is $q\left(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i, \mathbf{Y}_k\right) = p\left(\mathbf{X}_k^i|\mathbf{X}_{k-1}^i\right)$, whereby Eq. 34 becomes $w_k^i \propto p \left(\mathbf{Y}_k|\mathbf{X}_k^i\right) w_{k-1}^i \quad \text{with} \quad \sum_i w_k^i = 1$. The state and the uncertainty of the estimate can then be determined as follows:

$$\hat{\mathbf{X}}_k = \sum_i w_k^i \cdot \mathbf{X}_k^i \quad \text{und} \quad \hat{\mathbf{P}}_k = \sum_i w_k^i \cdot \left[\mathbf{X}_k^i - \hat{\mathbf{X}}_k\right] \left[\mathbf{X}_k^i - \hat{\mathbf{X}}_k\right]^T \tag{35}$$

A common issue encountered in the filtering algorithm is weight collapse, that is, the situation that all but one of the importance weights are close to zero.

Resampling is used to avoid the problem. It must be pointed out that this introduces a new problem called sample impoverishment, which must be considered when deploying such an approach.

4.3 Tracking with the Kalman Filter

In the simple case that the system dynamics and the observation equation are linear and system as well as observation noise can be assumed to be zero mean Gaussian white noise, Eq. 33 can be easily and efficiently implemented with the so-called Kalman filter. At each time *k*, the normal distribution is completely described by its mean $\hat{\mathbf{X}}_k$ and covariance matrix $\mathbf{P}_k$.

In the prediction step of the Kalman filter, the estimate of the prior time step $\hat{\mathbf{X}}_{k-1}$, $\mathbf{P}_{k-1}$, is projected with the linear system model to the current time step

$$\hat{\mathbf{X}}_k^- = \mathbf{F}\hat{\mathbf{X}}_{k-1} \quad \text{and} \quad \hat{\mathbf{P}}_k^- = \mathbf{F}\hat{\mathbf{P}}_{k-1}\mathbf{F}^T + \mathbf{P}_\mathbf{S} \tag{36}$$

with the state transition matrix $\mathbf{F}$ and the covariance matrix of the system noise $\mathbf{P}_\mathbf{S}$. In the subsequent innovation step, the latest observation $\mathbf{Y}_k$ is considered.

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_k^- + \hat{\mathbf{P}}_k^-\mathbf{G}^T\left(\mathbf{P}_\mathbf{V} + \mathbf{G}\hat{\mathbf{P}}_k^-\mathbf{G}^T\right)^{-1}\left(\mathbf{Y}_k - \mathbf{G}\mathbf{F}\hat{\mathbf{X}}_{k-1}\right) \tag{37}$$

$$\hat{\mathbf{P}}_k = \hat{\mathbf{P}}_k^- - \hat{\mathbf{P}}_k^-\mathbf{G}^T\left(\mathbf{P}_\mathbf{V} + \mathbf{G}\hat{\mathbf{P}}_k^-\mathbf{G}^T\right)^{-1}\mathbf{G}\hat{\mathbf{P}}_k^- \tag{38}$$

with observation matrix $\mathbf{G}$ and covariance matrix of the observation noise $\mathbf{P}_\mathbf{V}$.

In most of the practical applications of the filter framework derived above, the prerequisites can at best be guaranteed approximately. An obvious limitation is the descriptive power of the system and observation models itself, as they generally reflect the real-world conditions only very limited. The resulting inaccuracies often lead to a divergent behavior of the filter, which can be prevented in practice by so-called parameter tuning. In the specific design of a filter, this step states an essential part of the development process. Another limitation in the implementation of the filter is the rounding error of digital arithmetic units, which in turn can lead to a drastic divergence of the filter. A numerically stable and efficient version of the Kalman filter for processors with fixed-point arithmetic is based on the Cholesky factorization. This is achieved by algebraic decomposition of Eqs. 37 and 38 to efficiently realize the implicit matrix inversions there. On parallelizable signal processors, this variant with deterministic runtime behavior can be implemented very efficiently. A widely used filter variant based on the Bierman-Thornton UD algorithm has significantly lower runtime because no root function is calculated (Ghanbarpour and Pourtakdoust 2007; Liu et al. 2007). Due to the lower numerical dynamics, this approach is also more robust to rounding errors, respectively, the risks of indefinite covariance estimates. However, due to the increased use of processors with floating-point arithmetic, these widespread methods are becoming less important.

5 Application Examples

Looking back only a few years, the number of camera-based driver assistance systems on the market has been low and reserved to the vehicle premium segment. Today, most car manufacturers have customer functions based on camera technology on their roadmap, from the compact class up to the premium segment. The latter plays an important role in the development of new and innovative functionalities toward the vision of highly automated driving in the near future. However, legal regulations and an altered understanding of security among customers and authorities have led to a new market situation. This can exemplarily be seen from the development of the new evaluation scheme of the European New Car Assessment Program (Euro NCAP) for the years 2013–2017, in which systems for emergency braking and lane departure warning gained increasing impact to the testing and evaluation scheme. Exemplary, in Fig. 11 the market for camera-based driver assistance systems based on the indicators performance and cost is visualized. It turns out that in addition to the actual performance of a system, its flexibility and applicability will play a more important role in the future. In order to respond appropriately to potential market developments, this fact must also be considered in the development of new detection algorithms. In the past, single driver assistance functions have mostly been understood and developed as self-contained systems with an integrated and independent effect chain from the sensor to the actuator. With increasing market penetration, however, such a strategy cannot succeed, since in addition to the ever-increasing number of new assistance systems, the access to a specific market segment would no longer be manageable from a technological point of view.

By separating the cross-functional task of environment perception from the design of a specific function, the scalability of the overall system can be significantly increased. Furthermore, this architecture allows a partitioning of the algorithms on the control unit or even a distribution of it over several control devices. Algorithm validation and safety aspects of modern assistance functions should not be underestimated. The complexity of requirements can significantly be reduced by the decoupling of perception and function. The environment perception including sensor signal processing and data fusion has the task of providing a generic and internally consistent image of the traffic scene, which is matched in the further processing layers scene interpretation and behavioral decision to a particular function (Grewe et al. 2012).

With a camera, the definition of the term data fusion can be extended from the integration of different sensor signals, which already plays an important role today, to the delivery and interpretation of different recognition results toward a consistent description of the traffic scene around the vehicle (Stiller et al. 2011). This scene model

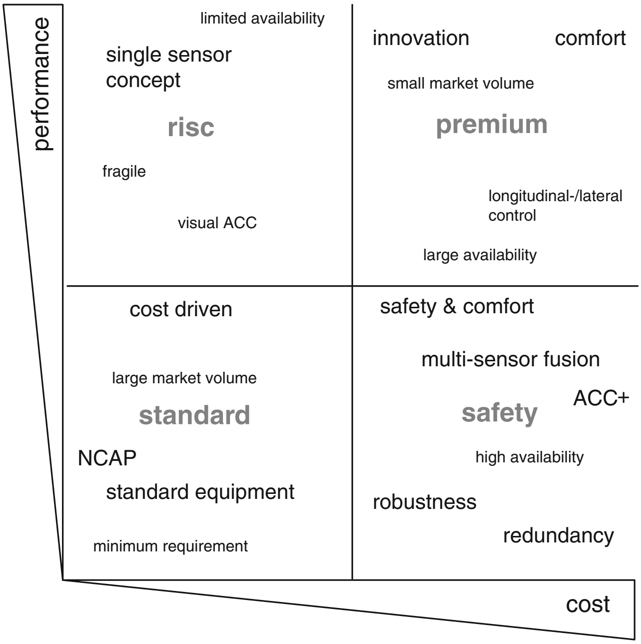


Fig. 11 The market for camera based driver assistance systems based on the indicators performance and cost

can be divided into six logical groups and acts as the central authority for environmental data throughout the vehicle, which is also known as Single Point of Truth.

The traversable vehicle environment is explicitly described by the so-called freespace model. For this purpose, different models exist, which differ from the classical object recognition in that they predominantly model the static environment of the vehicle (Saarinen et al. 2013; Triebel et al. 2006). Further, the scene model contains information about the road geometry such as curb stones and bumps, road markings, road signs, and predictive route data (in the form of navigation maps). The important category of movable or moving objects such as pedestrians or vehicles will be described in more detail below.

5.1 Object Detection

In principle, methods for object detection can be divided into the consecutive steps of detection and verification.

The aim of the detection step is to search in a data-driven process for stable sets of features that describe the objects to be recognized. Due to the low abstraction level of the data and the limitation of processing only a snapshot of the observed scene, the detection output usually suffers from a high false detection rate. This is acceptable in most cases, since faulty hypotheses are filtered out in the subsequent object verification by means of time integration or higher abstracted, object-specific properties. Such properties are, e.g., the symmetry of vehicles, the shadow underneath a vehicle, or characteristic grayscale or color gradients (Bertozzi et al. 1997; Kalinke et al. 1998).

Disparity-based methods for hypothesis generation have the ability to detect, localize, and model arbitrarily shaped objects in the scene. Due to the standard uncertainty of the distance measurement which is quadratically increasing with distance, such methods are primarily applicable in the close range around the vehicle. In addition, the grouping of individual object features in the disparity image is often hampered by gaps in the disparity image.

Motion-based methods make use of monocular image sequence information to detect objects. Mostly, optical flow as presented in Sect. 2.1 is used to generate object hypotheses. For each pixel, the flow vector is determined which describes the movement of the projected scene point in the image plane. In the subsequent motion segmentation step, the grouping of points having similar motion vectors is conducted. Usually, a motion segment is only included in the list of object hypotheses when its expected motion proves to be stable over a longer period of observations. Thus, the false detection rate can be reduced but is still very high.

Appearance-based methods detect the characteristic of an object type on the basis of a training data set. This data set includes typical patterns of a single object type and discloses the different appearances of a dedicated object category. In the training step, a characteristic feature set is generated from the training images. A collection of such features is then combined to form an overall description of the object type. In order to clearly assign a feature vector to an object type, either a classifier is trained or the probability distribution of the features has to be modeled. In Papageorgiou and Poggio (2000) a collection of Haar wavelets is used, which describe the appearance of an object class in the image plane. As classifier, the AdaBoost algorithm has been selected. To describe the characteristic of a vehicle, in Sun et al. (2002) Gabor filters are used to efficiently determine the pronounced grayscale gradient of vehicle edges and lines in a certain order. As a classifier, the support vector machine (SVM) is used. In Mohan et al. (2001) and Bachmann and Dang (2007), road users are described as a collection of image fragments containing portions of the respective object class. A major advantage of this description is the component-based architecture that allows partial occlusion of the object. The main disadvantage of appearance-based methods is the elaborate creation of a representative database for each object type and the time-consuming training of the classifier itself.

By combining the method presented above, the description power of an object hypothesis can be increased immensely as, e.g., through the integration of classification and data-driven signal processing.

In the verification phase, the plausibility of the erroneous and inaccurate object hypotheses from the detection step is checked. This can be achieved, e.g., by temporal integration as described in Sect. 4 or by using predefined, usually parameterized models. These models or templates are then compared with the image data and checked for consistency. Model-based methods have been proven in simple traffic scenarios with restricted diversity and occurrence of objects. Further

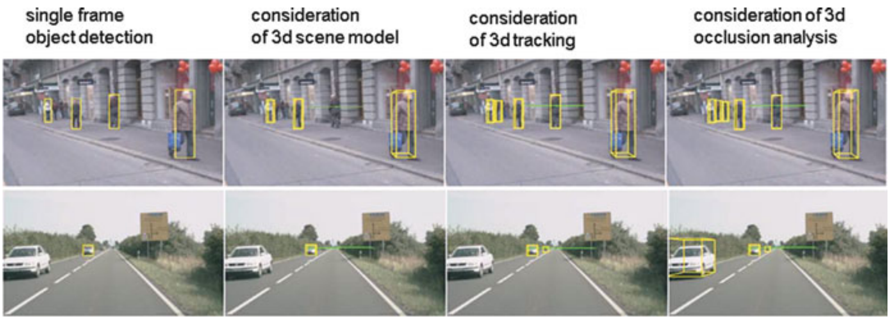


Fig. 12 Taken from (Wojek et al. 2013). In *yellow*, the detection results for the object categories vehicle (*top*) and pedestrians (*bottom*) are shown. With increasing plausibility level of the input data from *left* to *right*, scene model inconsistencies are identified enhancing model robustness

possibilities for verification of object hypotheses are the presence of license plates, lights, or the enclosing contour of equally moving points (Cuchiari and Piccard 1999; Papageorgiou and Poggio 2000).

In the prediction step of the tracking scheme from Sect. 4, the distinction between nonliving physical objects and subjects which have the ability to make independent situation-specific decisions can be beneficial. While with nonliving objects the prediction is made by extrapolation based on physical laws, for subjects it might be necessary to identify intentions, actions, and alternatives. Thus, the motion model can be adjusted depending on an object class and a specific situation to map the adequate motion characteristics. This approach is used already today to distinguish between pedestrians, cyclists, motorcycles, cars, and trucks.

The agglomeration of information in the scene model beyond a specific function or perception category allows the development of a consistent and uniform scene representation, as exemplified in Fig. 12. A balancing of different environment models, which are describing the same aspects of the real world, can be conducted. As a result, improved state estimates of the individual models can be achieved through self-diagnosis and a plausibility check of the estimated scene. The correctness and plausibility of, e.g., dynamics, geometry, and category of an object can be verified by putting it into relation to a detected lane. In return, the corresponding object hypothesis can also be used to support the lane detection module itself. It is also conceivable to exploit the footprint of an object to support the estimation of the road geometry and vice versa. For the intention or action recognition of pedestrians, it may be helpful to know more about its location in the scene and the context in which the pedestrian acts. The movement of a vehicle is also related to the estimated lane course because it can be assumed that drivers comply with traffic rules.

5.2 Intersection Understanding

Besides detecting and tracking other traffic participants, which can be achieved using the techniques described in Sect. 5.1, structural information such as the geometry and topology of the road ahead are essential for advanced driver assistance functions in modern vehicles. Extracting such information without costly infrastructure but solely based on sensors located inside the vehicle demands for solutions to a variety of challenging problems. For example, urban intersections frequently exhibit complex geometries and important cues about the geometry such as lane markings or other traffic participants are occluded by objects in the field of view. While autonomous driving on highways (Dickmanns 1994; Franke 1991), traversing simple and digitally annotated intersections (DARPA Urban Challenge) based on lane detection algorithms and laser scanners, as well as more complex scenarios such as the autonomous Bertha-Benz journey by Daimler and KIT/FZI (Ziegler et al. 2014) have already been demonstrated successfully, the general inner-city case hitherto remains a challenging problem. Simple color-, texture-, and lane-marking features are insufficient due to the high degree of occlusion, defects, and even absence of lane markings for smaller roads. The complexity and variety of possible scenarios thus requires a holistic approach, taking into account all dynamic and static objects as well as their interplay. While at least theoretically the uncertainty in the measurement of past trajectories of all traffic participants may be arbitrarily reduced with improving sensor technology, the inference of their intentions, future behavior, and trajectories is associated with a different kind of uncertainty, which poses challenges to both human and automatic driving (Liebner et al. 2013).

A holistic model for the interpretation of intersection scenarios based on stereo sequences has been presented in Geiger et al. (2014), and Geiger (2013). A camera system mounted on top of the experimental vehicle AnnieWAY (Geiger et al. 2013) supplies the required sensor information. Using a probabilistic generative model which describes the 3D scene geometry and the location and orientation of objects in the scene allows to faithfully consider uncertainties in the imaging process such as detection errors or uncertainty in depth.

The goal of probabilistic inference is then to estimate the most likely road topology (number of intersection arms) and geometry (position, orientation, and width of streets), and to simultaneously assign all traffic participants to the respective lane. Here, a holistic view plays a key role: On the one hand, the position of lane markings and buildings at the side of the road allows for drawing conclusions about the location of traffic participants; on the other hand, tracking dynamic objects in the scene leads to insights about the topology and geometry of the intersection. Information from other traffic participants is thus often complementary to classic lane and road features: While lane markings are often occluded by vehicles and other objects in heavy traffic, they provide important information when streets are empty. Further features such as vanishing points and occupancy grids of the environment provide additional information to the probabilistic inference process.

More specifically, Geiger et al. (2014) and Geiger (2013) integrate the features illustrated in Fig. 13 into a single model and investigate the utility of each feature with respect to estimating the intersection geometry, topology, and the location of traffic participants in 113 representative and challenging intersection scenarios. Each of the features is briefly described in the following:

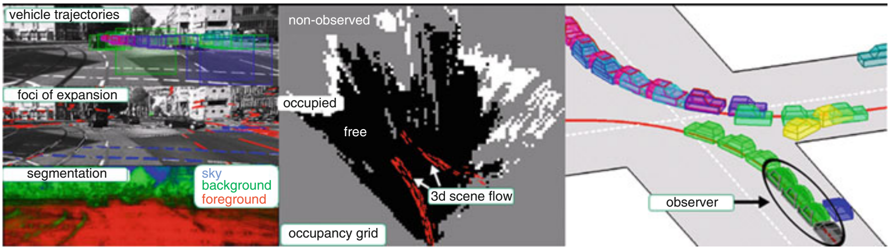


Fig. 13 Monocular features such as vehicle trajectories, vanishing points, semantic segmentation as well as stereo features such as 3d scene flow and occupancy grids computed from disparity maps allow for a holistic analysis of the scene at an urban intersection and yield information about the geometry and topology of the road as well as the location and motion of other traffic participants

Vehicle Trajectories: The motion of other traffic participants is a strong feature which conveys information about the location of streets, lanes, turns, and the status of traffic lights. A stereo camera is used to detect vehicles and to track them in 3D using methods akin to the particle filter described before. Besides depth information from the stereo sensor, appearance-based methods are leveraged to verify valid object hypotheses. The location of the camera system with respect to the road as well as stereo information can be used to estimate the distance of the observed object from the stereo camera rig. The orientation of the object can be extracted from appearance information as well as from the estimated motion of the object. Given the intersection topology and geometry, vehicles which are located at a lane and follow its orientation are considered likely to occur. In addition, the driving direction as well as the vehicle dynamics are taken into account. Parking vehicles at the side of the road are represented by a parking model.

Vanishing Points: Recognizing vanishing lines and vanishing points facilitates the precise estimation of the orientation of roads meeting at an intersection as many edges in the image (such as lane markings, building facades, walls, and windows) correspond to 3D lines which are oriented the same way as the respective roads or lanes. First, line segments are extracted from the image by clustering edge pixels. Next, each line segment is classified into either a structured edge or a structural outlier using its local appearance in the image in order to limit the influence of systematic errors due to cast shadows which occur frequently in urban domains. The verification of vanishing-point hypotheses is finally carried out by reprojecting them into the image using a voting-based method. Vanishing lines which are collinear in 3D increase the probability of a road oriented the same way while outliers are handled by a separate outlier model.

Semantic Image Segmentation: Further cues regarding the intersection geometry can be obtained by classifying each image pixel into one of the following three classes: street, background (buildings, vegetation), and sky. Comparing a virtually generated semantic segmentation with the image evidence for each pixel leads to another likelihood term in the probabilistic generative model. The sky region is often relatively easy to segment due to its unique color and location properties and leads to conclusions about the location of urban canyons when including some additional assumptions about the scene, such as an average building height of approximately four stories.

Scene Flow: While all previously mentioned features can be extracted from monocular image sequences, stereo images provide valuable depth information (see also Sect. 3) which can be used to estimate the 3D scene flow at each surface point in the scene. In order to calculate the 3D scene flow at a pixel, correspondences between the left and right image at two consecutive image frames need to be established. This results in the 3D location of a point in the world at two consecutive time steps and thus yields a 3D displacement vector (in contrast to the optical flow vectors described in Sect. 2.2). After compensating for the motion of the observer (which can be estimated using visual odometry) and thresholding, only those 3D scene flow vectors which originate from dynamic objects in the scene remain. These vectors are considered likely if they agree with the closest lane, both in terms of their location as well as their orientation.

Occupancy Grid: The static environment plays a central role when perceiving the scene as a whole: Inner-city buildings, for example, are often located close to the drivable road area, and situations where streets pass through or below buildings are typically rather rare. Occupancy maps can be leveraged for registering the static scene elements extracted via dense disparity maps. More specifically, at each pixel the viewing ray is traced until an object is reached. The Bresenham algorithm allows for discretizing this ray into pixels (2D) or voxels (3D), and the corresponding grid cells are marked as free or occupied. The state of the cells is tracked using a binary Bayes filter. In the probabilistic model, free space areas extracted using this approach represent candidates for road segments.

While each of these features is individually able to improve the inference results, the evaluation in Geiger et al. (2014) shows that the best results can be obtained by combining all features in a joint model. Vehicle trajectories, 3D scene flow, and occupancy grids have been determined as the most important features while the detected vanishing points and the semantic segmentation cues have been found to contribute relatively little to the performance of the full system. Fig. 13 illustrates an inference result of the proposed method. The reader is referred to Geiger et al. (2014), and Geiger (2013) for a mathematical formulation of the model and a detailed quantitative and qualitative evaluation.

6 Summary and Conclusions

Camera sensors and machine vision introduce perception techniques into automobiles that closely resemble the drivers' perception. The continuing decay of costs for camera and computing hardware makes camera systems the sensors of choice for a steadily growing variety of applications. A key advantage of video sensors in comparison to other environmental sensors is their most comprehensive acquisition of extensive information about the observed traffic scene. At the same time, the analysis of such extensive image information constitutes a major challenge to signal processing.

The wide and growing employment of vision sensors in driver assistance systems may be explained by the following properties:

- Being a passive measurement principle, vision sensors suffer neither from legal restrictions for their admission in traffic nor from possible interference effects.
- Humans have designed traffic infrastructure for visual perception. Hence, the relevant information in a traffic scene can be fully captured by vision sensors. The extensive information captured in image sequences allows for many tasks relevant to driver assistance systems.
- Being close to human perception, vision sensors may provide a high degree of transparency to drivers.

This chapter describes fundamentals of camera systems and associated machine vision techniques for driver assistance systems and automated driving. The characteristic loss of information by an entire dimension caused by the projection of the 3D world onto the 2D image plane has been discussed. Homogeneous geometry has been introduced that allows for an elegant mathematical description of perspective projection. In homogeneous coordinates, numerous mappings become linear. The intrinsic and extrinsic calibration parameters uniquely specify the orientation of any pixel's ray of sight in 3D space.

A wide variety of application-specific image processing methods has been proposed for the interpretation and analysis of image sequences. These are usually implemented in a modular architecture that consecutively reduces the amount of data processed. It typically begins with a preprocessing step that improves image quality and compensates for radiometric and geometric distortions of the acquisition process. Task-specific features are extracted during feature extraction that are fed to high-level analysis. This chapter has focused on the detection of salient image features and on image correspondence features expressed by disparity and optical flow. Temporal tracking is often applied to consistently accumulate information over time. Emerging from the general Bayes filter implementation considerations have been discussed for the particle filter and the Kalman filter, which are special practical realizations.

Practical examples and an overview of current research on object recognition and intersection recognition illustrate the interaction and performance of the methods described. In order to simultaneously handle the requirements and the complexity of future driver assistance systems, it appears important to embed classical image processing methods in an overall architecture that successively extracts and reduces information and that allows for holistic plausibilization of the individual processing results.

Acknowledgments The authors thank Dr. Christian Duchow for valuable contributions to an early version of this chapter.

References

Bachmann A, Dang T (2007) Improving motion-based object detection by incorporating object-specific knowledge. Int J Intell Inf Database Syst, Special issue on: Information Processing in Intelligent Vehicles and Road Applications 2(2):258–276

Barker A, Brown DE, Martin WN (1994) Bayesian estimation and the Kalman filter, technical report, IPC-94-002, 5

Bay H, Tuytelaars T, van Gool L (2006) SURF: speeded up robust features. In: European conference on computer vision, Berlin, Heidelberg

Bertozzi M, Broggi A, Castellucio S (1997) A real-time oriented system for vehicle detection. J Syst Archit 1:317–325

Calonder M, Lepetit V, Ozuysal M, Trzcinski T, Stretcha C, Fua P (2012) BRIEF: computing a local binary descriptor very fast. IEEE Trans Pattern Anal Mach Intell 7:1281–1298

Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Intell 6:679–698

Chen Z (2003) Bayesian filtering: from Kalman filters to particle filters, and beyond. Statistics 182 (1):1–69

Cuchiari R, Piccard M (1999) Vehicle detection under day and night illumination. In: International ICSC symposium intelligent industrial automation

Dang T, Hoffmann C, Stiller C (2009) Continuous stereo self-calibration by camera parameter tracking. IEEE Trans Image Process 18(7):1536–1550

Dickmanns ED (1994) The 4D-approach to visual control of autonomous systems. In: IAA NASA conference on intelligent robots in field factory service and space, Houston

Faugeras O (1993) Three dimensional computer vision: a geometric viewpoint. MIT Press, Cambridge, MA

Franke U (1991) Real time 3d-road modeling for autonomous vehicle guidance. In: 7th Scandinavian conference on image analysis, Aalborg

Geiger A (2013) Probabilistic Models for 3D urban scene understanding from movable platforms, PhD thesis, Karlsruhe Institute of Technology

Geiger A, Lenz P, Stiller C, Urtasun R (2013) Vision meets Robotics: the KITTI dataset. Int J Rob Res 32:1231–1237

Geiger A, Lauer M, Wojek C, Stiller C, Urtasun R (2014) 3D traffic scene understanding from movable platforms. IEEE Trans Pattern Anal Mach Intell 36(5):1012–1025

Ghanbarpour Asl H, Pourtakdoust SH (2007) UD Covariance factorization for unscented Kalman filter using sequential measurements update. Int J Eng Nat Sc 1:4

Gonzalez RC, Woods RE (2006) Digital image processing, 3rd edn. Prentice-Hall, Upper Saddle River

Grewe R, Komar M, Hohm A, Hegemann S, Lüke S (2012) Environment modelling for future ADAS functions. In: 19th ITS world congress

Harris CG, Stephens M (1988) A combined corner and edge detector. In: 4th Alvey vision conference, 1988, pp 147–151

Kalinke T, Tzokamkas C, von Seelen W (1998) A texture-based object detection and an adaptive model-based classification. In: IEEE intelligent vehicles symposium, pp 143–148

Lategahn H, Stiller C (2014) Vision-only localization. IEEE Trans Intell Transp Syst 15 (3):1246–1257

Lategahn H, Beck J, Kitt B, Stiller C (2013) How to learn an illumination robust image feature for place recognition. In: IEEE intelligent vehicles symposium, Gold Coast

Liebner M, Klanner F, Baumann M, Ruhhammer C, Stiller C (2013) Velocity-based driver intent inference at urban intersections in the presence of preceding vehicles. IEEE Intell Transp Syst Mag 5(2):10–21

Liu Y, Bougani CS, Cheung PYK (2007) Efficient mapping of a Kalman filter into an FPGA using Taylor expansion. In: International conference on field programmable logic and applications, pp 345–350

Longuet-Higgins HC (1981) A computer algorithm for reconstructing a scene from two projections. Nature 293:133–135

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60 (2):91–110

Lucas BD, Kanade T (1981) An iterative image registration technique with an application to stereo vision. In: International joint conference on artificial intelligence, Vancouver, pp 674–679

Meinhold RJ, Singpurwalla ND (1983) Understanding the Kalman filter. Am Stat 37(2):123–127

Mohan A, Papageorgiou C, Poggio T (2001) Example-based object detection in images by components. IEEE Trans Pattern Anal Mach Intell 23(4):349–361

Nistér D (2004) An efficient solution to the five-point relative pose problem. IEEE Trans Pattern Anal Mach Intell 26(6):756–777

Papageorgiou C, Poggio T (2000) A trainable system for object detection. Int J Comput Vis 38 (1):15–33

Saarinen J, Andreasson H, Stoyanov T, Luhtala JA, Lilienthal AJ (2013) Normal distributions transform occupancy maps: application to large-scale online 3d mapping. In: IEEE international conference on robotics and automation

Shi J, Tomasi C (1994) Good features to track. In: IEEE conference on computer vision and pattern recognition, Seattle

Stiller C, Konrad J (1999) Estimating motion in image sequences - a tutorial on modeling and computation in 2D motion. IEEE Signal Process Mag 7 and 9:70–91, 116–117

Stiller C, Puente Leon F, Kruse M (2011) Information fusion for automotive applications - an overview. Inf Fusion 12(4):244–252

Sun Z, Bebis B, Miller R (2002) On-road vehicle detection using Gabor-filters and SVM. In: IEEE international conference on digital signal processing, Santori, Greece

Tola E, Lepetit V, Fua P (2010) DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans Pattern Anal Mach Intell 32(5):815–830

Triebel R, Pfaff P, Burgard W (2006) Multi-level surface maps for outdoor terrain mapping and loop closing. In: IEEE/RSJ international conference on intelligent robots and systems, pp 2276–2282

Trucco E, Verri A (1998) Introductory techniques for 3–D computer vision. Prentice Hall, New York

Welch G, Bishop G (2006) An introduction to the Kalman filter. Technical report. http://www.cs.unc.edu/welch/kalman/kalmanIntro.html

Wojek C, Walk S, Roth S, Schindler K, Schiele B (2013) Monocular visual scene understanding: understanding multi-object traffic scenes. IEEE Trans Pattern Anal Mach Intell 35:882–897

Ziegler J, Bender P, Schreiber M, Lategahn H, Strauss T, Stiller C, Dang T, Franke U, Appenrodt N, Keller CG, Kaus E, Herrtwich RG, Rabe C, Pfeiffer D, Lindner F, Stein F, Erbs F, Enzweiler M, Knoppel C, Hipp J, Haueis M, Trepte M, Brenk C, Tamke A, Ghanaat M, Braun M, Joos A, Fritz H, Mock H, Hein M, Zeeb E (2014) Making Bertha drive - an autonomous journey on a historic route. IEEE Intell Transp Syst Mag 6(2):8–20

Stereovision for ADAS

21

Stefan Gehrig and Uwe Franke

Contents

Abstract

Camera-based driver assistance went from pure research level activities in the early 1990s to standard equipment products in vehicles nowadays. This change is due to both technological advances and algorithmic developments. Especially, stereo vision plays a vital role in advanced driver assistance systems. In this chapter, we lay the foundations of stereo vision, show the specific needs for driver assistance, and build up a processing chain combining motion estimation and stereo information. The combination is called 6D-Vision and allows velocity estimation on a pixel level. For efficiency reasons, an intermediate stereo representation called Stixels is introduced from which an object representation

---

S. Gehrig (✉) • U. Franke
Group Research and Advanced Engineering, Driver Assistance and Chassis Systems, Daimler AG, Sindelfingen, Germany
e-mail: stefan.gehrig@daimler.com; uwe.franke@daimler.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_22

is derived. The final section describes future trends in stereo vision such as a combination with machine learning approaches.

The first camera-based driver assistance systems were implemented as prototypes within the European project PROMETHEUS in the early 1990s. Back then, only a few optimists envisioned this technology to play a vital role in the real world some 20 years later. Moreover, this technology now acts as an enabler for high-end safety systems such as autonomous braking for pedestrians. No other sensor besides the camera could benefit from recent technological advances so much.

The costs for a camera have dropped from well above 1000$ in the beginning to a few dozen dollars these days. The limited dynamic range of early CCD sensors was dramatically improved, thanks to the modern CMOS-based consumer cameras. Back then, a presentation of in-vehicle computer vision algorithms was doomed when sunlight hit the camera. At the same time, the available computing power increased by five orders of magnitude, as predicted by Moore's law. Another important enabler is the FPGA (field-programmable gate array) technology, which is suitable for highly parallel image processing tasks on pixel level.

The enormous progress on the hardware side was complemented by innovations on the algorithmic level. This includes:

- The introduction of the Kalman filter in image sequence analysis by E.D. Dickmanns (Dickmanns and Zapp 1986) which allows the integration of dynamic systems knowledge in the processing chain
- The progress in the machine learning domain (cf. ▶ Chap. 22, "Camera Based Pedestrian Detection") which enables the utilization of image statistics and large image databases
- The step from local ad hoc methods to global optimal approaches that use well-founded mathematical methods for finding optimal solutions
- The introduction of stereo vision as the basis for robust image understanding

In contrast to the monocular approaches using only one eye, stereo vision allows a full 3D reconstruction of the environment within a single image pair, independent of the motion state of the observer and the observed objects. This enables advanced driver assistance systems, e.g., autonomous braking for pedestrians or collision avoidance for crossing objects, and partially automated driving functions in traffic jams. Moreover, stereo vision also precisely measures the three-dimensional shape of the road ahead, which allows an adaption of the active body control systems to compensate bumps. For these reasons, stereo cameras became available for Mercedes models in 2013.

In the literature, stereo vision is considered a largely solved problem. However, the outdoor traffic scenario and the usage in safety-critical driver assistance systems pose particularly hard requirements on current methods. These requirements include:

- Robustness: In traffic scenes, image disturbances occur due to illumination and weather circumstances that are not covered by most approaches in the literature.

Fig. 1 Safety-critical driver assistance applications pose high demands on the robustness of algorithms. (**a**) When raining, blinding and windshield smear can occur due to water on the windshield. (**b**) At nighttime, windshield wipers can generate disturbances with characteristic smear artifacts. (**c**) Snow fall and wet streets cause problems in disparity estimation. (**d**) Backlight can generate reflections on the windscreen that must not lead to erroneous obstacle detection (Figures courtesy of the HCI benchmark of D. Kondermann, University of Heidelberg (Meister et al. 2012))

Artifacts are, e.g., sun blinding, reflections, blurring caused by water on the windshield, spray water, partial occlusion due to the windshield wiper, snow, or darkness. Figure 1 shows four examples.

- Accuracy: The desired measurement range is rather large and ranges from the bumper (less than 2 m) to 50 … 80 m. Hence, the disparity estimate has to be sub-pixel accurate, which is not demanded in current stereo benchmarks.
- Real-time capability: In order to enable fast responses, the processing usually runs at 25 … 30 Hz. Many literature methods process one single image pair in seconds to minutes of runtime on high-end PCs.
- Long-term stability: vision systems are expected to work for the entire life cycle of the vehicle. For a stereo system, this includes the capability to perform online calibration in the field to maintain a well-calibrated system under varying temperatures and aging effects.

– Power consumption: For cost reasons, a one-box solution hosting camera and processing unit is desirable. If the camera system is mounted behind the rearview mirror, the power consumption has to be rather low (without active cooling) to avoid overheating.

The organization of this chapter is following the processing pipeline used by successful stereo vision systems. In Sect. 1, the traditional local methods of disparity estimation are compared to modern global optimal stereo methods. Since the accuracy of the disparity estimation is vital for the success of downstream object formation methods, we dedicate Sect. 2 to this topic. Especially, the demand for methods that also operate under adverse weather conditions is often overlooked in the literature. One reason for that is the lack of ground truth under adverse weather condition in outdoor scenarios.

Stereo vision delivers 3D measurements for every pixel. If the points are tracked over time, one can estimate the motion in 3D. This leads to Sect. 3 on 6D-Vision with fusion in space and time for a powerful basis to quickly detect moving objects. Especially, laterally moving objects such as crossing cars or kids running onto the street can be reliably detected this way.

Computer vision in recent years has exhibited a trend towards using superpixels as opposed to pixels. With the goal to represent the traffic scene compactly in 3D despite the growing imager resolution and to offer efficient subsequent processing steps, the Stixel World has been developed. This efficient representation tailored for traffic scene is introduced in Sect. 4, as well as the subsequent steps for the final object detection. This concludes the transfer from pixels to objects. The chapter ends with a summary and remarks for further research.

1 Local and Global Disparity Estimation Methods

The basic steps of disparity estimation were introduced in ▶ Chap. 20, "Fundamentals of Machine Vision". More similarity criteria common in the automotive field are described, and several options for disparity estimation via correlation are detailed in this chapter. For a categorization of stereo methods and the different processing steps, we refer to the seminal work of Scharstein and Szeliski (2002). According to the processing steps defined there, we distinguish between similarity criteria, disparity optimization, and sub-pixel estimation. We limit ourselves to spatially discrete methods that scan the disparity volume in discrete steps, since the continuous methods require significantly more computational resources to deal with the displacements present in traffic scenes. To compare published stereo methods, the Middlebury website (http://vision.middlebury.edu/stereo/) was launched and currently ranks some 150 methods. A similar benchmark for the automotive field is the KITTI benchmark launched in 2012 (Geiger et al. 2012). This data set contains images with some typical problems for traffic scenes, such as reflections. Since some methods are very sensitive to such disturbances, differences in the ranking between both benchmarks occur.

For automotive purposes, a real-time computation is of uttermost importance. Thus, mostly local methods such as correlation were considered until recently.

1.1 Local Correlation Methods

As shown in ▶ Chap. 20, "Fundamentals of Machine Vision", the classic method of disparity estimation works with correlation independently for every pixel. Independent refers to "independent of the results of neighboring pixels." For every pixel in the reference image (left image), the corresponding pixel in the search image (right image) that images the same world point is determined. For that purpose, the similarity of two pixels has to be established.

A large variety of criteria exist to determine the similarity of two pixels. We restrict ourselves to similarity metrics based on gray values. The simplest similarity metric is the difference of the two gray values in the left image (g_l) and the right image (g_r):

$$ABS(d) = |g_l\,(x; y) - g_r(x - d, y)| \tag{1}$$

This difference is computed line by line for all disparity hypotheses in the rectified image (▶ Chap. 20, "Fundamentals of Machine Vision"). After rectification, the images are aligned in standard stereo geometry, i.e., corresponding points lie on the same image row. In the automotive field, disparities down to 0 have to be evaluated (corresponding to infinite distance). The maximum disparity is defined by the smallest distance to be measured.

Since even high dynamic range cameras hardly deliver more than 12 bits, this metric is always ambiguous. In addition, cameras cannot be equalized well enough for a common world point to be imaged with exactly the same gray value. This ambiguity can be reduced by using a window around the considered pixel. This was exemplified in ▶ Chap. 20, "Fundamentals of Machine Vision" for the "sum of absolute differences" (*SAD*). Alternately, one can compute the "sum of squared differences (*SSD*)":

$$SSD(d) = \sum\nolimits_{x,y \in B} (g_l(x,y) - g_r(x - d, y))^2 \tag{2}$$

This computation scheme leads to strong deviations from the perfect similarity of 0, when the considered pixels differ by more than a few gray values. The authors prefer the similarity metric SAD that penalizes outlier less and leads to better results in practice. Independent of the chosen metric, one can find wrong correspondences when radiometric deviations between the cameras occur. This problem can be alleviated by subtracting the mean of the image window B, exemplified with the absolute difference:

$$ZSAD(x,y,d) = \sum\nolimits_{x,y \in B} \left| \left(g_l(x,y) - \overline{b}_l\right) - \left(g_r(x - d, y) - \overline{b}_r\right) \right| \tag{3}$$

For all introduced similarity metrics, one can find the optimal disparity value d^* for a pixel simply by finding the minimum over d:

$$d^*(x, y) = \min_d ZSAD(x, y, d) \tag{4}$$

A popular similarity metric that also delivers a confidence value (measure of agreement) besides the optimal disparity is the previously introduced mean-free cross-correlation function $p(d)$ (cf. ▶ Chap. 20, "Fundamentals of Machine Vision"). Its value ranges from -1 to 1. Good correspondences obtain values close to one; consequently, the maximum of the metric is searched. Most of the time, correspondences with $p(d) > 0.7$ are accepted. Although the metric yields a good measure of similarity, one has to make sure that high-contrast regions are not matched with low-contrast regions that look the same after normalization performed by the score function.

The image window is typically chosen to be rectangular with a size of 3 × 3–9 × 9 pixels. The above metrics require that corresponding points lie on exactly the same line. Similarity metrics on single points (i.e., 1 × 1 pixels) are particularly susceptible to small calibration errors. If the epipolar geometry is shifted by just one line, the corresponding point cannot be found since the search runs through the wrong image line. A larger image window reduces the sensitivity to calibration errors but induces higher computational load and causes the so-called foreground fattening: Since nearby objects usually have higher contrast than the background, background pixels next to the foreground are often associated with the foreground disparity.

When rank statistics are used instead of gray value information, the similarity metric becomes less sensitive to small rectification errors (Hirschmueller and Gehrig 2009). The Hamming distance of the census transform is a popular similarity metric that is more tolerant to such rectification errors and very efficient to compute. For every pixel within an image window, a bit is generated signifying whether the pixel's gray value is larger (1) or not (0) than the center pixel. This transform assigns a bitstring to every image window. The number of pixels within the considered window defines the bitstring length. The comparison of the two bitstrings from the reference image and the search image is straightforward. The Hamming distance provides an easy similarity metric for these two bitstrings. It just counts the number of different bits:

$$\mathrm{CENSUS}(x,\ y,\ d) = \mathrm{HAM}\left(T_{C(g_l, x, y)} - T_{C(g_r, x-d, y)}\right) \tag{5}$$

$T_c(g_l,x,y)$ represents the census-transformed image at location x based on the gray values of the left image. The smallest Hamming distance is the best disparity estimation. The census transform is invariant to linear transformations of the gray values in the left and right image, which occurs frequently due to sensitivity variations in cameras. Evaluations in Hirschmueller and Scharstein (2009) show that the census metric is superior to other metrics, especially when used with global stereo algorithms.

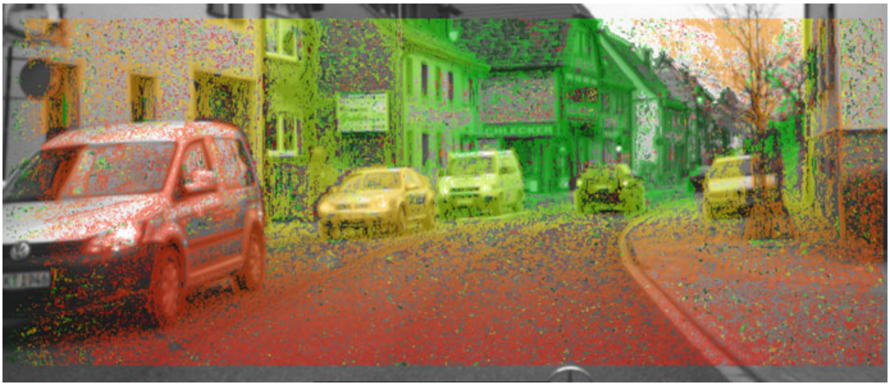


Fig. 2 Color-coded disparity image (*red* nearby . . . *green* far) overlaid on the original image. The disparity image does not contain a result for every pixel (no color overlay). Isolated outliers (*red* points in the background) are visible

Consistency Checks: The local methods introduced here cannot deliver meaningful results in texture-less areas. By applying a so-called interest operator, e.g., an edge filter (see Canny filter from ▶ Chap. 20, "Fundamentals of Machine Vision"), one can restrict the stereo analysis to areas with sufficient contrast. In addition, unreliable correspondences can be filtered out by computing the disparity image twice, swapping the reference and the match image: If pixel (x) in the left image obtains disparity d as a result, pixel (x-d) in the right image must have the disparity–d for the correspondence to be correct. This right-left check (RL check) can be applied without computing the disparity map twice (Muehlmann et al. 2001). For in-vehicle applications, one has to perform such consistency checks to deal with windshield wiper occlusions, since the camera is usually mounted in the wiper area of the windshield. A windshield wiper is only visible in one image, and the (wrong) disparities are removed with the RL check.

Figure 2 shows the result of a local correlation method with ZSAD used as similarity metric. Unreliable matches were removed with a RL check. Despite this check, occasional red (nearby) points remain in texture-less regions that are obviously wrong.

1.2 Global Stereo Methods

Correlation-based methods need sufficient image contrast to yield good results. This is not the case for traffic scenes with texture-less sky and pavements. Stereo analysis can be improved with additional constraints. In typical scenes, depth changes gradually and remains constant on planes parallel to the camera (frontoparallel). The only depth discontinuities occur at object boundaries. Global stereo algorithms try to incorporate this so-called smoothness constraint in the disparity estimation process and achieve improved results. This algorithm can be subdivided into 1D-optimizing methods along an image line and 2D-optimizing methods across the image.

1.2.1 1D Optimization

With the assumption of piecewise constant depth, a new smoothness term appears in addition to the so-called data term represented by the similarity metric. The disparity optimization is interpreted as an energy optimization task minimizing the total energy. The total energy E_{total} hence consists of the similarity E_{data} and the smoothness energy $E_{\text{smoothness}}$, summed up for all pixels:

$$E_{\text{total}} = \Sigma_{x,y}\left(E_{\text{data}} + E_{\text{smoothness}}\right) \tag{6}$$

For modeling the smoothness energy, the following principle has been established: Large depth discontinuities are penalized with a constant energy (P_2), small changes in disparity with a smaller energy P_1:

$$E_{\text{smoothness}} = \begin{cases} 0, \ \mathit{if} \ |d_1 - d_2| = 0 \\ P_1, \mathit{if} \ |d_1 - d_2| = 1 \\ P_2, \ \mathit{if} \ |d_1 - d_2| > 1 \end{cases} \tag{7}$$

d_1 and d_2 denote disparities of adjacent pixels. The smaller energy P_1 is designed to correctly reconstruct slanted surfaces that exhibit gradual disparity changes, in most cases smaller than 1 pixel. The simpler Gibbs potential, often used in global stereo schemes, can be obtained by setting $P_1 = P_2$. Often, P_2 is adapted to the image gradient. If an intensity edge is present, P_2 is reduced because a depth discontinuity is more likely to coincide with intensity edges. Such an energy optimization can be efficiently executed along one direction, e.g., along an image row. This method considers every row independently and is known as "scanline optimization" in the literature. If the optimal disparity is found at the end of the row, backtracking is started to find the optimal disparity along the row. This method hence employs the principle of dynamic programming and is known as "dynamic programming stereo" (Belhumeur 1996). Both methods can be efficiently implemented in embedded systems, but the results exhibit streaking artifacts since the optimization takes place line by line independently. Figure 3 shows a scanline optimization example with a 9 × 7 census mask as data term.

1.2.2 2D Optimization

The streaking artifacts are avoided when the optimization takes all adjacent neighbors into account, i.e., when a two-dimensional optimization is carried out. There are several approaches to find the energy minimum. The most common ones are sketched in the following:

GraphCut: The optimization method GraphCut conducts an energy minimization on a graph. There, pixels are considered as nodes and the connections in

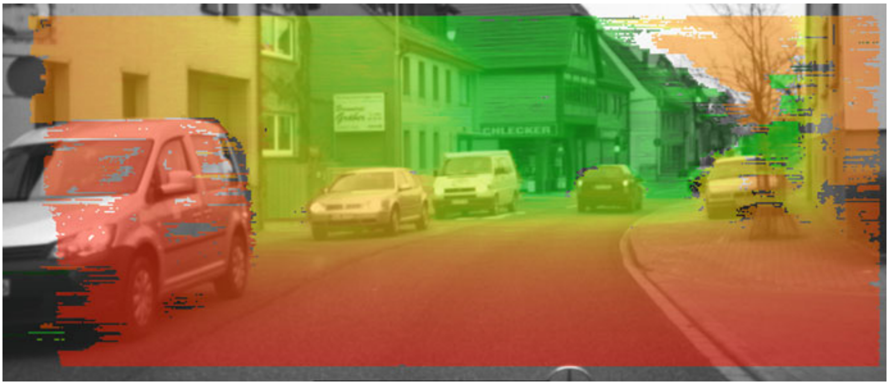


Fig. 3 Two color-coded disparity image (*red* nearby … *green* far) obtained by scan line optimization overlaid on the original image. The row-wise optimization leads to streaking artifacts that are even worse in low-contrast regions

Fig. 4 Two color-coded disparity image (*red* nearby … *green* far) obtained by GraphCut overlaid on the original image. A visually error-free disparity image is generated. However, computation time exceeds 10 s (in 2014)

between are considered edges. GraphCut is often used for foreground-background segmentation. For such a binary problem with two labels, GraphCut is guaranteed to find the optimal solution, i.e., the global minimum. The extension to more than two labels leads to an iterative scheme without optimality guarantee, usually giving very good results in practice. The labels in the stereo case are interpreted as discrete disparities (Boykow et al. 2001). Figure 4 shows an example result using the 9 × 7 census similarity metric as data term. The computation time depends on the number of pixels and disparity hypotheses. For typical image size and disparity ranges in

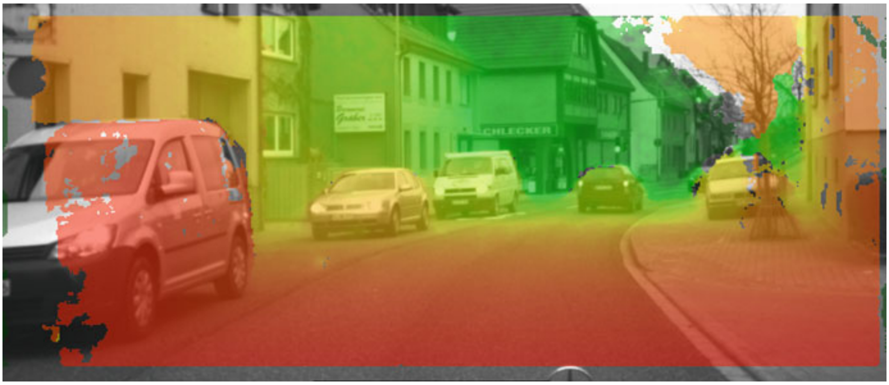


Fig. 5 Color-coded disparity image obtained by SGM (*red* nearby . . . *green* far) overlaid on the original image. A visually error-free disparity image is generated, similar to the GraphCut result

driver assistance, computation times are prohibitively large. The optimization method "belief propagation" solves the same task as GraphCut and delivers similar results with similar computation times when comparable parameters are used (Tappen and Freeman 2003).

Semi-global Matching: With semi-global matching (SGM), streaking artifacts are removed, but the computational efficiency of 1D-optmization methods is maintained. SGM conducts a scanline optimization as described above, but this time in multiple directions (typically 8), and the results are summed up (Hirschmueller 2008). This method, proposed by Hirschmueller in 2005, approximates the 2D optimization by multiple independent 1D optimization steps. Similar results to GraphCut at a fraction of the GraphCut runtime are obtained.

Hirschmueller originally proposed SGM with mutual information as similarity metric, a pixel-based metric very sensitive to calibration errors. In driver assistance scenarios and for real-time implementations, census is preferred as similarity metric. Real-time implementations exist for Intel processors (Gehrig and Rabe 2010), GPU (Banz et al. 2011), and reconfigurable hardware (FPGA) (Gehrig et al. 2009). This method is applied in the Daimler stereo camera. An example can be viewed in Fig. 5.

The key to success can be explained like this: Isolated pixels exhibit weak and ambiguous minima in low-texture areas. By comparing disparity hypotheses with the neighbors and the preference of smooth solutions, all pixels can find the correct disparity as a compatible solution, even when the disparity minima of the isolated pixels do not coincide.

An extension to color images is simple. For driver assistance, color images for stereo matching have not been proven useful when contrasted with the additional computation effort. Using color information leads to worse results than robust similarity metrics such as ZSAD when color constancy is not perfectly fulfilled (Bleyer and Chambon 2010).

2 Accuracy of Disparity Estimation

The global stereo methods described in the previous section deliver integer disparities, with the distance Z being inversely proportional to the disparity d via

$$Z = fB/d \tag{8}$$

With f being the focal length and B the baseline, integer disparities lead to undesired quantization artifacts of the estimated distance that are only tolerable in the near range. Figure 6 illustrates these quantization effects.

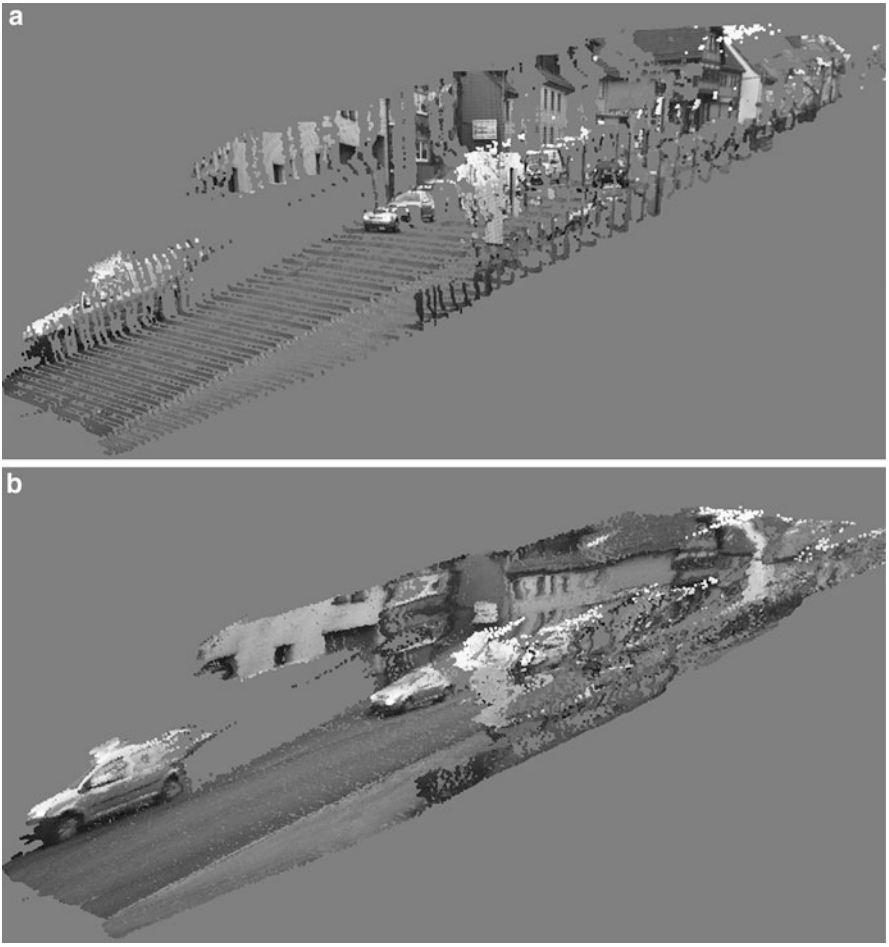


Fig. 6 3D reconstruction using SGM based on the disparity map of Fig. 5. At the *top*, the triangulated result without sub-pixel estimation and, at the *bottom*, the result with sub-pixel estimation are shown. Without sub-pixel estimation, the oncoming car "decays" into two parts

These quantization effects can be reduced by using a larger focal length and/or a larger baseline. However, this is counterproductive to the goals of large field of view (i.e., small focal length) and design-friendly small baselines. Alternately, one can increase the imager resolution. Unfortunately, this increases computation and memory demands and reduces the imager sensitivity, when the imager size remains constant due to cost reasons.

Driver assistance asks for large measurement range and high measurement accuracy, which makes sub-pixel estimation indispensable. At the same time, small decalibrations induce further systematic errors. The following section is devoted to these problems.

2.1 Sub-Pixel Estimation

Deriving the disparity-distance relationship from Eq. 8 yields

$$\frac{\partial Z}{\partial d} = -\frac{f \cdot B}{d^2} = -\frac{Z^2}{f \cdot B} \tag{9}$$

Accordingly, depth errors increase quadratically with the distance to the camera, a property of triangulation methods. This uncertainty as a function of distance Z leads to distance uncertainties of several meters in ranges above 50 m, given a stereo system with $f \cdot B = 250\ \mathrm{m} \cdot px$, roughly the stereo camera geometry used in Mercedes cars.

Independent of the applied stereo method, one can generate a sub-pixel estimation with little effort. For quadratic similarity metrics (e.g., SSD), the best-known sub-pixel estimation conducts a second-order Taylor approximation at the location of the minimum of the similarity function. The minimum of this quadratic polynomial marks the sub-pixel estimate. For linear similarity functions (e.g., SAD, Hamming distance of Census), the so-called equiangular fit has proven its usefulness (Shimizu and Okutomi 2001). In that case, the optimal disparity is determined by intersecting two lines defined by the minimum cost and its neighbors.

Shimizu (Shimizu and Okutomi 2001) shows that the proposed sub-pixel disparity estimation leads to a preference of integer disparities. The effect depends on the similarity metric and on the contrast within the considered pixel window. In certain cases with high contrast, the error can amount up to 0.15 px disparity. With the right choice of interpolation function, the effect is smaller than 0.1 px. More dramatic is the pixel-locking effect for global stereo methods that interpolate not only similarity costs but also smoothness costs. There, deviations of up to 0.3 px occur (cf. Fig. 7).

The attainable accuracy of sub-pixel estimation is limited. While theoretical publications predict accuracies of 0.05 px, even perfectly calibrated systems generate accuracies of about 0.25 px on average for all measured points when using spatially discrete stereo methods.

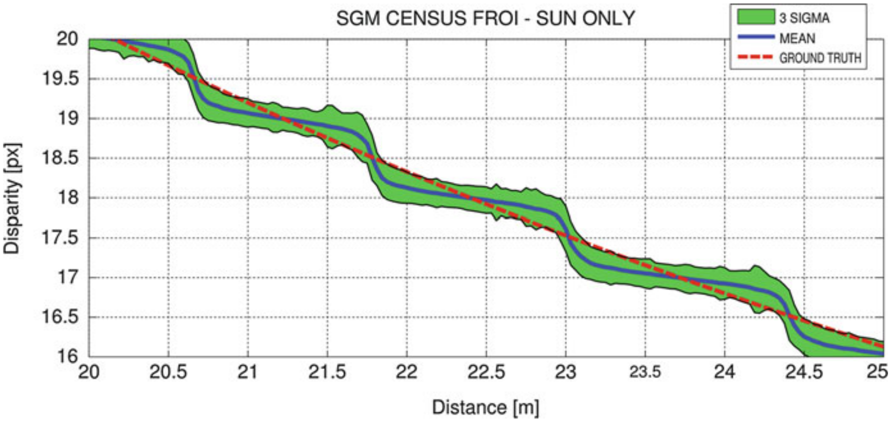


Fig. 7 Disparity progression approaching a planar obstacle. The reference disparity generated by a rendered scene and the measured disparity using SGM are shown. Deviations of up to 0.3 px occur due to the smoothness term that implicitly favors integer disparities

With more computational effort, better estimates can be generated when necessary. In Gehrig et al. (2012) for instance, the energy optimization is conducted on a sub-pixel level (e.g., in quarter pixel steps).

2.2 Effects of Decalibration

So far, the sub-pixel analysis of disparity errors assumed an ideal camera system with perfect compensation of all lens errors and camera orientation errors. Methods to determine these parameters, i.e., to perform a calibration, are introduced in ▶ Chap. 20, "Fundamentals of Machine Vision".

All proposed lens models can only approximately represent the reality, and the calibration methods have residual errors, even when the system is designed and analyzed very carefully. Residual errors in relative roll and pitch angle affect the correspondence analysis directly.

Roll and pitch angle errors: As soon as the corresponding points do not lie on the same line after rectification, the correspondence search is affected. For mostly vertical structures such as poles, the problem is not apparent. The effect is amplified for mostly horizontal structures and leads to invalid or even worse, to wrong disparities. An extreme case with mostly horizontal structure is shown in Fig. 8. At the top, the correct result with perfect calibration is shown. At the bottom, the result with only 0.2 px error in epipolar geometry is shown. The wall is triangulated significantly further away than it really is.

Squint angle error. Even more dramatic are small errors in squint angle or relative yaw angle of the camera system that generate a disparity offset Δd.

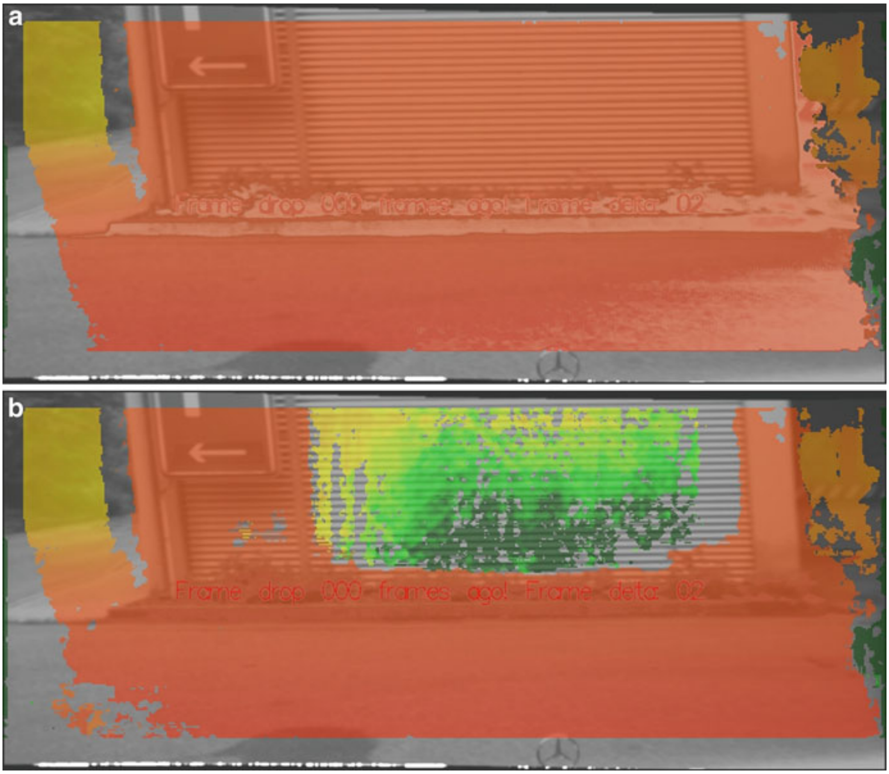


Fig. 8 Disparity map of a frontoparallel wall with horizontal structure with correct calibration (*left*) and with small decalibration (0.2 px in relative tilt angle) which causes the wall to be a very distant obstacle

If the object is to be found at distance Z with disparity d, one obtains an estimate $\hat{Z}$ via

$$\hat{Z} = \frac{Z}{\left(1 + \frac{\Delta d}{d}\right)} \tag{10}$$

Figure 9 shows that this effect is not negligible for larger distances. The shown curves are valid for the stereo camera geometry described above. If one underestimates the disparity by 1 pixel at 60 m distance, the distance is overestimated by 20 m! If such an error is not detected, it can cause problems in a sensor fusion stage.

This effect is amplified when velocities are computed based on the distance change (cf. Sect. 3). Let the relative velocity be v_{rel}, then one obtains for the estimate $\hat{v}_{\text{rel}}$:

$$\hat{v}_{\text{rel}} = \frac{v_{\text{rel}}}{\left(1 + \frac{\Delta d}{d}\right)^2} \tag{11}$$

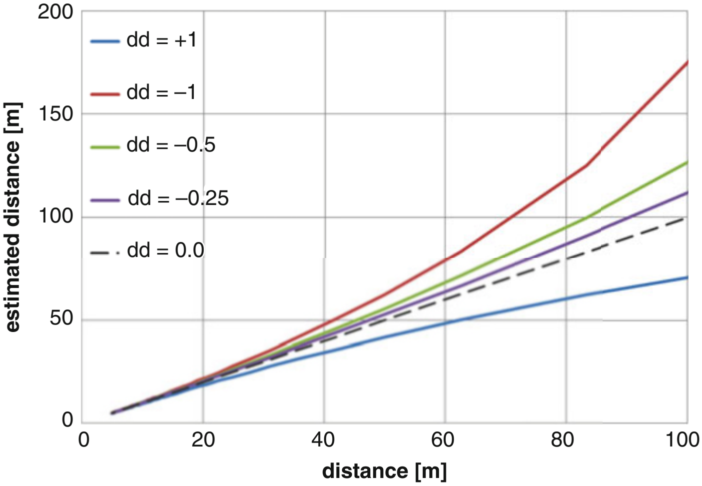


Fig. 9 Estimated distance over true distance as a function of disparity (squint angle) error Δd

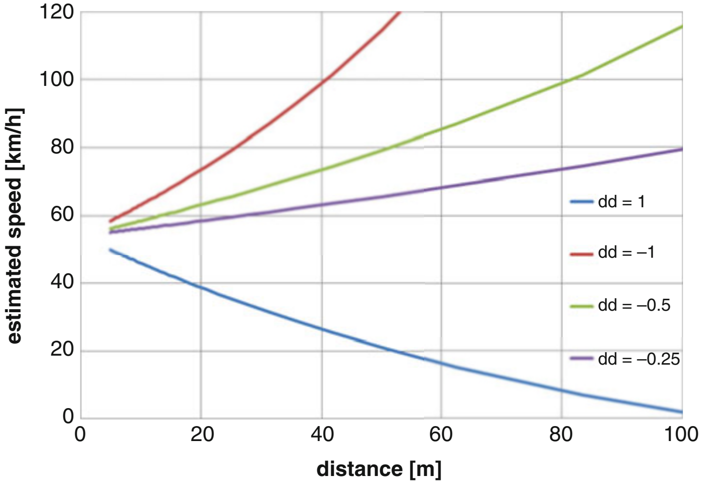


Fig. 10 Estimated velocity of an oncoming car driving 50 km/h with the observer moving at the same speed. At larger distances, the errors increase dramatically

Most of the time, the absolute velocity is searched. Subtracting the (known) ego-velocity on obtains

$$\hat{v}_f = v_{\text{ego}}\left(1 - 2\frac{\hat{v}_{\text{rel}}}{v_{\text{rel}}}\right) \tag{12}$$

If one follows a leading car with small relative velocity, this error is negligible. This is very different for oncoming cars as shown in Fig. 10: The estimated velocity

Fig. 11 Critical intersection scenario – the time to collision for the running kid behind the car is 1 s

is shown over the vehicle distance when both cars approach each other with 50 km/h. If the disparity offset is 1 px, velocities well above highway speeds are measured for trucks in urban scenarios.

This observation shows that squint angle errors are critical in stereo camera systems, especially for velocity estimation. The outer orientation of a stereo rig is changing due to aging and moreover due to temperature fluctuations. This makes an online calibration algorithm necessary for in-vehicle usage. By comparing with a reference sensor, it is possible to keep the disparity offset well below 0.1 px which is tolerable even for critical scenarios.

3 6D-Vision

Ideally, the disparity estimation delivers an unbiased estimate of the 3D position. Driver assistance systems, especially emergency brake systems, need a reliable detection of moving objects along with an accurate motion estimate and a detection confidence measure to assess the criticality of a traffic situation. The intuitive approach to extract objects from the disparity map and track them over time has not proven to be robust enough. The 3D object separation capability is very limited at larger distances due to the quadratically decreasing measurement accuracy. In an example shown in Fig. 11, the car and the child cannot be separated and the criticality of the situation might be overlooked.

Hence, it is necessary to estimate the three-dimensional motion of the image points directly in order to allow the detection of moving objects without error-prone prior segmentation. An example for a suitable algorithm is presented in the following.

3.1 The Principle

A substantial benefit of the camera sensor lies in the capability of finding corresponding points from the previous frame again, i.e., the points can be tracked.

Optical flow methods and feature tracking methods have been widely researched for many years and are well understood. For every corresponding image point pair with 3D information, one could determine the motion vector by simple differentiation. However, due to the distance uncertainty of the considered measurements and the small time difference of typically 40 ms, the obtained motion estimations are extremely noisy. This behavior can be improved by increasing the time basis to larger time differences, but this is counterproductive for fast responses when objects appear suddenly.

The central idea of the 6D-Vision concept (Franke et al. 2005) is to track relevant image points over several frames and to reduce the measurement uncertainty by temporal integration. The goal is to deliver an optimal estimate of the motion for every frame. For that purpose, the pixels are modeled as objects with mass, which move with constant velocity through space.

This allows to estimate the pixel motion by means of the Kalman filter (see ▶ Chap. 20, "Fundamentals of Machine Vision"). The 3D position $\vec{p} = (X, Y, Z)^T$ of an observed image point and its velocity vector $\vec{v} = \left(\dot{X}, \dot{Y}, \dot{Z}\right)^T$ are stacked to a six-dimensional state vector $(X, Y, Z, \dot{X}, \dot{Y}, \dot{Z})^T$. After a time interval Δt, the new position at time k+1 is

$$\vec{p}_{k+1} = R\vec{p}_k + \vec{T} + \Delta t R \vec{v}_k. \tag{13}$$

R denotes the rotation and T the translation of the scene, i.e., representing the inverse camera motion. For the velocity vector assuming constant motion, one obtains

$$\vec{v}_{k+1} = R\vec{v}_k \tag{14}$$

With that, the time-discrete linear system of the Kalman filter is

$$\vec{x}_k = A_k \vec{x}_{k-1} + B_k + \vec{\omega} \tag{15}$$

with the mean-free Gaussian noise term $\vec{\omega}$. The state transition matrix A_k is

$$A_k = \begin{bmatrix} R_k & \Delta t_k R_k \\ 0 & R_k \end{bmatrix} \tag{16}$$

and the control matrix is given by

$$B_k = \begin{bmatrix} \vec{T}_k & 0 \\ 0 & 0 \end{bmatrix} \tag{17}$$

The measurement vector $z = (u, v, d)^T$ consists of the tracker's current image position $(u, v)^T$ and the disparity measured by the stereo system. The easy-to-linearize measurement model reads as

$$z = \begin{bmatrix} u \\ v \\ d \end{bmatrix} = \frac{1}{Z} \begin{bmatrix} Xf \\ Yf \\ bF \end{bmatrix} + \vec{v} \tag{18}$$

The compensation of the ego-motion causes a correct measurement of $v = 0$ for stationary points. If the vehicle is equipped with suitable inertial sensors, the sensor values for translation and rotation can be used directly. If a simple planar rotation is assumed relying solely on a yaw rate sensor, pitch motion is not measured, which results in a misinterpretation of stationary points being vertically moving. Alternately, these parameters can be estimated via visual odometry from the camera images. The authors rely successfully on the method developed by Badino (Badino et al. 2006).

The literature offers a good selection of feature tracking methods. For driver assistance, tracking large displacements is vital, so descriptor-based feature trackers (cf. ► Chap. 20, "Fundamentals of Machine Vision") are of interest. The authors use a real-time variant of the method proposed by Stein in 2004 that is able to measure arbitrary large displacements with constant time and typically 10 % of the image points matched. Imagery with oncoming cars or driving through tight curves generates displacements of more than 100 px/frame. Another vital property of feature tracking is the insensitivity to illumination changes, which often occur in practice. If the prediction from the 6D Kalman filter is used, one can also successfully work with robust variants of the popular Kanade-Lucas-Tomasi Tracker (Shi and Tomasi 1994).

Figure 12 illustrates the performance of the described approach. The arrows depict the motion state of the considered pixels and point to the predicted position in 500 ms. For viewing reasons, every other frame is skipped so the time difference is 80 ms for the shown images. 200 ms are sufficient to robustly detect critical situations of that type, as proven by many experiments. The ego-vehicle from Fig. 12 drove about 30 km/h and would have collided with the pedestrian without intervention.

Another example is depicted in Fig. 13. The scenario shows a cyclist making a left turn in front of the approaching ego-vehicle. Here, the assumption of partial linear motion is confirmed: Although the cyclist turns (circular motion), the orientation of the estimated motion agrees well with reality.

The Kalman filter introduced above delivers position and velocity estimates of the tracked image point. For initialization, the position is readily available via stereo measurement. The initial velocity cannot be determined from a single position measurement. It must be computed with suitable velocity hypotheses and velocity variances. This aspect is covered by Franke et al. (2005) in more detail.

3.2 Dense6D

The previous section introduced 6D-Vision for isolated image points that can be localized in space and time. For reasons of robustness and accuracy, a high

Fig. 12 Four enlarged image sections show the result of the 6D-Vision estimate for the situation from Fig. 11. The arrows point to the predicted position in 500 ms

measurement density is desirable. Ideally, a measurement of position and motion vector is available for all pixels. For that purpose, both optical flow and stereo methods should determine results for every pixel. The SGM method described above is able to deliver depth information for almost every pixel. Several dense optical flow schemes have been published recently. However, they were not used due to high computational demands and lack of robustness.

Classical methods to determine dense optical flow use the "Constant Brightness Assumption," i.e., the intensity of corresponding points is assumed to be identical. From that, a data term is generated that minimizes the costs by varying the displacements. Similar to global stereo methods, a smoothness term is also introduced that penalizes changes of displacements of neighboring pixels. For computational reasons, smoothness deviations were penalized quadratically in older publications which

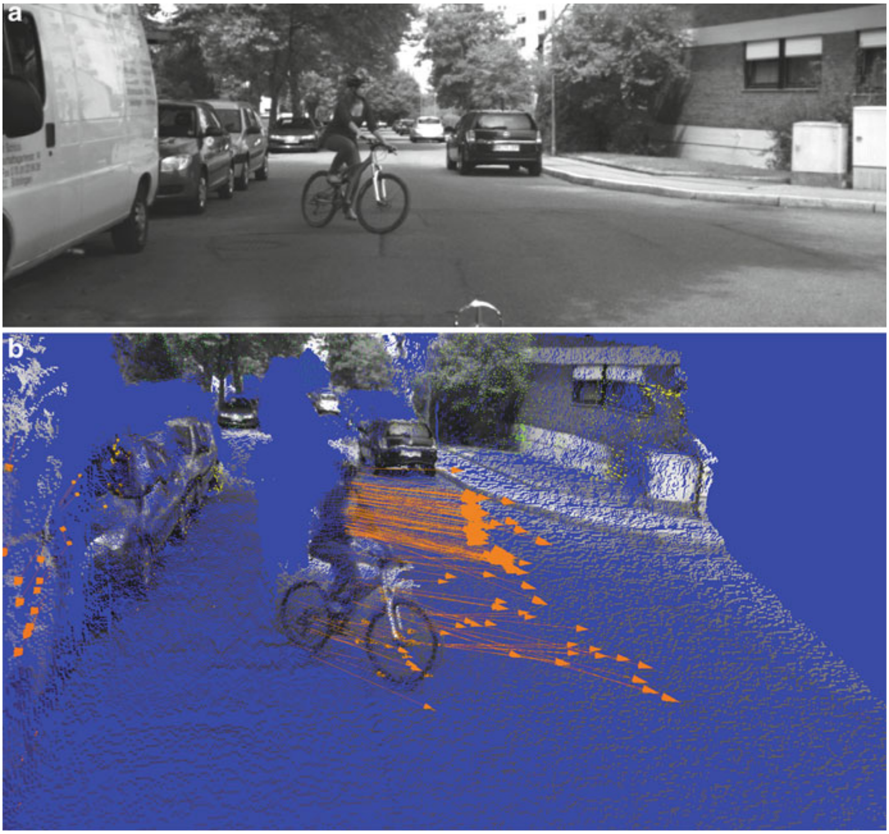


Fig. 13 Turning cyclist scenario depicted at the top. At the bottom, the 6D-Vision interpretation shown in a 3D viewer. The *arrows* point to the predicted position in 500 ms. The colors encode distance (*red* near … *green* far)

resulted in high sensitivity to outliers. Especially, the constant displacement assumption is not fulfilled for traffic situations and leads to oversmoothing.

Zach et al. (2007) proposes a dense optical flow method that can be computed in real-time on high resolutions. They employed the computing power of modern graphics card (GPU). The implemented TV-L1 method uses total variation and penalizes smoothness deviations linearly, not quadratically. This results in more accurate flow fields, but is still error prone to illumination changes. Müller et al. (2011b) shows that dense optical flow can be robustly computed under illumination changes by using the census operator.

Despite this progress, the TV-L1 method tends to oversmooth the flow field and exhibits difficulties with large displacements. Müller (Müller et al. 2011a, b) proposes two extensions to improve the method: Firstly, to estimate large displacements, measurements of a local flow method are used and, secondly, the

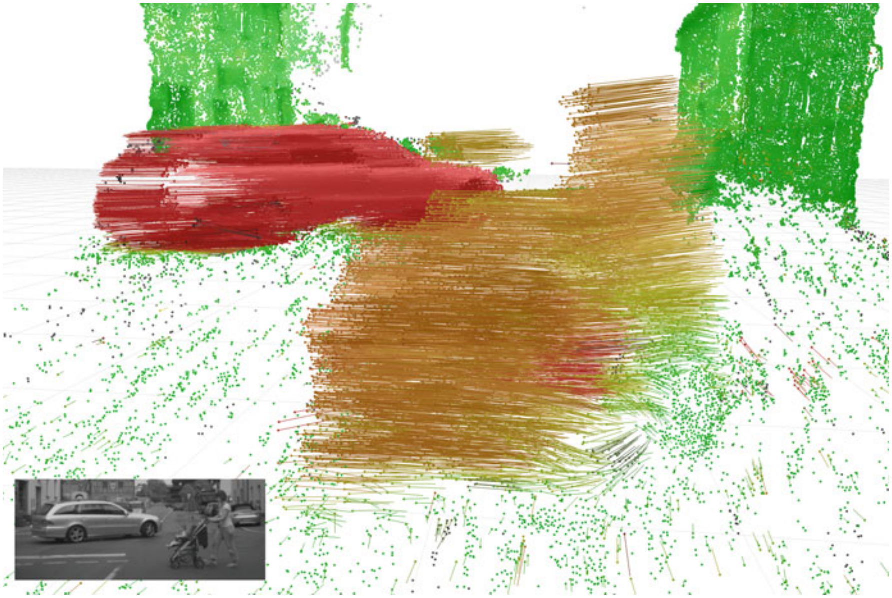


Fig. 14 Dense 6D result for a crossing scenario. Colors encode the velocity magnitude (*green* slow . . . *red* fast). The vectors depict the predicted position in 500 ms

smoothness constraint is modified to penalize the deviation from the expected flow field determined by stereo and ego-motion information under the assumption of a static world.

When optical flow and stereo information is available for almost every pixel, the 6D-Vision concept can be applied. Rabe et al. (2010) presented a real-time variant under the name Dense6D in 2010. The Kalman filters are organized as iconic filters in 2D, similar to the image structure. By computing the flow field from the current to the previous frame for every pixel, one can identify the predecessor and continue the tracking chain.

Another type of method to determine both motion and depth for every pixel is called scene flow, where disparity change and optical flow are estimated simultaneously. From the disparity change and the original disparity, one can compute the velocity relative to the observer. As Rabe et al. (2010) showed in their experiments, the velocity field obtained by these differential methods is significantly inferior to the Dense6D result. It seems that one gains more from the strong temporal smoothness of the 6D approach than from the optimized spatial smoothness of the scene flow approach.

A Dense6D example result is shown in Fig. 14.

4 The Stixel World

Thanks to the real-time capability of dense stereo methods, subsequent processing steps can use about 500,000 3D points per frame (as of 2014). In the upcoming years, the number of 3D points is expected to increase from one to two million

Fig. 15 Representation of the scene from Fig. 5 with Stixels. Distances are color coded; *white rectangles* define the Stixel outline. The *white arrows* indicate speed and motion direction

points. At the same time, more detection modules for pedestrians, cyclists, vehicles, stationary obstacles, free space, road height profile, etc., will use this information to improve their performance. This will lead to extreme demands on computation power and bandwidth.

This problem can be circumvented with a more compact representation that clusters image points to superpixels. A superpixel method suitable for traffic scenes is the Stixel World, proposed by Badino et al. (2009) and Rabe et al. (2010) in 2009. As shown in Fig. 15, the complete 3D information is approximated by a few hundred thin rectangular sticks defined by base point, distance, and height. If these Stixels are tracked over time following the 6D-Vision concept, their motion state is available as well. Since the number of Stixels is small, moving object segmentation can be performed with a global optimal method. The steps from pixels via Stixels to objects are described in the following.

4.1 Optimal Computation of the Stixel World

Traffic scenes are dominated by (approximately) horizontal and vertical planes. The most prominent plane is the street or ground plane. Objects such as cars, pedestrians, and infrastructure stand on this ground plane. Only in rare cases, e.g., for bridges, objects do not touch the ground. The Stixel World exploits this property to generate a compact and robust representation of the scene.

The most powerful method to date to compute this representation was introduced by Pfeiffer et al. (Pfeiffer and Franke 2011) in 2011. There, the computation was cast to a maximum a posteriori (MAP) problem solved by dynamic programming. For that purpose, small stripes of 5–9 pixels are considered independent of their neighbors, resulting in independent one-dimensional optimization tasks.

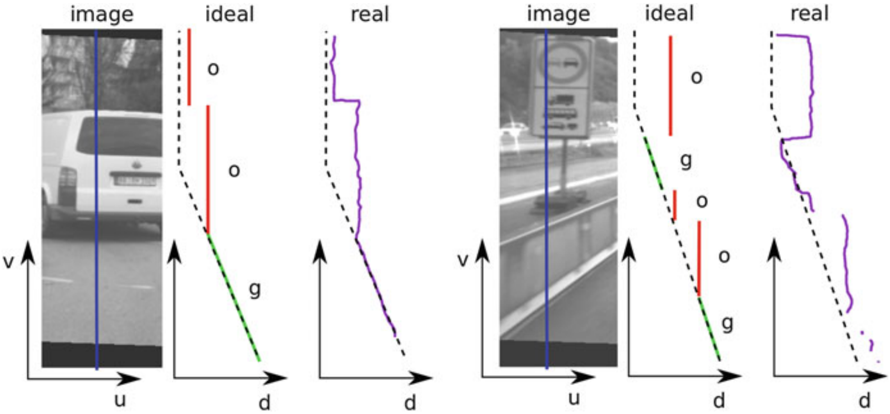


Fig. 16 Concept of the Stixel World: The purple disparity curve along a column is approximated by constant parts (objects, shown in *red*) and linear parts following the ground layout (ground, shown in *green*)

Figure 16 illustrates the basic idea of the Stixel World giving two examples. The blue line marks the considered image column/stripe. The measured disparities are shown in violet. A segmentation into the classes "object" and "ground" is sought. Pixels of an object have approximately the same disparity, while pixels of the class "ground" have a linear disparity progression given by the camera geometry. The left case depicts the standard case of a dominating foreground object in front of a background object. For the right example, we expect a representation of three objects and two parts labeled as "ground."

The task of the optimization step is to find the most probable approximation of the disparity measurement vector D complying with the model. So we seek the most probable labeling L^*, formally

$$L^* = \arg\max_L P(L|D) \tag{19}$$

Applying Bayes rule, the a posteriori probability distribution can be converted to $P(L|D) = \frac{P(D|L)P(L)}{P(D)}$. With that, the optimization can be written as

$$L^* = \arg\max_L P(D|L)P(L) \tag{20}$$

$P(D|L)$ represents the probability density of the observed disparity vector for any labeling L and can be considered the data term. The second term $P(L)$ accounts for the fact that not all possible object orderings are equally likely. This way, prior knowledge or statistics about the typical layout of traffic scenes can be incorporated. This represents the smoothness term of global stereo methods without limiting the term to smooth solutions. More constraints can be modeled with the following being the most powerful:

- **Bayes information criterion**: Within a column, only few objects are expected; solutions with few objects are preferred.
- **Gravitational constraint**: Hovering objects are unlikely. Objects with a base point close to the ground are preferred to touch the ground.
- **Ordering constraint**: The higher the Stixel in the image is, the further away it usually is. Bridges, trees, and other objects violating this constraint are still approximated correctly when the data term supports it sufficiently.

A more detailed description of the prior term $P(L)$ can be found in Pfeiffer and Franke (2011). Since the optimization task is a one-dimensional problem, dynamic programming can be efficiently used to solve it. For efficiency reasons, the disparity measurements are treated to be independent. The disparity distribution $P(D|L)$ is modeled as a Gaussian distribution overlaid with a uniform distribution to be robust against outliers.

Figure 16 shows the Stixel World for the scene introduced in Sect. 1. The distance is color coded as before. White vertical rectangles define the outline of the foreground Stixels; the ground plane is shown in gray. The vehicles are clearly separated from the background; the tree on the right is correctly modeled, despite violating the gravitational constraint.

In the example shown here, arrows on the ground depict velocity and motion direction of the Stixels. If the Stixels are considered “large” pixels, the 6D-Vision concept can be applied without changes. For the purpose of driver assistance, a four-dimensional state vector suffices since height and vertical motion are not important. With that, the pitch motion is not critical to estimate anymore. This “dynamic Stixel World” is the basis for subsequent high-level vision modules.

The Stixel World is an approximate representation of 3D data. It offers the following desired properties:

- **Robustness**: The implicit averaging of all disparities within a Stixel leads to disparity noise reduction. Thanks to the robust modeling, local disparity errors can be detected and eliminated. Larger occasional errors (e.g., from the windshield wiper) that are not consistent over time are removed by the temporal filtering.
- **Compactness**: The Stixel World is very compact. The 3D scene content of an image is represented by 300–600 Stixels, which can be encoded in a few kilobytes representing the full geometry of the scene including motion information.
- **Explicit representation**: The representation extracts the scene content. If the Stixel World shown in Fig. 16 is viewed without the grayscale image, humans are still able to understand the scene. The implicit classification for street and object also encodes the drivable free space, which is necessary for the trajectory planning of swerve maneuvers.

The Stixel World models 3D data and is hence not limited to disparity maps. Also, data from high-resolution laser scanners, e.g., the Velodyne HD64, can be represented that way.

4.2 Image Understanding in the Stixel World

As part of driver assistance, computer vision has to solve multiple tasks:

(a) Detection of moving objects and estimation of their motion states,
(b) Classification of objects (Pedestrians, cars, cyclists, . . .),
(c) Detection of the free space and
(d) Intention recognition of other traffic participants.

The concepts presented here are the basis of next-generation vision systems solving above tasks efficiently.

4.2.1 Stixmentation

The first task needs a segmentation of the dynamic Stixel World into moving objects and static background. A simple threshold on the Stixel velocity would ignore the fact that Stixels from one object have similar velocities. Again, global optimal methods deliver better results for this task.

Erbs et al. (2013) define the Stixels to be nodes of a conditional random field (CRF) and set up an optimization task, which is solved with GraphCut. For separating a single moving object from static background, an optimal solution of this binary problem is available. Erbs et al. describe an iterative scheme that is able to deal with an unknown number of objects and detects them reliably.

As shown in Fig. 17, the oncoming cars are detected and their motion states are estimated. It is recommended to apply a vehicle-specific tracker to these Stixel groups. Barth et al. (Barth and Franke 2009) show how to compute the complete motion state including yaw rate of oncoming cars based on such input data.

4.2.2 Object Classification

A strength of the camera sensor is its capability to detect interesting objects such as pedestrians and vehicles based on its appearance even in very complex scenes. Due to its relevance for driver assistance, a separate chapter (▶ Chap. 22, “Camera

Fig. 17 Segmentation of the dynamic Stixel World in single, independently moving objects

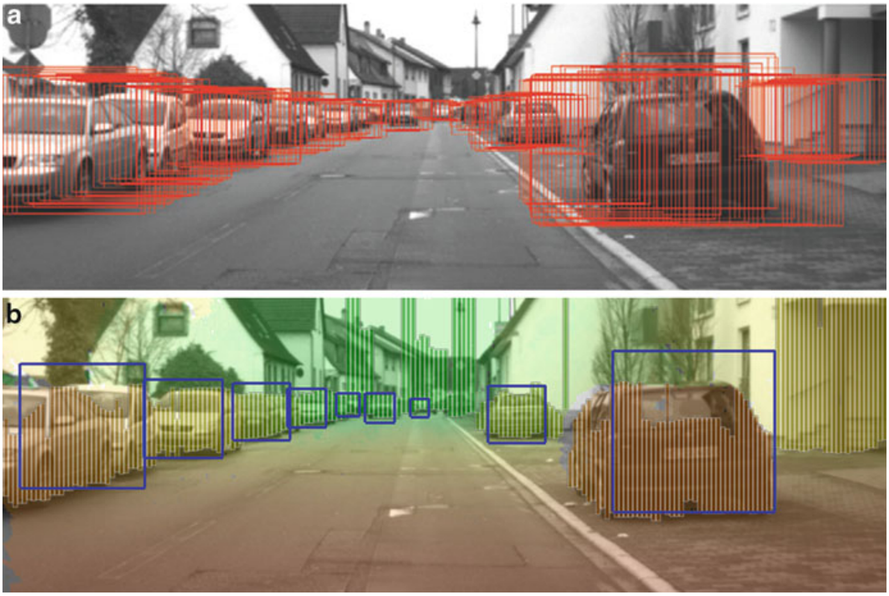


Fig. 18 (**a**) Regions of interest generated by Stixels for vehicle classification. (**b**) Final vehicle detections after non-maxima suppression

Based Pedestrian Detection") is dedicated to pedestrian detection. However, the computational effort for classification schemes increases proportional to the number of hypotheses to be checked. Therefore, powerful attention control mechanisms that place the classification hypotheses efficiently are needed.

The Stixel World allows for a very efficient reduction of hypotheses. For every Stixel, it is assumed that this Stixel is the center Stixel of an object type to be classified, e.g., a car. Distance and position of the Stixel define the size of the region-of-interest (ROI). If the Stixel height conforms to the hypothesis height, a classification is performed. Figure 18a shows the ROIs found this way. Typically, only a few hundred ROIs are generated. Enzweiler et al. (2012) show that this approach is not only efficient but also delivers better results than classic full search strategies. The reason for that is that all relevant hypotheses are tested, but far fewer requests are sent to the classifier, which inevitably ends in fewer false positive results. One obtains a system with comparable detection rate at significantly reduced false positive rates and less computation time.

The classification result is depicted in Fig. 18b. This approach easily extends to multiple classes such as pedestrians or cyclists.

4.2.3 Scene Labeling

Classification schemes operating on rectangular bounding boxes are well researched and are used in, e.g., face detection or traffic sign recognition systems. They are restricted to cases where the object of interest can be clearly marked by a

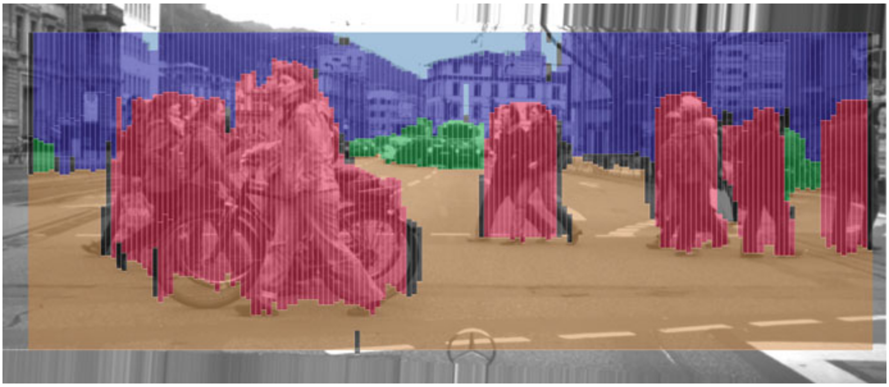


Fig. 19 Example of an automatic scene labeling result. The colors encode classes: vehicle, *green*; pedestrian, *red*; building, *violet*; and street, *brown*

box. When occlusions (e.g., a pedestrian behind a parking car), staggered object arrangements (e.g., a row of parking cars), or elongated objects (houses, guard rails) occur, such an approach comes to its limit.

The current trend in computer vision is to subdivide an image into superpixels and then determine the attributes of the superpixel such as street, vehicle, building, vegetation, etc. This discipline is called "scene labeling." The assignment of attributes to superpixels is usually obtained by classification based on color and texture information.

Scharwächter et al. (2013) use the Stixel World as an efficient basis for scene labeling. The grouping of Stixels yields image segments that coincide better with object boundaries compared to appearance-based superpixel algorithms. Height information is used in addition for the classification step. For an efficient feature, encoding a so-called random decision forest (Moosmann et al. 2007) is applied. The scene labeling results defined the state of the art at time of publication.

Figure 19 shows the scene labeling result using the Stixel World. The attributes used here are street, vehicle, pedestrian, building, and sky. The choice of classes is application specific. For highway scenarios, the "guard rail" class is more relevant than the "building" class.

5 Summary

The presented methods of stereo image processing are the result from years of research and have proven to be powerful in practice. Semi-global matching and the 6D-Vision concept became available in premium and midsize cars at Mercedes Benz in 2013. The applied classification methods benefit heavily from the dense stereo information. The false alarm rate is reduced by a factor of five at the same

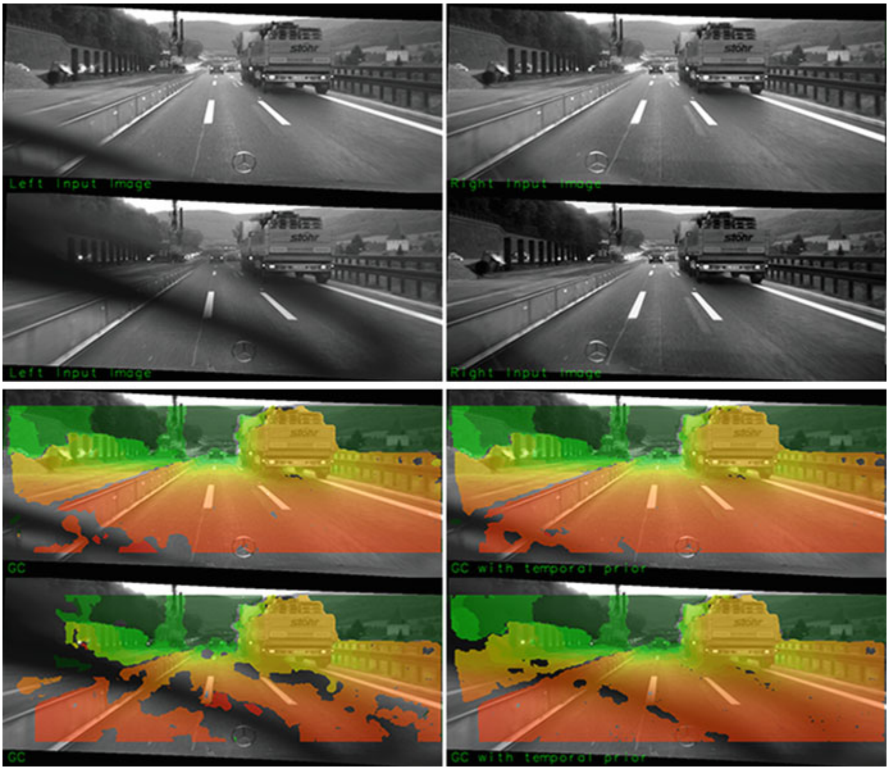


Fig. 20 The *top* row shows from *left* to *right*: a stereo image pair recorded during rain, the single-frame disparity map, and the disparity map with temporal smoothing. The *bottom* row shows the situation one frame later when the windshield wiper blocks the *left* image. The improvements of the method from Gehrig et al. (2013) are striking, resulting in an almost error-free disparity map

detection rate (Enzweiler and Gavrila 2011). Here, the advantage of a stereo-based approach compared to a monocular approach is clearly visible.

Also, the Stixel World and its subsequent processing steps have proven their effectiveness. The algorithms formed the basic scene understanding module within the Bertha-Benz autonomous drive in August 2013. There, a Mercedes S500 intelligent drive vehicle drove autonomously the famous route of Bertha-Benz first overland drive from Mannheim to Pforzheim, Germany, with sensors close to serial production. Due to the compact representation of the environment, computation-intense methods can be run without runtime problems (Franke et al. 2013). Autonomous driving is the focus of ► Chap. 62, "Autonomous Driving".

The vision of autonomous driving and the desire for more elaborate driver assistance systems will further drive the demands for more accuracy and robustness of stereo vision algorithms. New approaches for disparity estimation (Yamaguchi et al. 2013) or for scene flow (Vogel et al. 2013) are not real-time capable, but perform excellently on the KITTI benchmark.

A different approach is the temporal coupling of disparity estimation (Gehrig et al. 2013) targeted for extreme weather conditions (cf. Fig. 1). It reduces false-positive rates of obstacle detection while increasing obstacle detection rate. Figure 20 shows two consecutive stereo image pairs on the left and the disparity images without (second to right) and with temporal coupling (right).

The market penetration of stereo cameras will increase for modern vehicles in upcoming years, since many object detection tasks can be solved more efficiently and robustly than with monocular systems. Today's FPGA implementations will be replaced by ASIC solutions, and hence high-resolution disparity images can be generated with very low power consumption. The increasing sensor resolution will both enable larger detection ranges and larger fields of view.

References

Badino H, Franke U, Rabe C, Gehrig S (2006) Stereo vision-based detection of moving objects under strong camera motion. VisApp, Portugal

Badino H, Franke U, Pfeiffer D (2009) The stixel world – a compact medium level representation of the 3D-world. In: DAGM symposium, Jena, Ger

Banz C, Blume H, Pirsch P (2011) Real-time semi-global matching disparity estimation on the GPU. In: ICCV workshop on mobile computer vision, Barcelona

Barth A, Franke U (2009) Estimating the driving state of oncoming vehicles from a moving platform using stereo vision. IEEE Trans ITS 10:560–571

Belhumeur N (1996) A Bayesian approach to binocular stereopsis. Int J Comput Vis 19(3): 237–260

Bleyer M, Chambon S (2010) Does color really help in dense stereo matching? In: International symposium 3D data processing, visualization and transmission (3DPVT), pp 1–8, Paris

Boykov Y, Veksler O, Zabih R (2001) Fast approximate energy minimization via graph cuts. IEEE Trans Pattern Anal Mach Intell 23(11):1222–1239

Dickmanns ED, Zapp A (1986) A curvature-based scheme for improving road vehicle guidance by computer vision. SPIE 727:161–168

Enzweiler M, Gavrila DM (2011) A multi-level mixture-of-experts framework for pedestrian classification. IEEE Trans Image Process 20:2967–2979

Enzweiler M, Hummel M, Pfeiffer D, Franke U (2012) Efficient stixel-based object recognition. In: IEEE intelligent vehicles symposium, Alcala de Henares

Erbs F, Schwarz B, Franke U (2013) From stixels to objects. In: IEEE intelligent vehicles symposium, Brisbane

Franke U, Rabe C, Badino H, Gehrig S (2005) 6D-vision: fusion of stereo and motion for robust environment perception. In: DAGM symposium, Wien

Franke U, Pfeiffer D, Rabe C, Knöppel C, Enzweiler M, Stein F, Herrtwich RG (2013) Making Bertha See. In: ICCV workshop computer vision for autonomous driving, Sydney

Gehrig S, Rabe C (2010) Real-time semi-global matching on the CPU. In: CVPR embedded computer vision workshop, San Francisco

Gehrig S, Eberli F, Meyer T (2009) A real-time low-power stereo engine using semi-global matching. In: International conference on computer vision systems, Liege

Gehrig S, Badino H, Franke U (2012) Improving sub-pixel accuracy for long range stereo. J Comput Vis Image Underst 116(1):16–24

Gehrig S, Reznistkii M, Schneider N, Franke U, Weickert J (2013) Priors for stereo vision under adverse weather conditions. In: ICCV workshop computer vision for autonomous driving, Sydney

Geiger A, Lenz P, Urtasun R (2012) Are we ready for autonomous driving? The KITTI vision benchmark suite, computer vision and pattern recognition (CVPR), Providence, RI, USA

Hirschmüller H (2008) Stereo processing by semi-global matching and mutual information. IEEE Trans Pattern Anal Mach Intell 30(2):328–341

Hirschmüller H, Gehrig S (2009) Stereo matching in the presence of sub-pixel calibration errors. Computer vision and pattern recognition (CVPR), Miami, FL, USA

Hirschmüller H, Scharstein D (2009) Evaluation of stereo matching costs on images with radiometric differences. IEEE Trans Pattern Anal Mach Intell 31(9):1582–1599

Meister S, Jähne B, Kondermann D (2012) Outdoor stereo camera system for the generation of real-world benchmark data sets. J Opt Eng 51(02):021207–1

Moosmann F, Triggs B, Jurie F (2007) Fast discriminative visual codebooks using randomized clustering forests. In: Bernhard Schölkopf, John C. Platt, Thomas Hoffman (eds) Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 4-7, 2006. MIT Press 2007, ISBN 0-262-19568-2

Mühlmann K, Maier D, Hesser J, Männer R (2001) Calculating dense disparity maps from color stereo images, an efficient implementation. In: Stereo and multi-baseline vision, 2001 (SMBV 2001). Proceedings. IEEE workshop on, Kauai, 9–10 Dec 2001, pp 30–36. IEEE. doi:10.1109/SMBV.2001.988760

Müller T, Rannacher J, Rabe C, Franke U (2011a) Feature- and depth-supported modified total variation optical flow for 3D motion field estimation in real scenes. Computer vision and pattern recognition (CVPR), Colorado Springs, CO, USA

Müller T, Rabe C, Rannacher J, Franke U, Mester R (2011b) Illumination-robust dense optical flow using census signatures. In: DAGM symposium, Frankfurt, Ger

Pfeiffer D, Franke U (2011) Towards a global optimal multi-layer stixel representation of dense 3D data. In: British machine vision conference (BMVC), Dundee, UK

Rabe C, Müller T, Wedel A, Franke U (2010) Dense, robust, and accurate motion field estimation from stereo sequences in real-time. In: European conference on computer vision (ECCV), Heraklion, Gr

Scharstein D, Szeliski R (2002) A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Comput Vis (IJCV) 47(1):7–42, Kluwer

Scharwächter T, Enzweiler M, Franke U, Roth S (2013) Efficient multi-cue scene segmentation. In: DAGM symposium, Saarbrücken

Shi J, Tomasi C (1994) Good features to track. Computer vision and pattern recognition (CVPR), Seattle

Shimizu M, Okutomi M (2001) Precise sub-pixel estimation on area-based matching. In: International conference on computer vision (ICCV), Vancouver, CA

Stein F (2004) Efficient computation of optical flow. In: DAGM symposium, Tuebingen, Ger

Tappen M, Freeman W (2003) Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In: International conference on computer vision (ICCV), Nice, Fr

Vogel C, Roth S, Schindler K (2013) Piecewise rigid scene flow. In: International conference on computer vision (ICCV), Sydney, AUS

Yamaguchi K, McAllester D, Urtasun R (2013) Robust monocular epipolar flow estimation. Computer vision and pattern recognition (CVPR), Portland, OR, USA

Zach C, Pock T, Bischof H (2007) A duality based approach for realtime TV-L1 optical flow. In: DAGM symposium, Heidelberg, Ger

Camera Based Pedestrian Detection

22

Bernt Schiele and Christian Wojek

Contents

Abstract

Detecting pedestrians in street scenes is one of the most important but also one of the most difficult problems of computer vision. Ideally, all pedestrians should be robustly detected in order to provide optimal assistance to the driver regardless of visual conditions. Different environmental factors complicate this, however. Especially problematic are changing weather and visual conditions as well as difficult lighting situations and road conditions. Moreover, an individual's clothing and partial occlusions of pedestrians, for example, by parked cars,

---

B. Schiele (✉)
Computer Vision and Multimodal Computing, Max Planck Institute for Informatics, Saarbrücken, Germany
e-mail: schiele@mpi-inf.mpg.de

C. Wojek
Carl Zeiss AG, Oberkochen, Germany
e-mail: wojek@zeiss.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_23

further complicate the detection task. Also, in comparison to many other objects in street scenes, pedestrians are characterized by a high degree of articulation further complicating the task.

1 Introduction

Essentially, two types of detection tasks can be distinguished, depending on the type of sensor used:

- Video-based methods – during daylight
- Infrared camera-based methods – at night

Whereas sensors vary according to the light spectrum that is captured, similar principal processing methods have proved valuable in practice.

As mentioned above, a system used for robust detection of pedestrians in street scenes must meet high requirements. The following aspects are of particularly high importance:

- **Resolution and scale of pedestrians in the image:** The video image resolution and the focus of the camera used substantially affect the amount of representable information (cf. ▶ Chap. 19, “Automotive Camera (Hardware)”). While even people have difficulty detecting pedestrians in images of low resolution, it is possible to determine their position and their pose in images of high resolution. Consequently, different methods and models should be preferred for different ranges of detection and system functions. Figure 1 shows a scene with a pedestrian at different resolutions recorded by a standard onboard camera.
- **Robustness:** Robustness plays a crucial role in application scenarios. In particular, functionality needs to be achieved in various weather and visual conditions. At the same time, systems for pedestrian detection have to operate regardless of

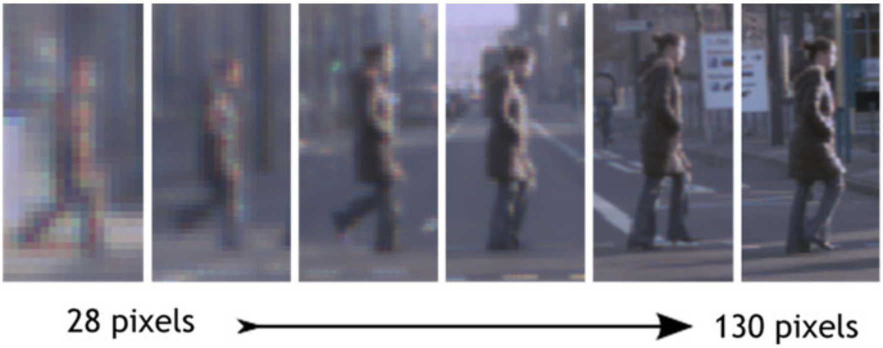


Fig. 1 Urban traffic scene with a pedestrian at varying resolution (normalized representation)

a pedestrian's clothing and articulation. This also influences choosing the right sensor. While pedestrians can be detected well in visible light during the day, visibility already decreases at twilight. Infrared cameras, however, can also register parts of the invisible light spectrum. Whereas background structures often have a similar signature as pedestrians during the day, pedestrians can be detected clearly at night due to the heat radiation they emit, making infrared cameras preferable over standard cameras.

- **Viewpoint invariance:** Pedestrians must be detected regardless of the angle of the camera relative to the pedestrian.
- **Partial occlusion:** The occlusion of pedestrians can hardly be avoided in realistic applications. Especially in complex urban scenarios, a functioning system must deal with these situations.
- **Pose estimation:** Pose estimation is necessary for the determination of the pedestrian's direction of movement. Especially when there is not much time before a collision, and thus the pedestrian is only a short distance from the vehicle, this aspect is of great importance for achieving a reasonable reaction strategy.
- **2D versus 3D modeling:** Although 2D approaches that model the environment in image coordinates achieve good results for small pedestrians at a greater distance, there remains uncertainty regarding their exact position in relation to the own vehicle. Modeling in global coordinates, especially for pose estimation, is therefore desirable for the close range.

2 Possible Approaches

There are three basic types of approaches in the literature on pedestrian detection:

- Sliding-window approaches
- Interest point- and body part-based approaches
- System-oriented approaches

Sliding-window methods successively move a window of predefined fixed size over the input image, independently determining for each sliding window, by means of a classifier, whether there is a pedestrian or not. In order to achieve scale invariance, i.e., to make the pedestrian's size in the input image independent of the size of the classifying window, the input image is rescaled and retested until it is smaller in dimension than the detection window (cf. Fig. 2).

Approaches using gradient histograms to enable generalization for various instances are especially popular (Shashua et al. 2004; Dalal and Triggs 2005; Maji et al. 2008). The major limitation of these methods is the global rigid description of a pedestrian by means of a window with a predefined aspect ratio. This can be countered, for example, by dividing the window into several parts.

In many cases, AdaBoost (Friedman et al. 2000) and support vector machines (SVMs) are used as classifiers (Schoelkopf and Smola 2001).

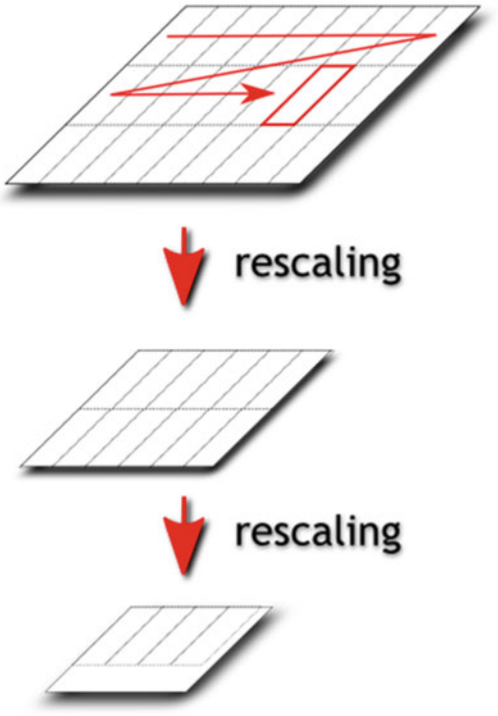


Fig. 2 Sliding-window object detection

AdaBoost is a classification method that obtains a so-called “strong” classifier by a weighted sum of weak classifiers. Weak classifiers are often decision tree stumps with a single decision node. In each training round a weak classifier is chosen that separates the training data along the most discriminative dimension.

In contrast, support vector machines optimize the global classification mistake by defining a hyperplane to separate training data optimally according to statistical learning theory. At the same time, so-called kernels can be used to separate the training data with a nonlinear decision surface.

In interest point-based approaches, distinctive image areas are first extracted. These can be corner points with a high intensity gradient in two directions (Harris and Stephens 1988) or circular areas (Lowe 2004). Subsequently, a canonical scale can be established by means of the Laplace function (Lindeberg 1998). The established points are then further characterized by so-called feature descriptors (Mikolajczyk and Schmid 2004) and combined to a person model. This group of methods comprises, for example, the *Implicit Shape Model (ISM)* by Leibe et al. (2005), Seemann et al. (2006), Seemann and Schiele (2006), and Andriluka et al. (2008). A comparison of various descriptors can be found in Seemann et al. (2005).

Closely related to this are the body part-based approaches, where an attempt is made to separately detect individual body parts like the extremities and torso. These are then combined through a probabilistic model. The advantage of these approaches is the robustness against partial occlusion and a good generalization for different articulations.

Finally, the system-oriented approaches deserve mentioning here. In contrast to the models above, they use prior knowledge regarding a concrete application in the automotive environment to construct a system. An example of this is the assumption of a ground plane on which vehicles as well as pedestrians move. Moreover, a computational step that automatically determines regions of interest in the image is often included. The PROTECTOR system by Gavrila and Munder (2007) is a prominent representative of this group.

The sliding-window-based methods dominate in the area of infrared camera-based pedestrian detection at night. Mählisch et al. (2005) and Suard et al. (2006) adapt similar approaches and features that have already proven valuable for the visible area of the spectrum. Bertozzi et al. (2007) also use heat radiation characteristics for pedestrian detection.

3 Description of Operating Principle

As mentioned above, depending on the video image resolution, different methods can be suitable for the requirements described. In the following, one work out of each category is described in more detail.

3.1 Sliding-Window Approaches

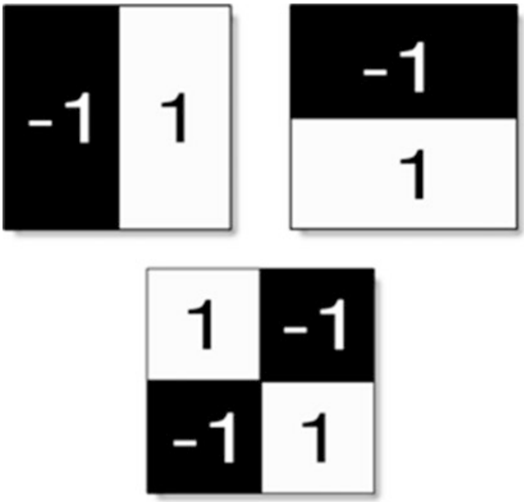
The performance analysis by Wojek and Schiele (2008) is presented in greater detail for the method group of sliding-window approaches.

An important difference between the methods is that different features are used. The applied classification methods also vary including AdaBoost and SVMs with different kernels. Table 1 gives an overview of the combinations in the original papers.

Table 1 Combinations of features and window classifiers

Feature/ classifier	SVM with kernel	AdaBoost	Others	Evaluation criteria
Haar wavelets	Polynomial kernel			ROC
Haar-like wavelets		Cascaded with decision trees		ROC
HOG	Linear and RBF kernel			FPPW
Shapelets		With decision trees		FPPW
Shape context			ISM	RPC

Fig. 3 Haar filter bank example



As can be seen from this table, many possible feature/classifier combinations were not evaluated. In this context, various combinations should be compared exhaustively on an established data set with a 64 × 128 pixel detection window. To begin with, the applied features are briefly presented.

Haar process latent variable in Papageorgiou and Poggio (2000) encode local image intensity differences. The applied filter mask scale is 16 and 32 pixels with an overlap of 75 % for the individual masks, thus enabling an overcomplete representation.

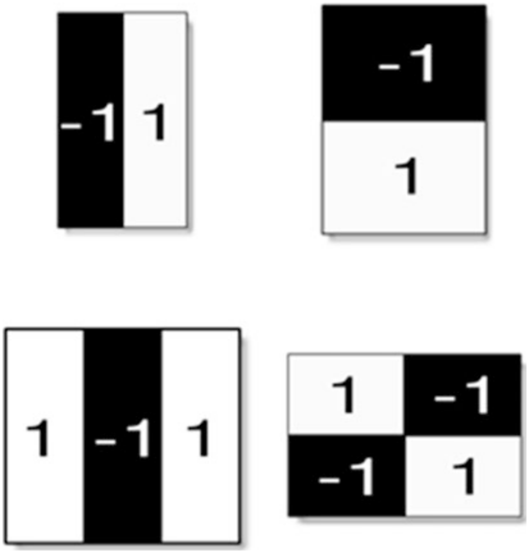
The filters shown in Fig. 3 (second to fourth basic function) are used; self-similarity (first basic function) is neglected. In order to cope with differences in illumination, all individual responses are normalized by the mean filter response for the corresponding filter type. Additionally, only the strength of the filter response is of significance, due to the variety in clothing. Further improvement can be achieved by a global L_2 length normalization.

Haar-like features in Viola et al. (2003) constitute a generalization of the Haar features to general rectangular features, which can occur at any location and in any size in the detection window (cf. Fig. 4).

Discriminative features are selected by AdaBoost during training. This is based on efficient feature computation by integral images. The exponentially increasing number of possible feature positions and sizes constitutes the limiting factor of this feature. Therefore, for the following evaluation, features for a 24 × 48 pixel window are determined and then scaled to the detection window size. This also shows that neglecting the sign of the filter responses is an advantage. Moreover, a L_2 length normalization of selected feature responses is superior to average and variance normalization.

Histograms of oriented gradients (HOG) have been proposed by Dalal and Triggs (2005) as a further feature. The gradients are initially computed in *x*- and *y*-

Fig. 4 Haar-like features



direction and then inserted into so-called cell histograms (of 8 × 8 pixels), interpolating in spatial coordinates and with regard to orientations. Subsequently, all cell histograms are normalized with respect to the neighboring cells to compensate for local illumination differences. An additional hysteresis step has proven useful to prevent one histogram entry from dominating (Lowe 2004). The feature vector is generated by concatenating all histogram entries (see Fig. 5).

Shapelets are another type of gradient-based feature that are automatically learned for local detection window areas (Sabzmeydani and Mori 2007). AdaBoost is employed for selecting the gradients, and it uses the gradients in multiple orientations (0°, 90°, 180°, 270°) as entries at scales from 5 × 5 pixels to 15 × 15 pixels. Figure 6 illustrates the selected discriminative gradients. Again, illumination invariance is achieved by normalizing the gradients with respect to the local neighborhood; adequate regularization is required to not amplify noise.

Shape context was originally proposed as a feature point descriptor by Belongie et al. (2002) and has shown excellent results in the ISM framework by Seemann et al. (2005). The descriptor is based on edges, which are extracted with a Canny detector. These are stored in a log-polar histogram with location being quantized in nine bins (see Fig. 7). To adapt the feature for the sliding-window approach, it was computed for lattice points with a distance of 16 pixels. Additionally, a principal component analysis was applied to reduce the feature dimensionality.

As a first step, the individual features are evaluated in combination with multiple classifiers on the basis of the “INRIAPerson” dataset from Dalal and Triggs (2005). Initially, 2,416 positive and 12,180 negative training samples were used.

Performance is shown as a precision-recall curve in Fig. 8. Recall is defined as

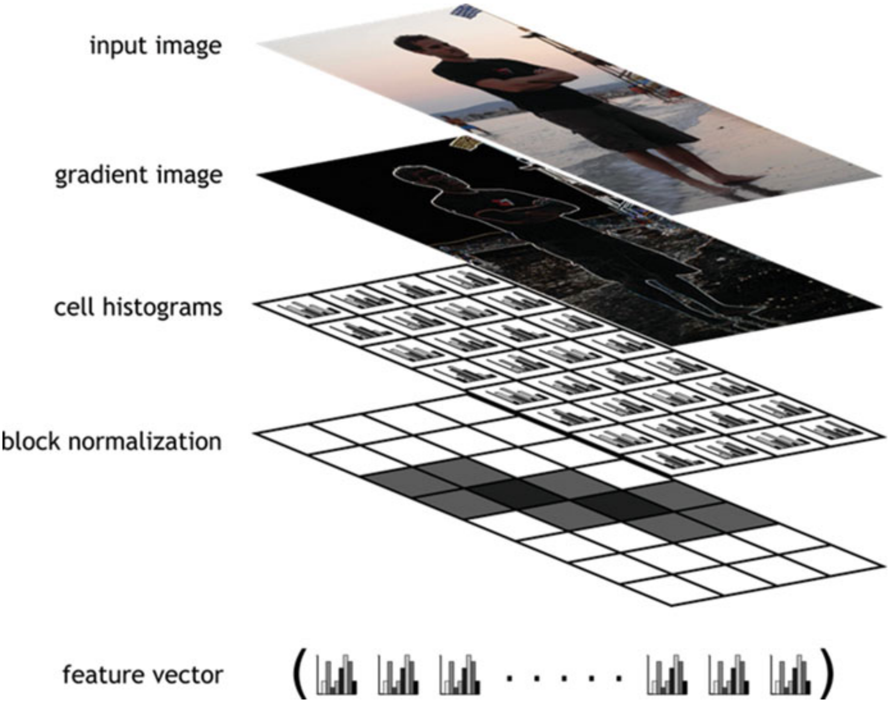


Fig. 5 HOG features

Fig. 6 Shapelet features

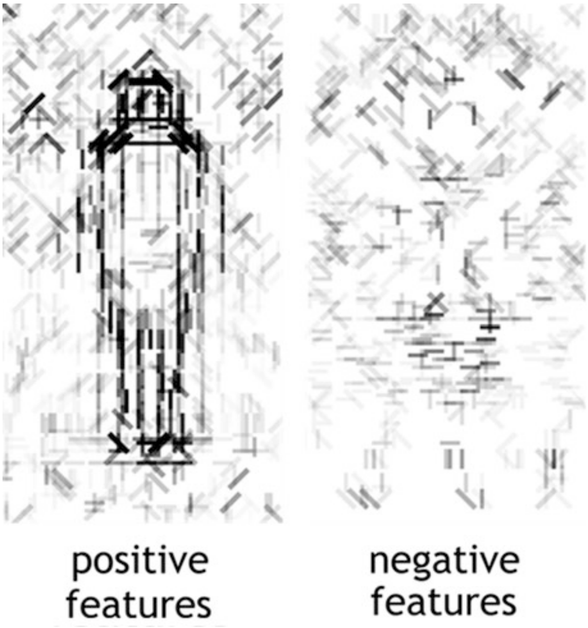
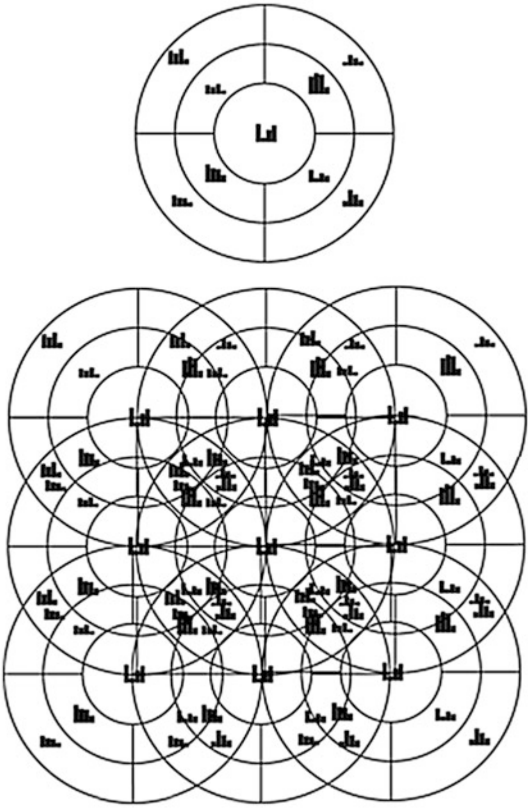


Fig. 7 Shape context features



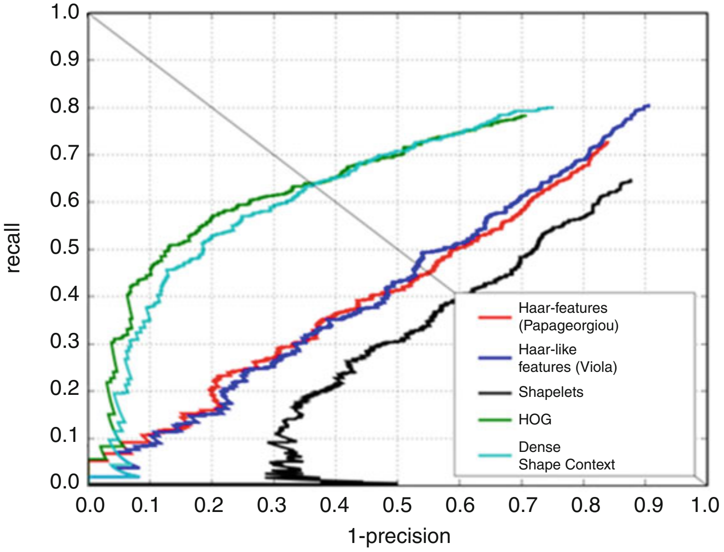
$$\frac{\#\text{correct detections}}{\#\text{correct detections} + \#\text{missing detections}}$$

Precision is defined as

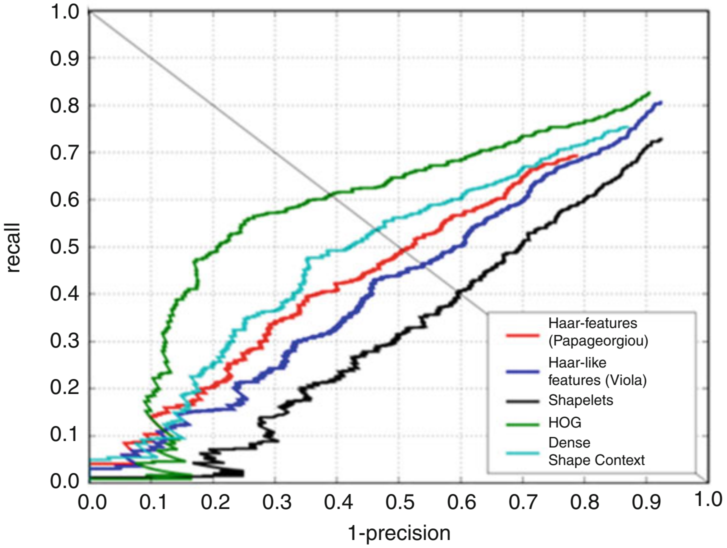
$$\frac{\#\text{correct detections}}{\#\text{correct detections} + \#\text{false detections}}$$

Clearly, the gradient-based features HOG and shape context achieve the best results, regardless of the choice of classifier. Additionally, Haar and Haar-like features perform similarly well, which is not surprising due to the similar design of their features. It should be noted that, in most cases, results can be improved using an RBF kernel.

In a second step, a so-called "bootstrapping" method is employed. For this, an initial model is trained, and all negative images are tested to find additional negative samples that are hard to classify. These are then added to the original training samples, thus multiplying their number. An analysis of individual detectors also



Feature performance with RBF-kernel SVM



Feature performance with AdaBoost

Fig. 8 Evaluation of features with multiple classifiers

shows that these detect different instances and that combining multiple features is therefore most promising. The results can be seen in Fig. 9, with the combination of HOG features and linear SVM (HOG-linSVM) as a baseline.

Evidently, a combination of HOG and Haar features achieves a better overall performance than HOG-linSVM; however, performance depends on the choice of

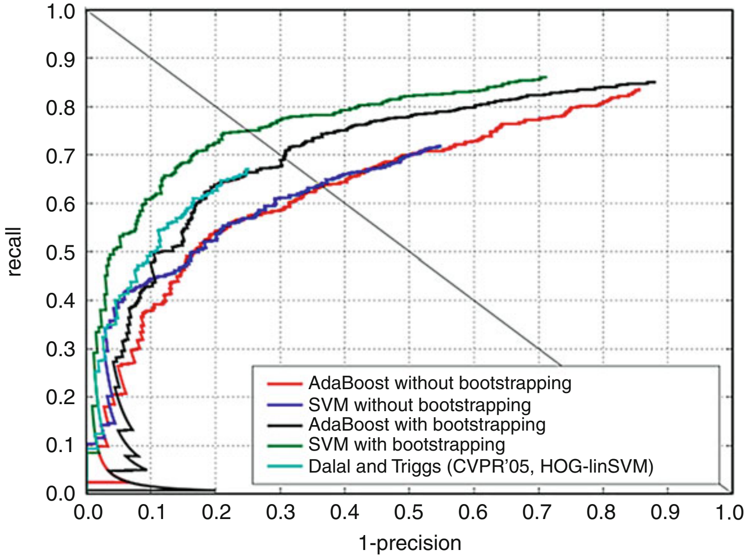


Fig. 9 Performance of feature combinations

the classifier. In combination with AdaBoost, bootstrapping achieves significantly better results, and without it, similar results. The combination with a linear SVM as a classifier also results in a performance comparable to HOG-linSVM.

Performance is similar for the combination of dense shape context features with Haar features without bootstrapping for linear SVMs and AdaBoost. However, in this combination, linear SVM profits more from the bootstrapping step and achieves a significantly better overall performance than HOG-linSVM. In contrast, AdaBoost only achieves an equally good performance.

3.2 Interest Point- and Body Part-Based Approaches

Interest point-based approaches are particularly suitable for large-scale pedestrians. In the approaches discussed here, a typical scale is, for example, 100 pixels and more. In contrast to sliding-window-based detectors, modeling of pedestrians is done locally instead of globally, making this group of approaches significantly more robust against occlusion and articulation. Additionally, some of these approaches allow the pose to be estimated (Seemann et al. 2006; Seemann and Schiele 2006; Andriluka et al. 2008), so that the moving direction of pedestrians can be estimated simultaneously. The following section examines the successful *Implicit Shape Model (ISM)* approach by Leibe and Schiele (Leibe et al. 2008) and further papers on the topic (Leibe et al. 2005; Seemann et al. 2005, 2006; Andriluka et al. 2008).

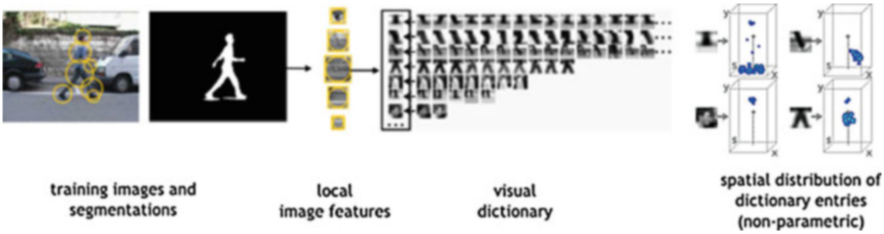


Fig. 10 Learning method for compilation of visual dictionary

The *visual dictionary* is an integral and central part of this group of approaches. It comprises a collection of object parts extracted from a training set of images.

As a first step, an interest point detector is used to determine distinctive pixels. Basically, Harris-Laplace, Hessian-Laplace (Mikolajczyk and Schmid 2004), DOG-detector (Lowe 2004), or a combination of these can be used for this purpose. The detector not only supplies the x-y position in the image, but also a generic scale, i.e., feature size.

In a second step, the feature size is normalized and then described with a feature descriptor. There are several possibilities for choosing a descriptor as well (Mikolajczyk and Schmid 2005; Seemann et al. 2005). SIFT (Lowe 2004), shape context (Belongie et al. 2002), or simply the gray-scale pixel values are most widely used. The descriptors are then clustered to visual dictionary entries. An agglomerative reciprocal nearest neighbor (RNN) pair clustering method is used, which is well adapted to large quantities of data. Next, the visual words are projected back into the images, and their spatial distribution is learned relative to the object center in a nonparametric form. The fact that this is a star-shaped model is particularly important, so that the dependencies for each dictionary entry are learned individually, and there is no modeling of dependencies between dictionary entries. When modeling of interdependencies between dictionary entries is neglected, a small amount of training is sufficient (210 instances). Figure 10 gives an overview of the presented learning method.

The following section describes pedestrian detection by means of the learned visual dictionary. Figure 11 gives an overview of the ISM detection procedure.

As already described for the learning procedure, interest points are again extracted and described by means of descriptors. Subsequently, these are compared to the entries of the visual dictionary and entered into a probabilistic matching space. The local maxima of this space represent detected object positions. A scale-adaptive mean-shift search is performed to determine these efficiently.

Finally, the interest points that support the located maxima can be projected back into the image. This results in a rough pedestrian segmentation in addition to object position and scale. Based on the common foreground/background segmentation of training data, the probability of belonging to the foreground can be determined for each pixel. For this purpose, segmentations are saved, which are then considered according to their contribution to the detection hypothesis, with each entry of the

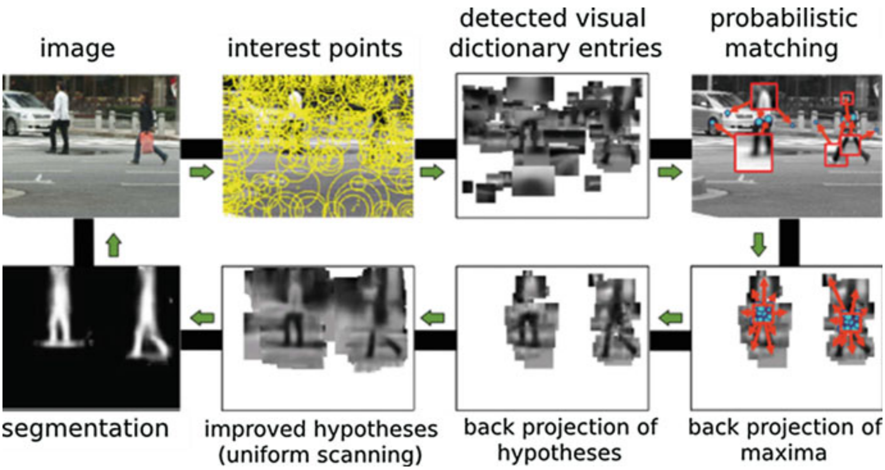


Fig. 11 Overview of the ISM detection procedure

visual dictionary. The probability ratio of foreground versus background is eventually a determining factor for object segmentation.

In particular, in the event of occlusion and multiple closely related hypotheses, inconsistent contributions during the probabilistic matching process occur. This can result in pedestrian hypotheses that are based on parts of another pedestrian, or false hypotheses arising between closely related detections. It has been shown, however, that these ambiguities can be resolved efficiently with the help of the inferred segmentation in a maximum description length (MDL) formulation.

There are several extensions to the general detection method described above, which are specifically tailored to pedestrians. Leibe et al. (2005) combine the local detection method with a global verification step, where the final segmentation of hypotheses is compared to the known silhouettes of training data using chamfer matching. Overall performance can be improved by combining global shape-based silhouette features and ISM detection, supported by local features.

Seemann et al. (2006) extend the matching space by an additional discrete dimension that describes the articulation of the detected pedestrian. For each entry in the visual dictionary, the articulation with which it occurs is saved additionally, thus ensuring articulation consistent occurrence of local features in the detection step. Head features, for example, are consistent with almost all articulations, but foot features only with specific articulations, so that a soft assignment is advantageous. This is confirmed by experimental validation. This approach outperforms the results achieved by the global chamfer verification strategy.

Seemann and Schiele (2006) suggest modeling the contribution of local interest points depending on local context. For each entry in the visual dictionary, the location where it occurs on the silhouette of the training instance is saved. At the time of detection, it is verified whether the context described by other dictionary

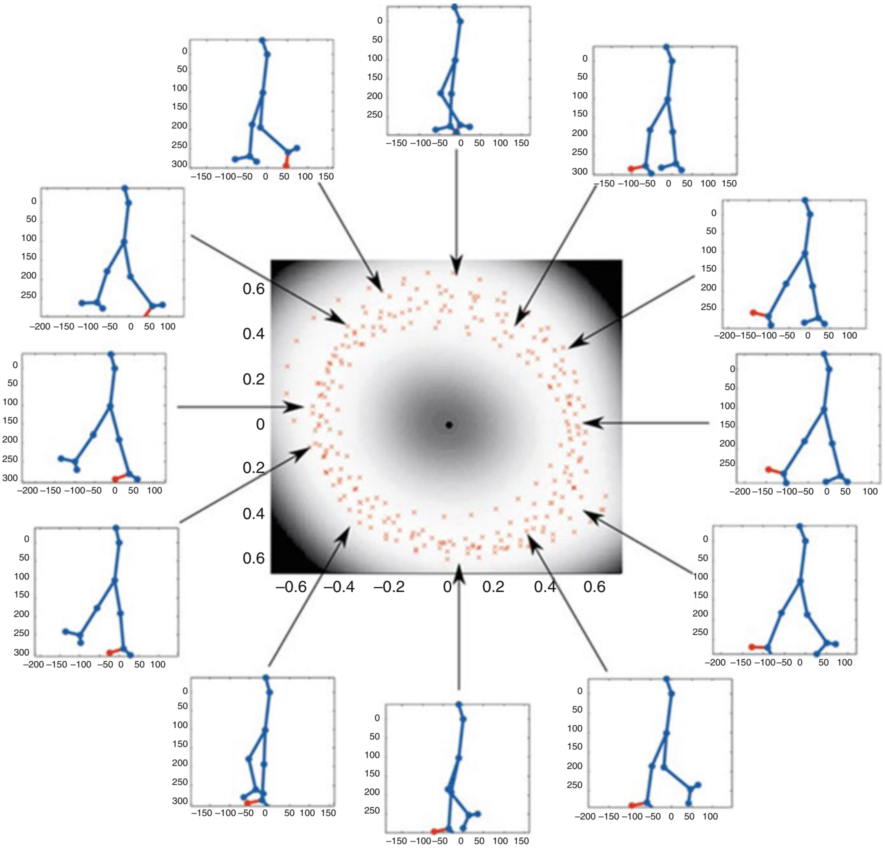


Fig. 12 Human walking sequence in the Gaussian process latent variable model

entries concurs within the user-defined radius. In the specific case that the radius is set to infinity, this approach is identical to the original ISM approach. Compared to the two previous approaches, this approach achieves improved performance.

Andriluka et al. (2008) suggest a further development of ISM for dynamic image sequences. For this approach, global modeling of pedestrians as an object is neglected; instead, individual body parts like feet, arms, torso, and head are separately detected. It is then possible to reconstruct the pose that is embedded in a low-dimensional 2D space (Lawrence 2005) using a nonlinear representation. This embedding is made possible by the interdependencies of different pose parameters, and it is well suited as a dynamic model of a human motion sequence (see Fig. 12).

This dynamic model can be used in a tracking framework that does not make a Markov supposition. Additionally, an instance-specific color model is learned to identify pedestrians in spite of complete long-term occlusion. Even without temporal integration, this body part-based model outperforms ISM; when temporal

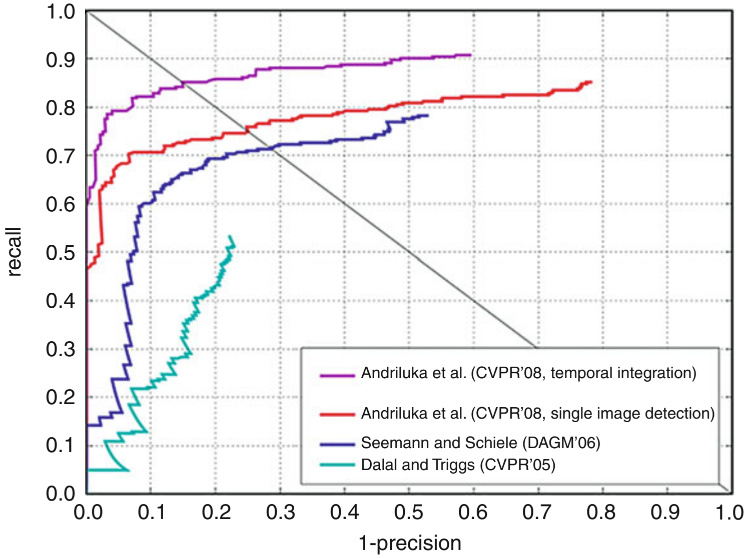


Fig. 13 Performance comparison on video sequence data of ISM variations (Seemann and Schiele 2006; Andriluka et al. 2008) and HOG (Dalal and Triggs 2005)

development is added, a further improvement of detection accuracy is possible (see the comparison in Fig. 13). Figure 14 shows some example detections for this approach.

3.3 System-Oriented Approaches

In contrast to the detectors described above, which work independently from an application, there are approaches that employ application knowledge. An example is the system by Gavrila and Munder, which comprises the following components (cf. Fig. 15):

- Stereo-based region-of-interest selection
- Shape-based detection
- Texture-based pedestrian classification
- Stereo-based pedestrian verification
- Tracking

In order to keep the amount of processing low for further steps, the stereo images are first rectified, and then a sparse disparity image is computed (Franke 2000). In a further step, initial pedestrian hypotheses are generated by means of chamfer matching (Gavrila and Philomin 1999). This is an instance-based matching method. In order to keep computational cost low, the multiple sample instances are clustered

Fig. 14 Example detections for the body part-based ISM approach with temporal integration (Andriluka et al. 2008)

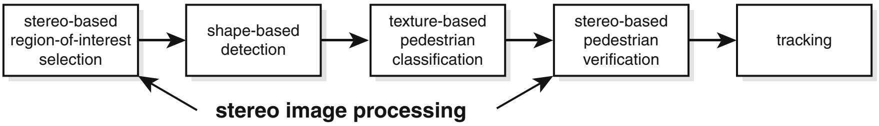


Fig. 15 Computational steps of the PROTECTOR system

hierarchically into three levels. The sliding-window method already discussed is used for this. To further reduce the amount of computation, detection is only performed at image locations where a pedestrian of average size matches the scene geometry. This assumes that camera and pedestrian are located on the same ground plane.

The initial hypotheses are then verified by a texture method. For this purpose, the texture is modeled by means of artificial neural networks (Wöhler and Anlauf 1999), and an SVM is used as a classifier. In contrast to conventional feed-forward networks, weights in local receptive fields are used jointly, so that this method requires less training instances without any trade-off in performance.

Subsequently, a further verification step using stereo image information is performed in order to reduce the number of false detections. In the area of the shape mask that generated the hypotheses, a second-degree polynomial is fitted over the distribution of the cross-correlation values in the range of the estimated depth in the second stereo image. The hypotheses that do not follow the expected distribution are then rejected.

Finally, the hypotheses are temporally smoothed with a Kalman filter. The state vector of the individual tracks consists of the image coordinates, the scale of the corresponding detection, and the estimated depth. In addition, the corresponding first derivative is modeled. Kuhn's classic method (Kuhn 1955) is used to avoid ambiguities when associating measurements with existing tracks. For establishing a cost matrix between tracks and measurements, a weighted linear combination of the chamfer similarity measure and distance of object centers is used.

The experimental validation is also performed for the full system. Contrary to most other studies, performance is validated here regarding the detection of pedestrians in 3D coordinates. Pedestrians were annotated in the 2D input images and then projected back into the 3D space by means of common scene geometry. The system was tested for the detection range of 10–25 m and a camera aperture angle of 30°.

The results show that in particular the stereo-based region-of-interest selection contributes significantly to overall performance. This is confirmed by much higher false detection rates, especially on structured background, if inactivated. When the detection range is limited to +/– 1.5 m lateral distance to the vehicle middle axle, all pedestrians are detected, with approximately five false detections per minute. Gavrila and Munder further demonstrate that computational speed can be increased by approximately 40 % by reparameterizing the region-of-interest selection. However, this affects detection performance, which is reflected in a 6–8 % decrease in system accuracy.

4 Description of Requirements for Hardware and Software

The computational cost of complex detection systems necessary for pedestrian detection is considerable, due to the substantial amount of data that needs to be processed. An essential concept for accelerating existing well-established algorithms is parallelization and using special hardware, for example, *Field-Programmable Gate Arrays (FPGA)* or *Application-Specific Integrated Circuits (ASIC)*. Basically, modern graphic boards are also able to parallelize existing algorithms. Specifically, by providing modern programming paradigms like *Compute Unified Device Architecture (CUDA)* by Nvidia or *Stream SDK* by ATI, relevant concepts can easily be validated at an early stage of development.

Whereas unoptimized implementations of the detection algorithms described above typically have a runtime of several seconds per image on standard hardware, real-time application is made possible by using modern graphics hardware. An empirical study by Wojek et al. (2008) demonstrates that for sliding-window methods, a great number of components can be parallelized and thus performed at greater speeds. For a minimum pedestrian scale of 96 pixels, Wojek et al. achieve an average computation time of 38 ms per image at a resolution of 640 × 480 pixels, which corresponds to 26 Hz. The distance between the processed scales was increased to achieve high computational speed. A respective analysis shows, however, that this is not at the expense of detection performance.

Table 2 compares the computational time of a standard implementation to an accelerated implementation (with small-scale gaps).

This table demonstrates the utmost importance of parallelization, by means of which an acceleration of the factor 109.5 can be achieved. Evidently, the complex computation of image features can benefit in particular, and it experiences the largest acceleration. Feature point-based algorithms, such as ISM, where runtime on standard hardware is also in the range of seconds, can most likely be accelerated by the use of parallel hardware as well. The largest share of runtime is needed for verification of the initial hypotheses in this approach, which requires computation-intensive pixel-wise segmentation of candidate hypotheses. However, this segmentation is well suited for parallelization.

Additionally, note that the quoted runtimes attempt to detect pedestrians at all possible locations and on all possible scales. Techniques like region-of-interest

Table 2 Pure algorithm runtime in milliseconds (without image acquisition)

Processing step/implementation	CPU	GPU	Acceleration
Padding	10.9	1.19	9.15
Gradient computation	3,083.9	20.71	148.9
Histogram computation	4,645.3	24.44	190.1
Normalization	95.8	5.67	16.9
Classification	970.1	27.15	35.7
Image scaling	128.5	2.47	52.0
Sum	**8,934.5**	**81.63**	**109.5**

selection or the assumption of a uniform flat ground plane, as discussed in the section on system-oriented approaches, further significantly reduce the number of operations to be performed and thus accelerate runtime.

Finally, there are further possibilities to accelerate sliding-window detectors, in addition to operation parallelization. However, this might have a detrimental effect on detector performance to some extent.

Coarse-to-fine methods, for example, provide for higher computational speed, but they are limited to a lower minimum detection rate (Zhang et al. 2007). Another possibility is to use detector cascades and to compute only discriminative features over partial areas of the detector window. This, however, requires very efficient feature computation so that the training, which is already complex, can be performed (Viola et al. 2003; Zhu et al. 2006).

5 Outlook

Although, in recent years, there has been considerable success in the area of camera-based pedestrian detection, performance is not yet sufficient for use in automotive applications. Gavrila and Munder, for example, report 0.3–5 false detections per minute for the PROTECTOR system. Moreover, the detection range is limited to a maximum of 25 m, which should be insufficient for most applications.

This is related to impaired pedestrian detection performance for low resolutions, which needs further improvement. This can be achieved by an improved understanding of the overall scenario, for example.

Motion features, such as optical flow, are hardly used in state-of-the-art systems; yet, integrating them could improve detection performance, especially for crossing pedestrians.

References

Andriluka M, Roth S, Schiele B (2008) People-tracking-by-detection and people-detection-by-tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, Anchorage, pp 1–8

Belongie S, Malik J, Puzicha J (2002) Shape matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24:509–522. IEEE Computer Society Press, Washington, DC

Bertozzi M, Broggi A, Caraffi C, Del Rose M, Felisa M, Vezzoni G (2007) Pedestrian detection by means of far-infrared stereo vision. Comput Vis Image Underst 106(2–3):194–204. Elsevier Science Inc., New York

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, San Diego, pp 886–893

Franke U (2000) Real-time stereo vision for urban traffic scene understanding. In: Proceedings of the IEEE international conference on intelligent vehicles. IEEE Computer Society Press, Dearborn, pp 273–278

Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting. Ann Stat 38(2):337–374. Institute of Mathematical Statistics, Beachwood

Gavrila D, Munder S (2007) Multi-cue pedestrian detection and tracking from a moving vehicle. Int J Comput Vis 73(1):41–59. Springer, New York

Gavrila D, Philomin V (1999) Real-time object detection for "smart" vehicles. In: Proceedings of the IEEE international conference on computer vision. IEEE Computer Society Press, Kerkyra, Korfu, pp 87–93

Harris C, Stephens M (1988) A combined corner and edge detector. In: Proceedings of the fourth alvey vision conference. Organising Committee AVC 88, Manchester, pp 147–151

Kuhn HW (1955) The Hungarian method for the assignment problem. Naval Research Logistic Quarterly 2:83–97. Institute for Operations Research and the Management Sciences, Hanover

Lawrence ND (2005) Probabilistic non-linear principal component analysis with Gaussian process latent variable models. J Mach Learn Res 6:1783–1816. MIT Press, Cambridge, MA

Leibe B, Seemann E, Schiele B (2005) Pedestrian detection in crowded scenes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, San Diego, pp 878–885

Leibe B, Leonardis A, Schiele B (2008) Robust object detection with interleaved categorization and segmentation. Int J Comput Vis 77:259–289. Springer, New York

Lindeberg T (1998) Feature detection with automatic scale selection. Int J Comput Vis 30(2):77–116. Kluwer, Hingham

Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60(2):91–110. Kluwer, Hingham

Mählisch M, Oberlander M, Lohlein O, Gavrila DM, Ritter W (2005) A multiple detector approach to low-resolution FIR pedestrian recognition. In: Proceedings of the IEEE international conference on intelligent vehicles. IEEE Computer Society Press, Las Vegas, pp 325–330

Maji S, Berg AC, Malik J (2008) Classification using intersection kernel support vector machines is efficient. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, Anchorage, pp 1–8

Mikolajczyk K, Schmid C (2004) Scale and affine invariant interest point detectors. Int J Comput Vis 60(1):63–86. Kluwer, Hingham

Mikolajczyk K, Schmid C (2005) A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 27(10):1615–1630. IEEE Computer Society Press, Washington, DC

Papageorgiou C, Poggio T (2000) A trainable system for object detection. Int J Comput Vis 38(1):15–33. Kluwer, Hingham

Sabzmeydani P, Mori G (2007) Detecting pedestrians by learning shapelet features. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, Minneapolis, pp 1–8

Schoelkopf B, Smola AJ (2001) Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge, MA

Seemann E, Schiele B (2006) Cross-articulation learning for robust detection of pedestrians. In: Pattern recognition: proceedings of DAGM symposium. Springer, Berlin/New York, pp 242–252

Seemann E, Leibe B, Mikolajczyk K, Schiele B (2005) An evaluation of local shape-based features for pedestrian detection. In: Proceedings of the British machine vision conference. British Machine Vision Association, Oxford, UK, pp 11–20

Seemann E, Leibe B, Schiele B (2006) Multi-aspect detection of articulated objects. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, New York, pp 1582–1588

Shashua A, Gdalyahu Y, Hayun G (2004) Pedestrian detection for driving assistance systems: single-frame classification and system level performance. In: Proceedings of the IEEE international conference on intelligent vehicles. IEEE Computer Society Press, Tokyo, pp 1–6

Suard F, Rakotomamonjy A, Bensrhair A, Broggi A (2006) Pedestrian detection using infrared images and histograms of oriented gradients. In: Proceedings of the IEEE international conference on intelligent vehicles. IEEE Computer Society Press, Tokyo, pp 206–212

Viola PA, Jones MJ, Snow D (2003) Detecting pedestrians using patterns of motion and appearance. In: Proceedings of the IEEE international conference on computer vision. IEEE Computer Society Press, Washington, DC, pp 734–741

Wöhler C, Anlauf JK (1999) An adaptable time-delay neural-network algorithm for image sequence analysis. IEEE Trans Neural Netw 10(6):1531–1536. IEEE Computer Society Press, Washington, DC

Wojek C, Schiele B (2008) A performance evaluation of single and multi-cue people detection. In: Pattern recognition: proceedings of DAGM symposium. Springer, New York/Munchen, pp 82–91

Wojek C, Dorko G, Schulz A, Schiele B (2008) Sliding-windows for rapid object-class localization: a parallel technique. In: Pattern recognition: proceedings of DAGM symposium. Springer, New York/Munchen, pp 71–81

Zhang W, Zelinsky G, Samaras D (2007) Real-time accurate object detection using multiple resolutions. In: Proceedings of the IEEE international conference on computer vision. IEEE Computer Society Press, Rio de Janeiro, pp 1–8

Zhu Q, Avidan S, Yeh MC, Cheng KT (2006) Fast human detection using a cascade of histograms of oriented gradients. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE Computer Society Press, New York, pp 1491–1498

Part V

Data Fusion and Environment Representation

Data Fusion of Environment-Perception Sensors for ADAS

23

Michael Darms

Contents

Abstract

More and more driver assistance systems are based on a fusion of multiple environment perception sensors. This chapter gives an overview about the objectives of sensor data fusion approaches, explains the main components involved in the perception process, and explains the special topics that need to

---

M. Darms (✉)
Continental AG, Lindau, Bavaria, Germany
e-mail: mdarms@gmx.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_24

be taken into consideration in developing a multi-sensor fusion system for driver assistance systems. Focus is put on the topics of data association, tracking, classification, and the underlying architecture. The architecture strongly influences the costs, performance, and the development process of a multi-sensor fusion system. As there are no deterministic methods that guarantee an optimal solution for developing an architecture, the chapter gives an overview of established, general architecture patterns in the field of sensor data fusion and discusses their benefits and drawbacks.

1 Introduction

Driver assistance systems exclusively based on single-sensor solutions are known from prior art. Examples include applications such as adaptive cruise control, which relies on a single RADAR or laser sensor, for example, or lane departure warning, which typically relies on a video sensor system.

As described in the previous chapters, the various sensor technologies all have specific advantages and disadvantages. For example, a RADAR sensor can be used to determine the longitudinal distance and velocity of a vehicle driving ahead with a degree of accuracy sufficient for the adaptive cruise control application (see ▶ Chaps. 45, “Adaptive Cruise Control,” and ▶ 17, “Automotive RADAR”). However, the relevant object, to which a certain distance must be kept, can only be selected with a certain degree of precision due to the lateral resolution, the ambiguities in signal evaluation, and the lack of lane marker detection; interference from vehicles on adjacent lanes must be thus accepted in system operation. On top of this, there are limits to the ability to classify the detected object, so the control algorithm typically only uses objects for which motion has been detected.

The missing information can be provided, for example, by data from a video sensor (see ▶ Chaps. 19, “Automotive Camera (Hardware),” ▶ 20, “Fundamentals of Machine Vision,” and ▶ 48, “Lateral Guidance Assistance”). Lane marker detection provides information which can be used for lane assignment. Classification algorithms allow vehicles in the video image to be distinguished from other objects, while image processing technologies enable to determine the position of vehicles in the video image. In contrast to RADAR sensor systems, the distance and speed cannot be measured and must therefore be estimated. The achievable precision is significantly lower with current sensor systems, especially in the long-distance range. The functionality of an adaptive cruise control system purely based on video sensors is thus restricted to a smaller speed range.

Combining the information from both sensors helps to leverage the benefits of both technologies. For example, the RADAR sensor’s distance measurements can be combined with the classification information and vehicle position measurements in the video image. This makes it possible to reduce false interpretations and improve accuracy in terms of the lateral position and distance. At the same time, lane assignments, and thus the ability to detect the reference object with the help of the video sensor data, become more robust.

Various research works (see, e.g., Darms 2007; Holt 2004; Stüker 2004; Becker 2002; Bender et al. 2007), confirm the capability of data fusion approaches of this type; environmental sensor data fusion is used in production vehicles (see, e.g., Schopper et al. 2013). This means both the fusion of RADAR and camera sensors, as in the example given here, and other combinations, for example, short-range and far-range RADAR. The principle of fusion can be extended to other sensor technologies. The current research and development focuses include the fusion of various imaging sensors and the fusion of data from environmental sensors with stored map data.

The following sections provide an introduction to the basic principles of sensor data fusion in driver assistance systems. Firstly, the term sensor data fusion is defined, and the objectives of fusion are stated. The main components of environmental data processing are then explained with a view to fusion of data from multiple sensors. Finally, established architectural patterns for sensor data fusion are presented. Part of the text in this chapter is orientated on the text provided in Darms (2007).

2 Definition and Objectives of Sensor Data Fusion

2.1 Definition of Sensor Data Fusion

According to Steinberg et al., the process of data fusion is defined as follows:

> Data fusion is the process of combining data or information to estimate or predict entity states. (Steinberg et al. 1998)

The generic term "entity" is used to describe an abstract object to which information can be assigned. In the world of driver assistance systems, this can mean a physical object in the vehicle's environment, such as an observed vehicle, but also an individual state variable, such as the pitch angle.

The following text mainly refers to the former case and thus directly uses the term "object." The focus is on track estimation, which is also referred to as tracking, and on object discrimination (see also Klein 1999). Tracking means estimating the states of an object in terms of control theory (e.g., position and speed). Object discrimination is further broken down into detection and classification (Klein 1999). In the course of detection, a decision is made as to whether an object exists, while classification assigns the object to a predefined class (e.g., vehicle, pedestrian). However, the considerations presented here can also be generalized to apply to abstract objects (see also the discussion in Dietmayer et al. (2005)).

2.2 Objectives of Data Fusion

The primary objective of data fusion is to merge the data from individual sensors so as to combine their strengths in a beneficial way and reduce individual weaknesses.

The following aspects can be distinguished (see also Lou and Kay 1991; Joerg 1994).

Redundancy Redundant sensors provide information relating to the same object. This helps to improve the quality of the estimation. An estimation algorithm must take the measuring error dependencies into consideration (see, e.g., Bar-Shalom and Li 1995). One risk is the multiple introductions of artifacts and misinterpretations into the fusion process (see below).

Redundancy can also help to improve the error tolerance and availability of the system in cases of individual sensor failure on the one hand – assuming that the system can still provide data of sufficient quality without the information from the failed sensor – and for artifacts or misinterpretations by individual sensors on the other. Redundancy can reduce the influence of an individual single error on the system as a whole.

Complementarity Complementary sensors deliver different, supplementary information into the fusion process. This can happen from a spatial point of view, where the same sensors deliver information with different field of views. Particular attention should be focused on data processing of the peripheral zones of the detection area in this case (see, e.g., Stüker 2004). This can also mean data that relate to the same object. The information content can be enhanced by detecting different properties. It is possible that a combination of the individual items of data is required to provide the information required by the application.

The use of different sensor technologies can also improve the robustness of the overall system in terms of detecting individual objects that may not be reliably detected by single-sensor technology. For example, the beam from a laser sensor penetrates glass, or the beam from a RADAR sensor penetrates various plastic materials without detecting the object in question. Combining the sensors reduces the probability of not detecting the object at all.

Temporal Aspects The overall system's speed of acquisition can be improved by a fusion approach. This can be achieved firstly by parallel processing of information from the individual sensors and secondly by appropriate timing of the acquisition process (e.g., by sensors measuring alternately).

Improved precision, or the introduction of complementary information, also influences the dynamic of the estimation. It must also be noted that different applications can pose different requirements in terms of estimation dynamic and accuracy and that it can still make sense, even in a sensor fusion system, to use different estimation algorithms for different applications (see Sect. 3.2).

Costs When designing any sensor system, the costs are a decisive factor in deciding its practical feasibility. The use of a fusion system can help to reduce the costs, compared with an individual sensor. However, this is not true in all cases because, for example, improvements can also be achieved by developing new algorithms for evaluating the data from a single sensor or by hardware advances.

The decision to develop a single- or multi-sensor system will thus always be multidimensional and must be based on the aspects stated above.

The costs of a sensor fusion system are substantively influenced by the architectural structure of the system (see, e.g., Hall 2001; Klaus 2004). Thus far, a uniform architecture has not been specified in the automotive industry in the form of a mandatory or de facto standard. This makes cross-enterprise cooperation between suppliers and vehicle manufacturers, strategic development of sensors and algorithms adapted for a common architecture, and the migration to new assistance functions and sensor generations more difficult (see also Hall 2001).

Modularity and the ability to economically extend the system are critical to its practical feasibility. The aim is to realize migration to new assistance functions economically and make it possible to source sensors and modules from various suppliers, an aspect which is especially important to vehicle manufacturers.

3 Main Components in Sensor Data Processing

3.1 Overview

The following section summarizes the main components in environment sensor data processing. The structure is generic and applies also to single-sensor systems. The special features that need to be taken into consideration in developing a multi-sensor system are pointed out at the appropriate places.

3.2 Signal Processing and Feature Extraction

In the scope of signal processing and feature extraction (see also Hall and McMullen 2004), information from the vehicle's environment is acquired by sensors. Figure 1 shows the process. In the first step, which is referred to as measure, the receiver element of the sensor (signal reception) receives payload signals (energy) overlaid with interference signals (noise) and converts them into raw signals (e.g., voltages, currents). The raw signals are interpreted as physical measurements (e.g., intensities, frequencies, etc.), which finally form the sensor's

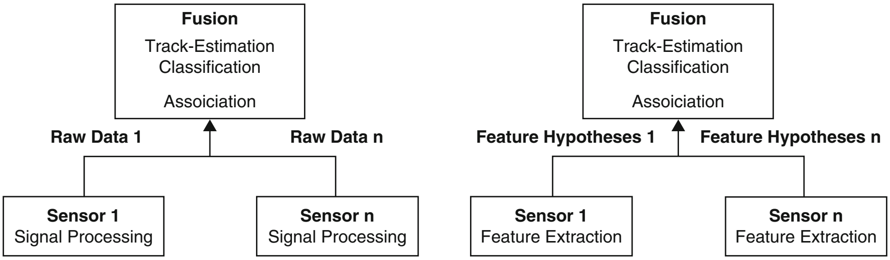


Fig. 1 Perception process: measure and perceive (see Darms 2007, p. 9 and Darms et al. 2009)

raw data. During signal processing, (physical) assumptions for interpretation are made (e.g., maximum reception level, impulse forms, etc.). Where these assumptions are breached, artifacts (system-specific weaknesses) occur.

In the second step, termed perceive, features (e.g., edges, extreme values) are extracted from the raw data on the basis of assumptions and models/heuristics. An object hypothesis, an assumed object, is derived from these feature hypotheses. Misinterpretations can occur due to the use of heuristics.

Where the information from multiple sensors is used in the estimation process, it is necessary to find a common reference for the information. This task is in particular made more difficult if the information is not orthogonal, that is, statistically independent.

One fundamental problem here is that of transferring the data to a coordinate system with a common reference point. In the case of a single sensor, the effect of adjustment errors can only be a negligible offset. However, maladjustment of a multi-sensor system can make it impossible to align the data from various sensors or cause systematic errors and deviations. This can impair the quality of the evaluation (see below). Suitable adjustment processes and (online) adaption algorithms are thus a central development focus for a multi-sensor system.

On top of this, various sensors can measure different attributes, even if this is not desired. This occurs in particular with non-orthogonal sensors. For example, if the distance to a vehicle is measured by a laser sensor and a RADAR sensor, it is possible that the sensors detect different parts: the laser sensor might detect the rear reflectors on a truck, while the RADAR sensor detects the rear axle. This effect can also be observed for identical sensors. One reason for this is that an object is detected from different angles of view. It is aggravated by sensor-specific artifacts during measurement and feature extraction, which can also have an effect despite the use of identical sensors.

Special care needs to be taken for the perceive part in multi-sensor systems. For example, the extracted feature hypotheses from various sensors will ideally relate to the same physical object. Due to different sensor resolutions, and misinterpretations, for example, in data segmentation (see, e.g., Holt 2004; Streller 2006), the object hypothesis can differ between sensors. For a system with unsynchronized sensors, the extracted features can also originate at different points in time. To be able to combine the data from the various sensors, one thus at least needs a mutual time base and sufficiently accurate time stamping (see also Kampchen and Dietmayer 2003).

The topic of temporal and spatial association of data from various sensors is also summarized in the referenced literature under “sensor registration” (see, e.g., Hall 2001).

3.3 Data Association

The feature hypotheses gained from signal processing and feature extraction are associated with object hypotheses already known to the system in the data

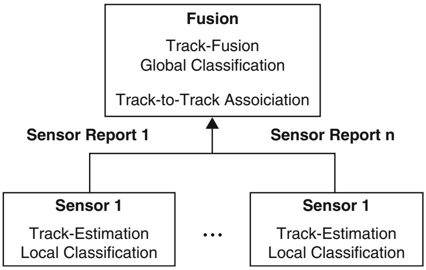


Fig. 2 Breakdown of the data association process (see Darms 2007, p. 45)

association step (see, e.g., Bar-Shalom and Li 1995). The quality of the estimation is significantly influenced by the data association process (see Holt 2004; Stüker 2004; Bar-Shalom and Li 1995). If an incorrect association is made, information loss occurs or false information is introduced in the estimation process (see, e.g., Stüker 2004).

Hall and Llinas break the data association process down into the following three steps (see Hall and Llinas 1997 and Fig. 2); special algorithms used in automotive applications can be found in Holt (2004), Becker (2002), and Streller (2006), for example.

1. **Generating association hypotheses**. Theoretically possible associations of feature hypotheses to object hypotheses are found. The results are one or multiple matrices with theoretically possible associations (association matrices).
2. **Evaluating the association hypotheses**. The association hypotheses found are evaluated with the aim of quantitative evaluation or ranking. The results are quantitative values (e.g., costs) in the association matrix or matrices.
3. **Selection of association hypotheses**. A selection is made from the evaluated association options; downstream data processing and thus, in particular, data filtering are based on this.

The three processing steps do not need to be implemented separately; on the contrary, they can depend on one another. However, it is advisable to decouple the steps in the development process (see Hall and McMullen 2004). The quality and performance of the available resources (e.g., computing capacity, resolution and usable raw data of a specific sensor, artifacts and potentially false interpretations) play a role in designing the algorithms. Depending on these boundary conditions, various solutions are possible (see Hall 2001).

Hypothesis generation itself can be broken down into two sub-steps: postulating the association hypotheses and selecting the theoretically possible hypotheses. Various methods can be used for postulating the association hypotheses. They include (see Hall and McMullen 2004):

Physical models. Fields of view and occlusions of the sensors used can be calculated. Object hypotheses that lie significantly outside the field of view are not considered in generating hypotheses.

Scenario knowledge. The behavior and potential location of objects on the basis of the observed scenario can be leveraged, for example, areas for finding road markers or traffic signs.

Probabilistic models. The expected number of false detections can be factored into the process.

Ad hoc methods. One example of this is postulating all possible association options. No prior knowledge needs to exist for this. However, it does make the process of selecting the correct associations more difficult.

The following methods are possible for selecting possible hypotheses (see, e.g., Hall and McMullen 2004):

Pattern detection algorithms. Associations can be ruled out using the raw signals and raw data (e.g., via correlation techniques).

Gating techniques. Physical models, for example, can be used to compute an area in which object hypotheses, or the feature hypotheses derived from them, can exist with a specific probability at the current time of measuring (prediction). Feature hypotheses originating from the current measurement cycle that lie outside of such an area are not associated with the corresponding object hypothesis.

Hypothesis evaluation can be based on probabilistic models based on Bayes' theorem, possibilistic models based on the Dempster–Shafer theory, neuronal networks, or even ad hoc techniques, such as unweighted distance computation between a prediction of the features and the features themselves (see, e.g., Hall and McMullen 2004).

Finally, a variety of mathematical algorithms exists for *hypothesis selection* (see Hall and McMullen 2004). This solution requires a large amount of computing time with increasing dimensions and in particular if data from multiple cycles are considered in the selection algorithm.

Simple approaches, where hypothesis selection only considers the data from the current cycle, are manageable in terms of complexity. Stüker provides an overview of various association methods (see Stüker 2004). A problem that is frequently found here is that of associating n object hypotheses with m feature hypotheses where $m \geq n$, and where one object hypothesis is associated with precisely one feature hypothesis.

Precise methods exist for this that minimize the aggregated costs in the association matrix. One example is the Munkres algorithm which has a complexity of $O(n^2m)$) (see Becker 2002). Less complex algorithms also exist, but they only provide approximated solutions. One example is the iterative nearest neighbor method, which successively selects associations with the lowest cost, or the highest probability, at a complexity of $O(m^2 \log 2m)$ (see Becker 2002). Depending on the sensor technology, various algorithms are used (see Darms 2007).

As the discussion shows, data association can also be optimized by means of sensor-specific algorithms. Without access to the raw data, and if the sensor-specific conditions are not taken into account, the quality of the data association can degrade (see also Darms 2007).

Data association is also related to feature extraction and object hypothesis generation. Again, a variety of sensor-specific options for optimizing or reconciling individual processes exist with a view to achieving the best possible association of feature hypotheses to object hypotheses given the existing resources. This approach makes it possible to identify artifacts, e.g., duplicate measurements in the scope of data association, and to exclude them from the fusion process (see, e.g., Darms et al. 2008).

Knowledge of the way the data are generated, such as potential artifacts and typical misinterpretations, can thus be used for optimizing the algorithms. In addition, special properties of a sensor technology, such as the resolution capability, can be taken into account when designing the algorithms. The data association algorithm design is thus related to the knowledge of how the data are generated and thus of the hardware of the sensor being used. In a modular setup, it can thus be useful to encapsulate the association algorithms in sensor-specific modules (see Darms 2007).

3.4 Data Filtering

The feature hypotheses that have been extracted and associated with an object hypothesis are processed downstream by a filter or estimation algorithm. This algorithm is used to improve the information, but also to gain new information (see Bar-Shalom et al. 2001; Hänsler 1997). Examples include:

- Signal and noise separation
- Reconstructing state variables that cannot be measured directly

For an overview of filter algorithms for sensor data fusion, see Holt (2004), Klein (1999), and Bar-Shalom and Li (1995). The filter parameters are designed and configured to suit the optimization criteria that need to be defined for the individual application (see Hänsler 1997). If the filter is part of the control loop, it influences the dynamic behavior of the entire system (see, e.g., Lunze 2006; Föllinger 1990). In this case, the filter parameters must be adapted to suit the control loop's requirements (e.g., ACC). It is important to find a compromise between the filter dynamic and the achievable estimation accuracy (see Lunze 2006). If a state controller is used, the separation theorem (Lunze 2006; Föllinger 1990) at least ensures the stability of the overall system, assuming that the estimator is stable. The control and estimation parameters can be designed separately (see Lunze 2006; Föllinger 1990); this offers benefits in terms of architecture.

To save costs, the data from a multi-sensor system can be provided to various applications (see, e.g., Darms 2007; Dietmayer et al. 2005). It is important to consider the fact that, depending on the sensor accuracy, areas can exist in which various applications cannot be operated with a common filter algorithm or in which shared operation of the application with one filter necessitates finding a compromise that is not optimal for individual applications in terms of dynamics (see Darms 2007).

Development of data filtering algorithms cannot be completely abstracted from the data association design. This is true of the design process, in which mutually compatible algorithms must be found (see Bar-Shalom and Li 1995), and also of the runtime behavior, given that the data filtering dynamic influences the quality of the association process. Again, depending on the sensor accuracy, it is possible that different filter algorithms for applications and data association make sense (see Darms 2007).

3.5 Classification

During classification, object hypotheses are assigned to a predefined class on the basis of associated properties (see, e.g., Klein 1999). The properties can come from the sensor's raw data, but also from the estimated state variables of the object hypothesis.

In a multi-sensor system, the input data from various sensors are available. In terms of the architectural design, it is beneficial for the data included in the fusion process to be mutually orthogonal. A multiple implementation of a classification on the basis of state variables can be avoided, given an appropriate architecture design (see Sect. 4).

3.6 Situation Analysis

Situation analysis determines the overall behavior of the driver assistance system. For example, adaptive cruise control (ACC) is backed up by a state machine that defines the application's behavior in various scenarios (see, e.g., Mayr 2001).

Situation analysis is thus the link between environment sensor data processing and the assistance function. Algorithms for situation analysis need to consider both the capability of the environmental data acquisition system and the application's boundary conditions. In the case of automatic emergency braking, for example, a decision to intervene is taken as part of the situation analysis; this decision is driven both by the accuracy with which the potential collision object is estimated and by the potential, vehicle-specific evasion trajectories.

4 Architecture Patterns for Sensor Data Fusion

4.1 General Overview

The architecture documents the structure and the interactions between the individual components for the persons involved in developing the system (see Starke 2005). The architecture of the system also contributes toward structuring the development process (see Starke 2005). This is also true beyond corporate boundaries, as the architecture and the degree of coupling (see Vogel 2005) within the system influence the extent to which components can be manufactured by various suppliers.

There is no deterministic method that guarantees an optimal solution for developing an architecture (Starke 2005). The following section lists established, general architecture patterns in the field of sensor data fusion and discusses the benefits and drawbacks.

4.2 Decentralized–Centralized–Hybrid

The distinction into decentralized, centralized, and hybrid fusion relates to the module view of the system (Vogel 2005). It is based on the degree of data processing in the sensors, the results of data processing in the sensors, and the point at which the data are merged in the fusion process (Klein 1999). It is typically used in conjunction with tracking (Hall and Llinas 1997).

Figure 3 shows a **decentralized architecture**. This approach is referred to in the referenced literature as *sensor-level fusion, autonomous fusion, distributed fusion, or post-individual sensor processing fusion* (Klein 1999). The individual sensor modules handle object discrimination and tracking. The results are merged in a central module, possibly involving feedback of results from the centralized fusion to the sensors (Bar-Shalom and Li 1995). In this case, each decentralized module can additionally handle the central module functions, thus achieving redundancy (Bar-Shalom and Li 1995).

In terms of object discrimination, this type of architecture is optimal, given that the sensors are mutually orthogonal for this operation. This is the case, for example,

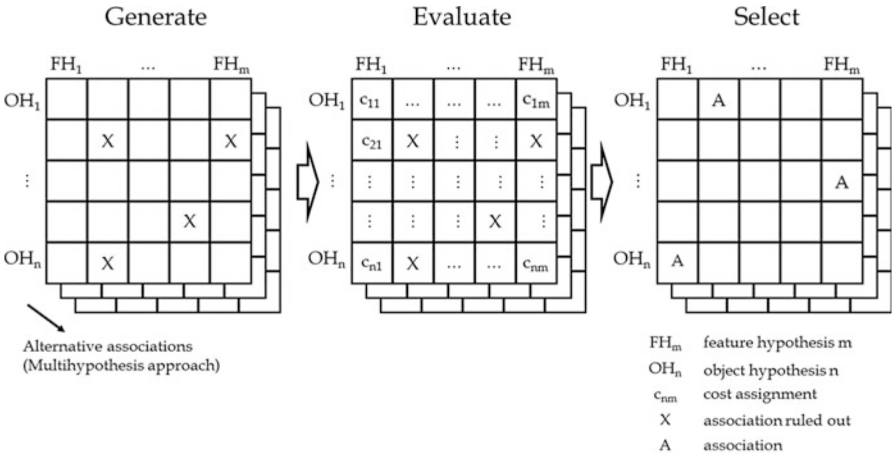


Fig. 3 Decentralized architecture (see Darms 2007, p. 16)

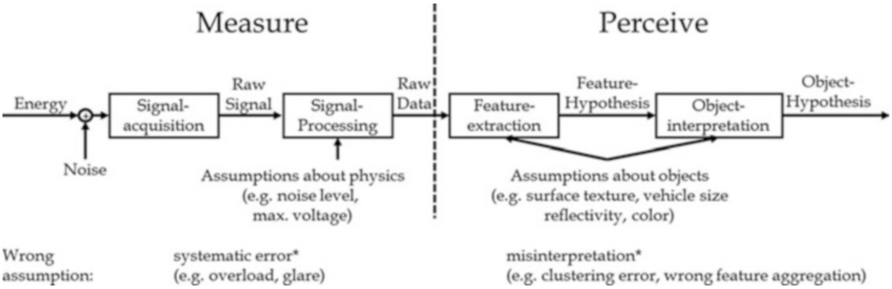


Fig. 4 Centralized architecture. (**a**) Fusion at raw data level. (**b**) Fusion at feature level (see Darms 2007, p. 17)

if sensor principles based on different physical effects are used that do not cause artifacts due to identical phenomena (Robinson and Aboutalib 1990). Two pieces of information are required for fusion: firstly the discrimination decision and secondly a metric for the decision quality (Klaus 2004).

The architecture can also be optimal for tracking, in the sense of minimizing the estimation error (Bar-Shalom and Li 1995). However, this is only true given relatively restrictive preconditions, which rarely exist in practical applications. If the sensors' measuring times also differ, again, the solutions are only approximately ideal in terms of the achievable accuracy (Bar-Shalom and Li 1995).

Figure 4 shows a **centralized architecture**. This is referred to in the referenced literature as *central-level fusion*, *centralized fusion*, or *pre-individual sensor processing fusion* (Klein 1999). The data only go through minimal preprocessing in the sensor modules (feature or raw data level) and are then merged in a centralized module, possibly involving feedback to the sensor modules (Klein 1999).

In terms of object discrimination, this type of architecture is superior to a decentralized architecture if the sensors are not mutually orthogonal. If the sensors are orthogonal, the results do not differ (Klein 1999).

A centralized architecture is optimal for tracking, without the restricting prerequisites that apply for a decentralized architecture. Additionally, measurements not taken at the same time can be optimally merged (Bar-Shalom and Li 1995).

The main drawbacks of a centralized architecture are firstly restrictions in terms of flexibility, as the internal algorithms of the central module may need to be modified to accommodate extensions, and secondly a higher data volume that occurs at the interfaces between the sensor modules and the fusion module (Klein 1999).

A **hybrid architecture** combines the centralized and decentralized approaches. In addition to minimally preprocessed data (raw data), data preprocessed by the sensors (tracks) can be fed to the central fusion module. Tracks can in turn provide input for a decentralized fusion module in the same system. The results from this decentralized module can flow into the central fusion module's fusion algorithm (Klein 1999).

As an example of the use of hybrid architecture, Bar-Shalom and Li describe a scenario that is broken down into various acquisition areas, each of which is covered by a multi-sensor platform. A centralized architecture is used within the platform, while the overall estimation is determined by a decentralized architecture across the areas (Bar-Shalom and Li 1995).

4.3 Raw Data Level–Feature Level–Decision-Making Level

The distinction into fusion at raw data level, feature level, and decision-making level relates to the resolution of the data fed to the fusion algorithm and the degree of sensor data preprocessing (Klein 1999). It thus relates to the runtime view (Starke 2005) and is typically used in the context of object discrimination algorithms (Hall and Llinas 1997).

In the case of **fusion at raw data level**, minimally preprocessed data that exist at the resolution of the sensors involved (e.g., pixels in image processing) are fused in a centralized architecture. This means that, for example, information from various spectra (infrared, visible light) can be fused prior to image processing (Klein 1999). The advantage this approach offers is the availability of complete sensor information to which the fusion algorithm can be adapted. The main disadvantages are the large data volume between the sensors and the centralized module, as well as the difficulty of changing and extending the optimized algorithms in the centralized module.

In the case of **fusion at feature level**, the features are first extracted before fusion is performed. In a centralized architecture, this reduces the communication bandwidth between the sensor modules and the central module at a price of losing information.

Fusion at decision-making level is equivalent to a decentralized architecture. In contrast to fusion at feature level, object discrimination is already performed in the sensor modules. The results are then merged together with the tracking information in a centralized module (Klein 1999). Tracking in this case does not need to follow the decentralized architecture principle.

Table 1 summarizes the architectural principles decentralized–centralized–hybrid and raw data level–feature level–decision-making level, as well as their dependencies.

4.4 Synchronized–Unsynchronized

In terms of the system's dynamic interaction, a distinction can be made between synchronized and unsynchronized sensors. The distinction relates to the temporal sequence in which the data are acquired by the sensors (see, e.g., Bar-Shalom and Li 1995; Narbe et al. 2003a; Narbe et al. 2003b; Mauthener et al. 2006).

In **synchronized sensors** data acquisition is temporally aligned. Synchronous sensors are a special case of synchronized sensors in which data acquisition occurs

Table 1 Fusion architectures (see Darms 2007, p. 19, following Hall and McMullen 2004, pp. 360–361; see also Klein 1999, p. 73)

Type	Description	Fusion level	Comment
Centralized	Raw data fusion	Raw data	Minimal information loss In comparison, needs the greatest communication bandwidth between the sensor modules and the centralized module Optimal for orthogonal and non-orthogonal sensors
	Feature fusion	Feature	Requires lower communication bandwidth than fusion at raw data level Information loss due to feature extraction The benefits of fusion at raw data level cannot be leveraged for non-orthogonal sensors
Decentralized	Fusion of state variables and discrimination decisions	Decision-making level	Information loss due to feature extraction Optimal object discrimination for orthogonal sensors Optimal tracking only under restrictive conditions Dependency on the results determined in the sensor modules must be taken into consideration on fusion Redundancy can be achieved by allowing multiple decentralized modules to compute the fusion
Hybrid	Combination of centralized and decentralized	Combination possible on all levels	Combines the properties of centralized and decentralized architecture High complexity of the architecture in comparison

simultaneously. With **unsynchronized sensors**, data acquisition occurs in an individual sensor cycle that is not aligned with the other sensors and does not need to be constant.

The drawback of synchronization is the additional overhead in terms of hardware and possibly software; the advantage is that the system's timing behavior is already known at the design stage (see also Kampchen and Dietmayer 2003).

4.5 New Data–Data Constellation–External Event

Events which cause data fusion to be performed can be grouped into three classes: the occurrence of new data, the occurrence of a specific data constellation, and the occurrence of an external event.

If fusion occurs whenever **new data** occur, no information is lost. Depending on whether synchronized or unsynchronized sensors are used, the fusion process needs to find solutions for processing data that do not arrive at the fusion modules in the temporal sequence of data acquisition (Stüker 2004; Bar-Shalom 2002). In a decentralized structure, the latest fusion data can be fed back to the sensors so that the sensors always have the latest prediction, for example, for preconditioning algorithms.

If the fusion process always occurs for **specific data** constellations, for example, whenever the data for specific sensors occur, then data caching resources must be reserved. Additionally, the fusion data are not available at the earliest possible point in time. If unsynchronized sensors are used, a decision as to the filtering state in which the data are input into the fusion process must be made (see Sect. 4.6).

If the results of a centralized fusion module are not fed back to the sensors, fusion can be triggered at arbitrary points in time by an **external event**. This allows the data rate to be accommodated to match downstream processing, thus allowing for the resources to be accommodated. However, in terms of tracking accuracy, this is not an optimal solution (Bar-Shalom and Li 1995).

4.6 Original Data–Filtered Data–Predicted Data

In terms of the filtering state of data input into the fusion process, a distinction can be made between original data, filtered data, and predicted data.

In case of **original data**, the temporally unfiltered data are fed into the fusion process. This allows for optimal tracking.

If **filtered data** are used (e.g., in a decentralized architecture), optimal tracking can be achieved under restrictive conditions. However, if the filtered data are treated like unfiltered data and passed to a further filter for estimation, a chain of filters is established. This generally leads to higher signal propagation delay. Additionally, errors are now correlated; for an optimal estimation result, the filter model needs to take this into consideration.

It is also possible to use **predicted data** (e.g., on the basis of models). This approach is often used to relate the acquired measurement data to a point in time when a specific data constellation occurs and to consolidate the different measurements to so-called super measurements. Bar-Shalom and Li are of the opinion that this approach does not lead to optimal results in terms of the achievable estimation error for unsynchronized sensors (Bar-Shalom and Li 1995).

4.7 Parallel–Sequential

Another distinction which can be found in the referenced literature relates to the fusion algorithm. A distinction is made here between **parallel fusion**, where fusion of the existing measurements occurs in a single step, and **sequential fusion**, where the measurements are merged in multiple, sequential steps. If the systems are linear and the sensors are synchronized, then the two methods are equivalent (Bar-Shalom and Li 1995).

Dietmayer et al. also refer to explicit fusion in the case of synchronized sensors and parallel fusion and to implicit fusion for unsynchronized sensors and sequential fusion (Dietmayer et al. 2005).

5 Conclusions

In the author's opinion, data fusion is essential for meeting the requirements for future driver assistance systems and automated vehicles. This is particularly true of systems designed to improve safety.

Given an appropriate architecture design, the sensor fusion data system can represent an abstraction of the environmental perception of the deployed sensors. The applications can thus be developed independently of the environment acquisition system. The design of the situational analysis as an interface between fusion and application plays a central role then.

However, experience also shows that the equation "more sensors equals a better system" does not apply without restrictions in practical applications. For example, the overall system complexity increases with each sensor. Each sensor adds sensor-specific properties to the system. If this is not modeled or taken into consideration with a sufficient degree of accuracy, it may still be possible to partially improve certain aspects, but will at the same time impact the overall performance. Hall (2001) provides an overview of the typical pitfalls in designing a multi-sensor system.

References

Bar-Shalom Y (2002) Update with out-of-sequence measurements in tracking: exact solution. IEEE Trans Aerosp Electron Syst 3:769–777

Bar-Shalom Y, Li XR (1995) Multitarget-multisensor tracking – principles and techniques. YBS, Storrs

Bar-Shalom Y, Li XR, Kirubarajan T (2001) Estimation with applications to tracking and navigation – theory, algorithms and software. Wiley, New York

Becker JC (2002) Fusion der Daten der objekterkennenden Sensoren eines autonomen Straßenfahrzeugs (Fusion of data from object detecting sensors in an autonomous road vehicle). VDI-Verl, Düsseldorf

Bender E et al (2007) Anti collision system proreta – on the way to the collision avoiding vehicle – part 1: basics of the system. ATZ Wordwide 109(4):20–23

Darms M (2007) Eine Basis-Systemarchitektur zur Sensordatenfusion von Umfeldsensoren für Fahrerassistenzsysteme (A baseline system architecture for sensor data fusion of environment sensors for driver assistance systems) Fortschrittberichte VDI, Series 12, vol 653

Darms M, Rybski P, Urmson C (2008) Vehicle detection and tracking for the urban challenge, AAET 2008, 9th symposium, 13/14 Feb 2008, Braunschweig

Darms M, Rybski P, Baker C, Urmson C (2009) Obstacle detection and tracking for the urban challenge. IEEE Trans Intell Transp Syst 10(3):475–485

Dietmayer K, Kirchner A, Kampchen N (2005) Fusionsarchitekturen zur Umfeldwahrnehmung für zukünftige Fahrerassistenzsysteme (Fusion architectures for environmental perception in future driver assistance systems). In: Mauerer M (ed) Fahrerassistenzsysteme mit maschineller Wahrnehmung (Driver assistance systems with mechanical perception). Springer, New York, pp 59–87

Föllinger O (1990) Regelungstechnik – Einführung in die Methoden und ihre Anwendung (Control theory – introduction to methods and their application), 6th issue. Hüthig Buch Verlag, Heidelberg

Hall DL (2001) Handbook of multisensor data fusion. CRC Press, Boca Raton

Hall D, Llinas J (1997) An introduction to multisensor data fusion. Proc IEEE 85(1):6–23

Hall DL, McMullen SA (2004) Mathematical techniques in multisensor data fusion, 2nd edn. Artech House, Boston

Hänsler E (1997) Statistische Signale: Grundlagen und Anwendungen (Statistical signals: basics and applications), 2nd issue. Springer, Berlin

Holt Vv (2004) Integrale multisensorielle Fahrumgebungserfassung nach dem 4D-Ansatz (Integral multi-sensorial environmental acquisition based on the 4-D approach), Diss. Univ. der Bundeswehr, Munich (online publication), URL: urn:nbn:de:bvb:706–1072. http://athene-forschung.unibw.de/doc/85338/85338.pdf. Accessed 31 Jan 2015

Joerg KW (1994) Echtzeitfähige Multisensorintegration für autonome Mobile Roboter (Real-time capable multi-sensor integration for autonomous mobile robots). BI-Wiss.-Verl., Mannheim

Kampchen N, Dietmayer K (2003) Data synchronization strategies for multi-sensor fusion. In: 10th world congress on intelligent transport systems. Band Proceedings of IST 2003. Madrid

Klaus F (2004) Einführung in Techniken und Methoden der Multisensor-Datenfusion (Introduction to the techniques and methods of multi-sensor data fusion). Habil.-Schr. Univ. Siegen, online publication, URL: urn:nbn:de:hbz:467–575; http://www.ub.uni-siegen.de/pub/diss/fb12/2003/klaus/klaus.pdf. Accessed 31 Jan 2015

Klein LA (1999) Sensor and data fusion concepts and applications, 2nd issue, vol 35. SPIE, Bellingham

Lou RC, Kay MK (1991) Multisensor integration and fusion in intelligent systems. In: Autonomous mobile robots, Bd. 1. IEEE Computer Society Press, Los Alamitos

Lunze J (2006) Regelungstechnik Mehrgrößensysteme, digitale Regelung (Control theory on multiple value systems, digital control), vol 2. Springer, Berlin

Mauthener M et al (2006) Out-of-sequence measurements treatment in sensor fusion applications: buffering versus advances algorithms. In: Stiller C, Maurer M (Hrsg) 4. Workshop Fahrerassistenzsysteme FAS2006 (Driver assistance systems workshop), pp 20–30. fmrt, Karlsruhe

Mayr R (2001) Regelungsstrategien fur die automatische Fahrzeugführung: Längs- und Querregelung, Spurwechselund Überholmanöver (Control strategies for automatic vehicle guidance: longitudinal and latitudinal control, lane changing and overtaking manoeuvres). Springer, Tokyo

Narbe B et al (2003a) Datennetzkonzepte für die Sensordatenfusion – Teil 1 (Data network concepts for sensor data fusion – part 1). Elektronik Automotive 4:54–59

Narbe B et al (2003b) Datennetzkonzepte für die Sensordatenfusion – Teil 2 (Data network concepts for sensor data fusion – part 2). Elektronik Automotive 5:40–44

Robinson G, Aboutalib A (1990) Trade-off analysis of multisensor fusion levels. In: Proceedings of the 2nd national symposium on sensors and sensor fusion, No 2, Chicago, IL, pp 21–34

Schopper M, Henle L, Wohland T (2013) Intelligent Drive Vernetzte Intelligenz für mehr Sicherheit (Intelligent Drive – networked intelligence for improved safety). ATZExtra 5:106–114

Starke G (2005) Effektive Software-Architekturen – Ein praktischer Leitfaden (Effective software architectures – a practical guide), 2nd issue Hanser, Wien

Steinberg A, Bowman C, White F (1998) Revisions to the JDL data fusion model. Quebec City

Streller D (2006) Multi-Hypothesen-Ansatz zur Erkennung und Verfolgung von Objekten in Verkehrsszenen mit Laserscannern (Multi-hypotheses approach for detection and tracking of objects in traffic scenarios with laser scanners). VDI-Verl, Düsseldorf

Stüker D (2004) Heterogene Sensordatenfusion zur robusten Objektverfolgung im automobilen Straßenverkehr. (Heterogenous sensor data fusion for robust object tracking in automotive road traffic), Diss. Univ. Oldenburg, (online publication), URL: urn:nbn:de:gbv:715-oops-2319; http://oops.uni-oldenburg.de/201/1/stuhet04.pdf. Accessed 31 Jan 2015

Vogel O (2005) Software-Architektur – Grundlagen – Konzepte – Praxis (Software architecture – basics – concepts – practical applications), 1st issue Elsevier, Spektrum, Akad. Verl., Munich

Representation of Fused Environment Data

24

Klaus C. J. Dietmayer, Stephan Reuter, and Dominik Nuss

Contents

Abstract

The requirements for a vehicle environment representation increase with the complexity of advanced driver assistance systems and automatic driving. The ability of the current traffic situation to interpret and predict is essential for being able to automatically derive reasonable decisions. As a consequence, a state of the art vehicle environment representation has to incorporate all relevant

---

K.C.J. Dietmayer (✉) • S. Reuter • D. Nuss
Institut für Mess-, Regel- und Mikrotechnik, Universität Ulm, Ulm, Germany
e-mail: klaus.dietmayer@uni-ulm.de; stephan.reuter@uni-ulm.de; dominik.nuss@uni-ulm.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_25

dynamic objects as well as static obstacles and context information. While dynamic objects are typically described by an object-based representation using state variables, static obstacles as well as free space area are commonly modeled using grid-based methods. This chapter gives an introduction into both of these concepts.

The chapter is organized as follows: First, the difference between function-oriented and modular fusion architectures is discussed. Afterwards, the joint integrated probabilistic data association (JIPDA) filter is introduced, which is one method to realize an object-based environment model incorporating both state and existence uncertainties. Further, the representation of static obstacles with occupancy grids is described in detail and the incorporation of measurements of different sensor types is illustrated. Finally, several hybrid environment representations are introduced and an example for a strictly modular architecture, the hierarchical modular environment perception, is presented.

1 Requirements for Vehicle Environment Representations

A vehicle environment representation, often also referred to as a vehicle environment model, is understood to be a dynamic data structure in which all relevant objects and infrastructure elements in the proximity of the ego vehicle are contained. All elements have to be represented in a common reference system as accurately as possible with regard to position and time. The detection and the temporal tracking of the objects and infrastructure elements are performed continuously using on-board sensors such as cameras and RADAR (see ▶ Chaps. 17, “Automotive RADAR,” ▶ 18, “Automotive LIDAR,” ▶ 19, “Automotive Camera (Hardware),” and ▶ 20, “Fundamentals of Machine Vision”). In the future increasingly more information from high-precision, attributed digital maps and, if applicable, external information based on Car2x communication will be available and can be incorporated in the data fusion. Figure 1 shows examples of common elements contained in a vehicle environment representation.

The objects and structure elements which are relevant for a vehicle environment representation depend to a large extent on the functions that will be implementing it. For example, a blind spot assist requires only the information on whether there are currently objects in the rear or side area of the vehicle, the type of object is immaterial. More complex assistance systems, for example automatic emergency steer assist all the way through to automatic driving, require more extensive perceptiveness and information. In these cases, the distances, speeds, and dimensions of all traffic participants in the immediate surroundings as well as the drivable free space must be reliably recognized along with the lane markings. These complex driver assistance functions additionally require an effective situational evaluation which interprets the vehicle environment model and predicts the near future with a certain reliability based on the current situation. Economic reasons will also drive the transition from the previously dominating function-oriented

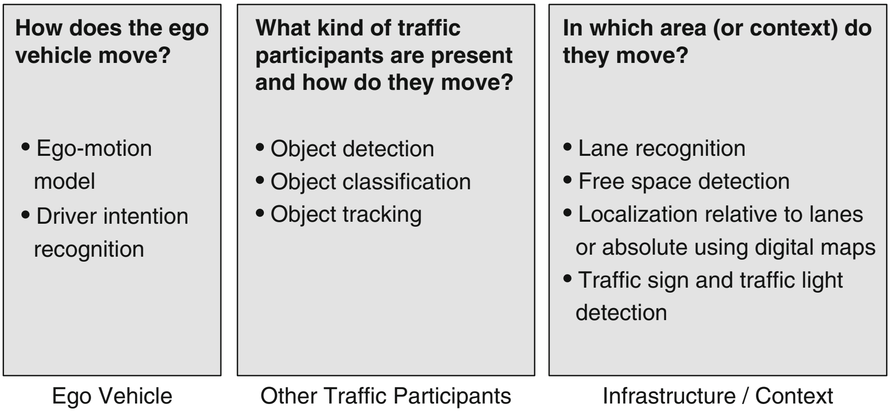


Fig. 1 Common elements of a vehicle environment representation

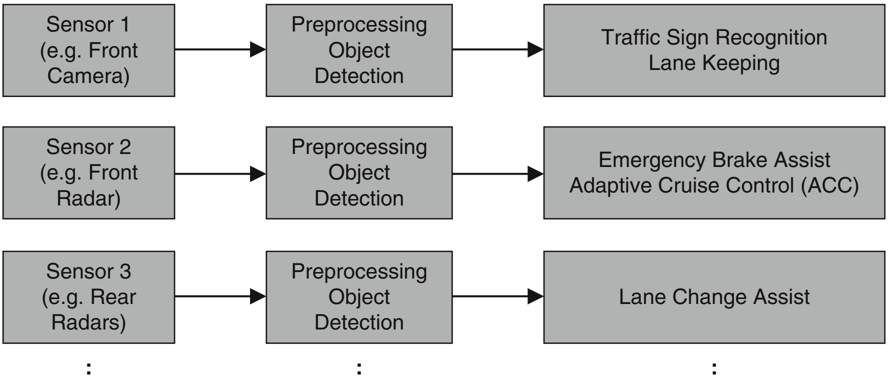


Fig. 2 Function-oriented, component-based architecture for driver assistance systems

architectures (Fig. 2) to a modular, generic architecture for vehicle environment representation which includes data fusion and that can preferably serve all driver assistance functions in a vehicle by means of a suitable interpretation and prediction of the current situation (Fig. 3). Adaptability to different function requirements is therefore a significant requirement for the architecture and data structure used in vehicle environment representation and it can only be warranted through consistent modularity and largely generic interfaces to the sensors as well as to the further processing stages of the situational interpretation. To date, no universally applicable fixed architecture principles exist and therefore various aspects of potential vehicle environment representation manifestations are discussed within the framework of this chapter.

In an object-based representation, all other traffic participants that are relevant for the representation, all relevant infrastructure elements as well as the ego vehicle

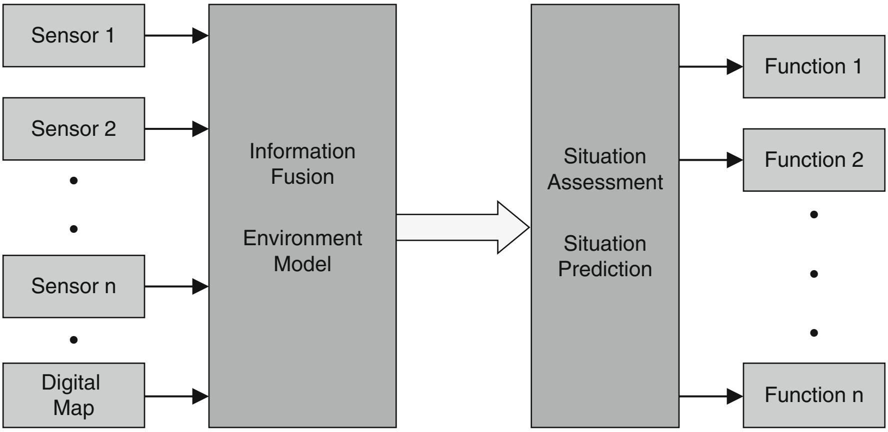


Fig. 3 Data flow in a modular architecture for driver assistance systems

are described by their own dynamic object model, generally a time-discrete state space model. Its states such as position, speed or 2D/3D object dimensions are updated continuously using sensor measurements and appropriate filtering processes (Sect. 2.3). Because measurements are fundamentally erroneous, these filtering processes should provide information on the current uncertainty of the vehicle environment representation to the further processing stages. This applies on the one hand to the uncertainties of the states themselves, generally expressed by the variance or covariance, but also to the existence probability, i.e., a gauge of whether the object detected by the sensor actually exists.

Since on-board sensors are preferred in environment perception, the object-based representation is given relative to the ego coordinate system. In principle, this description is sufficient for solving all driving tasks. It reaches its limits when context knowledge, for example from a high-precision digital map or through Car2x communication flows in as well. In these cases an absolute reference, i.e., a high-precision determination of the global location of the ego vehicle, is necessary.

A grid-based representation implements grid maps to divide the environment into fixed, identically dimensioned cells. The vehicle travels across this grid and the on-board sensors provide information on whether specific cells are free and therefore are drivable or whether there is an obstacle in the cell. This type of representation is suited primarily for representing static scenarios. It requires no model hypotheses and can therefore be classified as very robust with regard to model errors.

The following chapters will address in greater detail these two principles of representation as well as their applicable algorithms. There are also very promising approaches which combine the two representation forms. These will be briefly introduced at the end of the chapter.

2 Object-Based Representation

2.1 Sensor-Specific Object Models and Coordinate Systems

For object-based representation, the on-board sensors used to detect the environment make it practical to use the common frame of reference of the transported coordinate system of the ego vehicle. In accordance with the usual conventions for vehicles, a right-handed coordinate system will be used. The x-axis is in the direction of motion, the z-axis is vertical. The direction of the y-axis follows the definition of the right-handed coordinate system. Rotation about the x-axis is designated as rolling, about the y-axis as pitching, and about the z-axis as yawing.

In vehicle dynamic models, the center of mass of the vehicle is usually selected as the origin of the coordinate system. Such complex vehicle dynamic models are generally not necessary in the field of environment perception which requires merely comparably simple ego-motion estimation. For driving situations away from dynamic boundary conditions, at least, it is common to use purely kinematic, linear single-track models neglecting slip and skew of the wheels for ego-motion estimation models (Bar-Shalom and Fortmann 1988). A suitable reference point for multiple sensor applications (fusion systems) is the center of the rear axle of the ego vehicle projected onto the plane of the road. This choice has the advantage that given the mentioned model assumptions, the vehicle speed vector v always shows in the longitudinal direction (x-axis) of the vehicle, simplifying conversions. For isolated driver assistance functions with a single sensor (e.g., a purely RADAR-based ACC, see Fig. 2), the reference system is often selected as identical to the sensor coordinate system for pragmatic reasons.

Besides the ego vehicle, all other objects in the vehicle surroundings must be described by models as well. On account of the algorithm for object tracking which is described in greater detail below, these are generally formulated as time-discrete state space models in the plane.

A fundamental distinction is made between point models and spatially extended models whose length and width (2D) and possibly even height (3D) are modeled. A rectangle (2D) or cuboid (3D) serves as the basic form for extended models. However, the use of an extended model is only useful if the contour of the object can be observed by the sensors, which also depends inherently on the viewing angle. For example, the length of a vehicle driving immediately ahead cannot be determined independently of the implemented sensors, merely its width and height and even these only in the case that the appropriate contour-resolving sensors such as (stereo) cameras, laser scanners or high-resolution RADAR sensors are implemented. Because of this time-variant observability problem, the expansion estimation is often performed in parallel to the state estimation by a separate algorithm.

With regard to the state estimation, point models are modeled as so-called free mass models, which means that through the restriction to planar motion, the translational motion in the x and y directions as well as the rotation about the height axis $\dot{\psi}$ (yaw rate) are not coupled. Extended models (2D/3D) can be formulated as free mass models or as kinematic single-track models (see above) with corresponding coupled degrees of freedom. The latter is usually the best choice for all wheel-bound objects such as vehicles or bicyclists.

The relative position of each object is continually estimated based on the sensor measurements. Using the "constant velocity and yaw angle" model, the estimated state variables are the position in the plane, the speed in the plane, and the yaw angle of each object. A broadening of these models to consider the translational acceleration components as well as the yaw rate in the state vector is possible but only useful if sensors such as RADAR which can directly measure velocity components are available. Otherwise the acceleration data contains a noise effect, resulting from the two-fold differentiation in the filter of position measurements containing errors, which is so strong that the complete estimation result tends to be degraded. The selection of the best suited object model therefore is extremely dependent on the available sensor configuration and its measurement capabilities (see also ▶ Chaps. 17, "Automotive RADAR," ▶ 18, "Automotive LIDAR," ▶ 19, "Automotive Camera (Hardware)," and ▶ 20, "Fundamentals of Machine Vision"). Details on the formulation of object models can be found, for example, in Thrun (2005) or Bar-Shalom and Fortmann (1988). Particulars on the integration of various sensor types can be found in ▶ Chap. 23, "Data Fusion of Environment-Perception Sensors for ADAS".

2.2 State and Existence Uncertainties

For an object-based representation of the vehicle surroundings, it is necessary to specify the state and existence uncertainties for the individual objects in order to safeguard the triggering decisions of safety-relevant driver assistance functions based on the environment representation.

The state uncertainty of an object is usually described by a probability density function. In the case of a multidimensional, normally distributed probability density function, the state uncertainty is represented completely by the covariance matrix P. When estimating static parameters, the state uncertainty can be continually reduced through repeated measurements and the estimated value converges to the actual value, if no systematic errors such as an offset are present. In the estimation of dynamic states, the convergence to the actual value is not possible because of the movement of the object during the time between two subsequent measurements. Therefore, when evaluating the state estimation it is stipulated that the expected value of the estimation error is zero and that the variance of the actual estimation error corresponds to the estimated variance. An estimator with these characteristics is known as a consistent estimator.

The existence uncertainty, however, is at least as relevant for the realization of safety-relevant driver assistance functions as the state uncertainty. It expresses with which probability the object in the vehicle environment representation actually corresponds to a real object. Emergency braking, for example, should only be performed if there is a very high existence probability of the tracked object.

Whereas the state-of-the-art estimation of the state uncertainty is performed using the theoretically well-founded methods of the recursive Bayes estimation (see Sect. 2.3.1), current systems usually determine the existence of an object based on a heuristic quality criterion $q(x)$ of an object hypothesis. An object is considered confirmed if the quality criterion exceeds a sensor and application-dependent threshold ϑ. The quality criteria are based, for example, on the number of successful measurement value associations since the initialization of the object or the time between initialization of the object and the current time. In many cases the state uncertainty of the object or the quality criterion of another system are used for validation.

An alternative approach is the estimation of an object-specific existence probability. This first requires an application-specific definition of the object existence. Although some applications consider all real objects as existing, object existence can also be restricted to the relevant objects in the current application. A further restriction to the objects which can be detected with the current sensor setup is also possible. Contrary to the threshold ϑ of the quality criterion, determining the existence probability allows an interpretation based on probability. An existence probability of 90 % means, for example, that there is a 90 % probability that the measurement history and also the motion history of the object were generated by a real object. Consequently, assistance functions can use those objects whose existence probability exceeds an application-specific threshold.

2.3 Fundamentals of Multi-Object Tracking

The goal of object tracking is to estimate the time-variant state of all objects located in the field of view of the sensors. Changes in the object states result on the one hand through the motion of the objects as well as through the ego-motion of the observing vehicle. State-of-the-art object tracking is based on the recursive Bayes filter consisting of two parts: prediction and innovation. The prediction step models the movement of the object between two subsequent measurements in time based on an object-specific motion model (compare Sect. 2.1). Further, the ego-motion of the observing vehicle must be compensated in the prediction step. The ego-motion results in a change in the states (e.g., position) estimated relative to the ego vehicle as well as in an increased estimation uncertainty. In the subsequent innovation step the predicted object state is updated on the basis of the current sensor measurement, taking the measurement uncertainty into account.

In the following, the Bayes filter will first be introduced. Subsequently, the Kalman filter, which allows the analytic implementation of the Bayes filter for linear systems, will be explained and possibilities for the application to multi-object tracking introduced. The joint integrated probabilistic data association filter in particular will be discussed. It realizes a probabilistic association of the obtained measurements to the currently tracked objects and also estimates the objects' existence probability as called for in Sect. 2.2.

2.3.1 Recursive Bayes Filter

In the recursive Bayes filter (Mahler 2003), the estimated states of an object and the associated spatial uncertainty are represented by a probability density function (PDF):

$$p_{k+1}(x_{k+1}) = p_{k+1}(x_{k+1}|Z_{1:k+1}), \tag{1}$$

which is dependent on all measurements $Z_{1:k+1} = \{z_1, \ldots, z_{k+1}\}$ received up to point in time $k + 1$.

The motion model of an object for the time between two subsequent measurements is given by:

$$x_{k+1|k} = f(x_k) + v_k, \tag{2}$$

where v_k is an additive noise process. Alternatively, the motion equation can also be described by a Markov transition density:

$$f_{k+1|k}(x_{k+1}|x_k). \tag{3}$$

Applying the first-order Markov characteristic, the predicted state x_{k+1} of the object depends only on the state x_k since the latter implicitly contains the measurement history $Z_{1:k} = \{z_1, \ldots, z_k\}$. The prediction of the current object state x_k to the next measurement at time $k + 1$ is made based on the Chapman-Kolmogorov equation:

$$p_{k+1|k}(x_{k+1}|x_k) = \int f_{k+1|k}(x_{k+1}|x_k)\, p_k(x_k)\, dx_k. \tag{4}$$

Subsequently, the predicted PDF of the object state is updated using the measurement z_{k+1}. The measurement process of the sensor is described by the measurement equation:

$$z_{k+1|k} = h_{k+1}(x_{k+1}) + w_{k+1}. \tag{5}$$

The stochastic noise w_{k+1} represents the error of the measurement model. The measurement equation transforms the state of an object into the measurement space of the sensor and therefore allows the innovation of the object state in the measurement space. The innovation in the measurement space is an advantage since the transformation of the measurement into state space generally is not possible due to the non-invertible measurement equation. An alternative representation of the measurement equation is the likelihood function:

$$g(z_{k+1}|x_{k+1}),$$

which results from the measurement Eq. 5. The state innovation is subsequently performed using the Bayes equation:

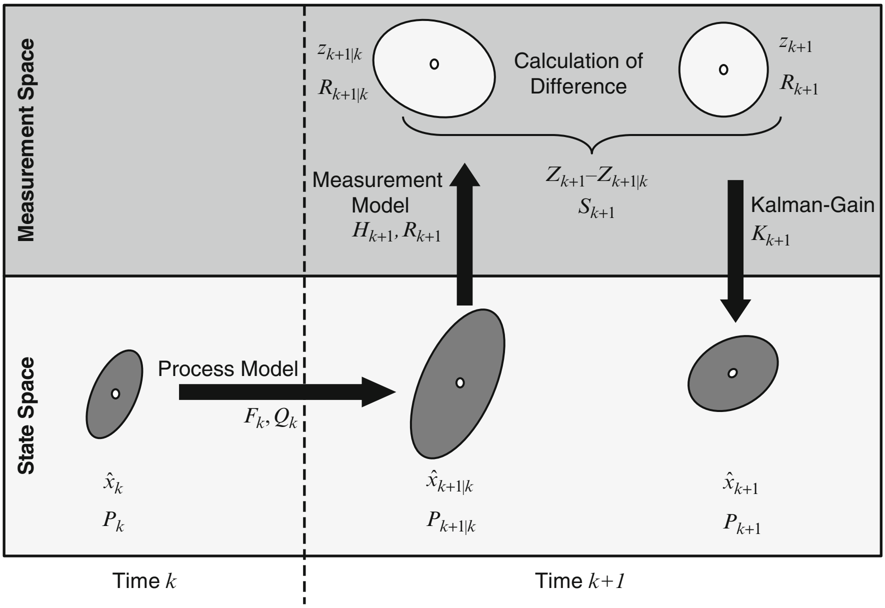


Fig. 4 Process of the Kalman filter: state prediction based on the process model, state innovation through transformation into measurement space and the Kalman gain

$$p_{k+1}(x_{k+1}|z_{k+1}) = \frac{g(z_{k+1}|x_{k+1})p_{k+1|k}(x_{k+1}|x_k)}{\int g(z_{k+1}|x_{k+1})p_{k+1|k}(x_{k+1}|x_k)dx_{k+1}}. \tag{6}$$

The recursive estimation process described by the prediction step Eq. 4 and the innovation step Eq. 6 is known as the Bayes filter. Besides the process and measurement equations, the process requires only an a priori PDF for the object state $p_0(x_0)$ at time $k = 0$.

2.3.2 Multi-Instance Kalman Filter

Under the assumption of normally distributed signals as well as linear process and measurement models, the Kalman filter (Kalman 1960) allows the analytic implementation of the Bayes filter. Since a Gaussian distribution is completely described by its first two statistic moments, i.e., the mean $\hat{x}$ as well as the corresponding covariance matrix P, the temporal filtering of the moments presents an exact mathematical solution. It follows that given these assumptions, the Kalman filter is a Bayes-optimal state estimator. Figure 4 illustrates the process of the Kalman filter, which will be described in detail in the following section.

The initial state of an object in the Kalman filter is given by a multi-dimensional Gaussian distribution:

$$N(x, \hat{x}_k, P_k) = \frac{1}{\sqrt{\det(2\pi P_k)}} \exp\left(-\frac{1}{2}(x - \hat{x}_k)^T P_k^{-1}(x - \hat{x}_k)\right), \tag{7}$$

with mean $\hat{x}_k$ and covariance P_k. For the case of linear process and measurement models, Eqs. 2 and 5 can be written as follows:

$$x_{k+1|k} = F_k x_k + v_k \tag{8}$$

$$z_{k+1|k} = H_{k+1} x_k + w_{k+1}, \tag{9}$$

where F_k and H_{k+1} represent the system matrix and the measurement matrix for the current measurement. The process noise v_k and measurement noise w_{k+1} are assumed to be zero-mean, Gaussian noise and the two noise processes are uncorrelated.

The prediction step of the Kalman filter is performed through the independent prediction of the expected value and the covariance:

$$\hat{x}_{k+1|k} = F_k \hat{x}_k \tag{10}$$

$$P_{k+1|k} = F_k P_k F_k^T + Q_k. \tag{11}$$

The covariance matrix $Q_k = E\left\{v_k v_k^T\right\}$ of the process noise represents the uncertainty of the process model, for example, the maximum possible acceleration of an object when using a process model for constant velocity.

By applying the measurement matrix H_{k+1}, the predicted state $\hat{x}_{k+1|k}$ can be used to determine the predicted measurement:

$$z_{k+1|k} = H_{k+1} \hat{x}_{k+1|k}, \tag{12}$$

as well as the corresponding covariance matrix:

$$R_{k+1|k} = H_{k+1} P_{k+1|k} H_{k+1}^T. \tag{13}$$

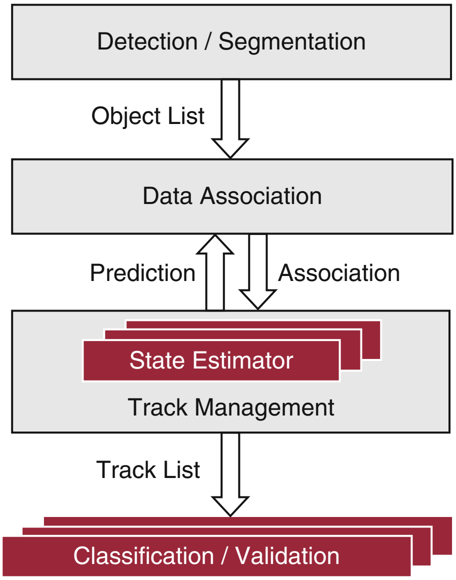
In the following innovation step, a measurement z_{k+1} with corresponding covariance $R_{k+1} = E\left\{w_{k+1} w_{k+1}^T\right\}$ is used to update the state. Applying the expected value of the predicted measurement $z_{k+1|k}$ as well as the actual measurement value z_{k+1} yields the measurement residual γ_{k+1} and the corresponding innovation covariance matrix S_{k+1}:

$$\gamma_{k+1} = z_{k+1} - z_{k+1|k} \tag{14}$$

$$S_{k+1} = R_{k+1|k} + R_{k+1} = H_{k+1} P_{k+1|k} H_{k+1}^T + R_{k+1}. \tag{15}$$

The filter gain K_{k+1} can be calculated using the innovation covariance matrix S_{k+1}, the predicted state covariance $P_{k+1|k}$ as well as the measurement matrix H_{k+1}:

Fig. 5 Multi-instance Kalman filter



$$K_{k+1} = P_{k+1|k} H_{k+1}^T S_{k+1}^{-1}. \tag{16}$$

The update of the object state estimate and corresponding covariance resulting from the current measurement is obtained using the following equations:

$$\hat{x}_{k+1} = \hat{x}_{k+1|k} + K_{k+1}\left(z_{k+1} - z_{k+1|k}\right) \tag{17}$$

$$P_{k+1} = P_{k+1|k} - K_{k+1} S_{k+1} K_{k+1}^T. \tag{18}$$

The application of the Kalman filter to systems with nonlinear process or measurement equations can be realized by means of the extended Kalman filter (EKF) (Bar-Shalom and Fortmann 1988) as well as the unscented Kalman filter (UKF) (Julier et al. 2000). While the EKF linearizes the process matrix F_k or measurement matrix H_{k+1} using a Taylor series approximation, the goal of the UKF is a stochastic approximation based on so-called sigma points (Julier et al. 2000).

The Kalman filter presents an optimal state estimator for an object and a measurement. Within the context of vehicle environment perception, however, it is necessary to simultaneously track multiple objects. In the literature, multi-object tracking is often realized using multi-instance Kalman filters as shown in Fig. 5, in which every object is tracked by an object-specific Kalman filter. Since not every Kalman filter represents a relevant or an actually existing object, a subsequent classification and validation of the estimated tracks is necessary. In order to reduce the amount of processed data, object hypotheses are generated in the detection or segmentation step. One example of a detection algorithm is the pedestrian detection in video images introduced in ▶ Chap. 22, "Camera Based Pedestrian Detection".

In the data association step, the obtained measurements are assigned to the available Kalman filters and the state of the objects are updated with the measurements, whereby the data association is ambiguous in many cases due to missed detections and false alarms.

The association process which requires the lowest computational effort is the nearest neighbor (NN) algorithm, which updates each object with the closest measurement relative to the state. The Mahalanobis distance (Bar-Shalom and Fortmann 1988) is commonly used to obtain the closest measurement. In scenarios with objects located close together, the NN algorithm often leads to a single measurement being used to update multiple objects. However, this contradicts the assumption that a measurement is generated by one object at the most. The global nearest neighbor algorithm guarantees the compliance with this assumption through the calculation of an optimal association of all tracks and measurements. Both algorithms make a hard and possibly erroneous association decision at a given time step which cannot be reversed and in the event of a wrong decision often leads to the loss of a currently tracked object.

The basic idea of the probabilistic data association (PDA) (Bar-Shalom and Tse 1975) is therefore to perform a weighted update of the object state using all association hypotheses in order to avoid hard, possibly erroneous association decisions. The association matrix $A = \beta_{ij}$ of the PDA algorithm represents all association probabilities:

$$\beta_{ij} = p\left(x_i \leftrightarrow z_j\right), \tag{19}$$

for the object hypotheses $x_1, \ldots, x_n$ and the measurements $z_1, \ldots, z_m$. Besides the association uncertainty, it is also possible that an object does not generate a measurement at time $k + 1$. In the event of such a so-called missed detection with weight β_{i0}, the improved state corresponds to the predicted state.

On account of the weighted update of an object with the m measurements obtained at time $k + 1$, the a posteriori probability density of the object state is given by the weighted superposition of the individual association hypotheses:

$$p\left(x_{i,k+1} | z_1, \ldots, z_m\right) = \sum\nolimits_{j=0}^{m} \beta_{ij} p\left(x_{i,k+1} | z_j\right). \tag{20}$$

Here, the PDF $p\left(x_{i,k+1} | z_j\right)$ represents the state innovation of the object x_i with measurement z_j. For the missed detection $j = 0, p\left(x_{i,k+1} | z_0\right)$ represents the predicted state estimate.

Because of the probabilistic data association, the a posteriori PDF contained in Eq. 20 no longer follows a Gaussian distribution since the superposition of multiple Gaussian distributions is generally multimodal. It is therefore necessary to approximate Eq. 20 by a single Gaussian distribution in order to be able to use the Kalman filter equation in the next innovation step again. For every innovation hypothesis, the updated state is first calculated by applying Eq. 17:

$$x_{ij} = \hat{x}_{i,k+1|k} + K_{ij}\left(z_j - z_{i,k+1|k}\right). \tag{21}$$

In the event of missed detection,

$$x_{i0} = \hat{x}_{i,k+1|k}. \tag{22}$$

The updated expected value for the state of object x_i is now determined from the weighted average of the innovated states of all association hypotheses:

$$\hat{x}_{i,k+1} = \sum\nolimits_{j=0}^{m} \beta_{ij} x_{ij}. \tag{23}$$

The innovation of the state covariance is given by the weighted accumulation of the state uncertainties of the individual association hypotheses:

$$P_{i,k+1} = \sum\nolimits_{j=0}^{m} \beta_{ij}\left[P_{i,k+1|k} - K_{ij}S_{ij}K_{ij}^{T} + \left(x_{ij} - \hat{x}_{i,k+1}\right)\left(x_{ij} - \hat{x}_{i,k+1}\right)^{T}\right]. \tag{24}$$

The third addend in Eq. 24 describes the additional uncertainty which results on account of the approximation by a single Gaussian distribution.

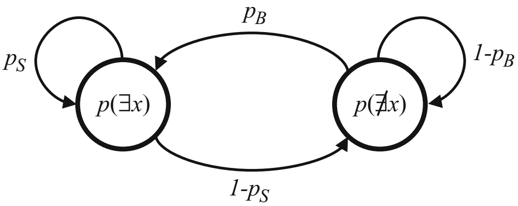
Because of the weighted updates, the PDA algorithm is very well suited for tracking a single object in scenarios with missed detections and a high number of false positive detections (false alarms). A disadvantage of the PDA algorithm, however, is the fact that a measurement can have a high association probability for more than one object. This contradicts the assumption made for the standard measurement model that a measurement originated from at the most one object. An enhancement of the PDA algorithm is the joint probabilistic data association (JPDA) filter algorithm which calculates the association weighting based on global association hypotheses (Bar-Shalom and Fortmann 1988). The calculation of the required association weights β_{ij} is presented in the following section using the example of the joint integrated probabilistic data association (JIPDA) filter, which is closely related to the PDA and JPDA filters.

2.3.3 Joint Integrated Probabilistic Data Association (JIPDA) Filter

The joint integrated probabilistic data association (JIPDA) algorithm (Musicki and Evans 2004) embodies a multi-object tracking algorithm which besides calculating the probabilistic data association weights also supplies an estimate for the object-specific existence probability. Since the existence of an object is dependent upon the detection and false alarm probabilities of the sensor as well as the data association, the integrated existence estimation is a useful enhancement to the probabilistic association process presented in the previous section.

The innovation of the object existence is completed analogously to the updating of the state in a prediction and an innovation step. The existence prediction is made based on a first-order Markov model (see Fig. 6). The predicted existence of an object is given by the Markov chain:

Fig. 6 Markov chain for predicting the existence probability



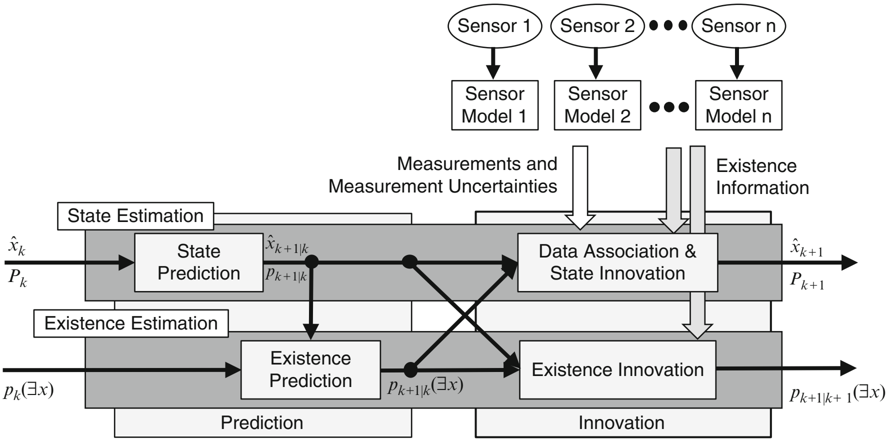


Fig. 7 Coupled Markov chains for the state and existence estimation

$$p_{k+1|k}(\exists x) = p_S p_k(\exists x) + p_B p_k(\nexists x), \tag{25}$$

where the probability p_s represents the persistence probability of the object and p_B the probability of the appearance of an object. It follows then that the probability of an object disappearing is given by $1 - p_S$. In the innovation step, the data association weights are calculated along with the a posteriori existence probabilities $p_{k+1}(\exists x)$. The latter depends upon how many association hypotheses validate the existence of the object.

Since the persistence probability of an object depends on the current object state and the a posteriori existence probability in turn depends on the data association, the JIPDA filter can be interpreted as the coupling of two Markov chains as illustrated in Fig. 7. The upper Markov chain represents the state prediction and innovation given by the Kalman filter while the lower Markov chain represents the prediction and innovation of the existence probability.

It follows then that in the prediction step of the JIPDA filter, the states of all objects are first predicted using Kalman filter Eqs. 10 and 11 from the preceding section. Subsequently, the predicted existence probability of object x_i is calculated based on the state-dependent persistence probability $0 < p_S(x_i) < 1$:

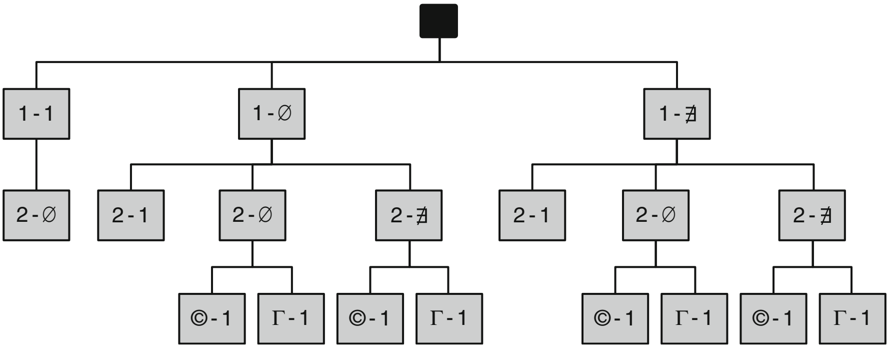


Fig. 8 Association tree for two objects and one measurement. The first symbol of each node represents an element from the set of objects, the second symbol an element from the set of measurements

$$p_{k+1|k}(\exists x_i) = p_S(x_i)p_k(\exists x_i). \tag{26}$$

The explicit dependency of the predicted existence probability on the object state represents the first coupling of the two Markov chains. After the state and the existence have been predicted to the time of the next measurement, the data association is carried out by means of the process introduced in Munz (2011). To this end, it is first necessary to define the set of all measurements:

$$Z = \{z_1, \ldots, z_m, \emptyset, \nexists\}, \tag{27}$$

where the two pseudo-measurements $\emptyset$ and $\nexists$ represent the missed detection and the non-existence of an object, respectively. Further, set X consists of the n objects currently being tracked as well as the two special elements © and $\varGamma$ for the false alarm source and additional new objects, respectively:

$$X = \{x_1, \ldots, x_n, \copyright, \varGamma\}. \tag{28}$$

The assignment of an object to a measurement is therefore given by a pair $e = (x \in X, z \in Z)$. In a complete assignment hypothesis given by the set $= \{e_i\}$, the objects $x_1, \ldots, x_n$ as well as the measurements $z_1, \ldots, z_m$ must be assigned exactly once. The special elements $\emptyset$, $\nexists$, ©, and $\varGamma$ may be used multiple times in an assignment hypothesis.

An intuitive representation of all possible assignment hypotheses is possible using a hypothesis tree. Figure 8 shows the example of the hypothesis tree for a situation with two objects and one measurement, where each node of the tree represents an element assignment e. Every path from the root node to a leaf node represents a complete assignment hypothesis $E_l = \left\{e_0, \ldots, e_{L(l)}\right\}$. The probability of an assignment hypothesis E_l can therefore be calculated using the product of the element assignment probabilities:

$$p(E_l) = \prod\nolimits_{e \in E_l} p(e). \tag{29}$$

On account of the combinatorial complexity, the number of assignment hypotheses grows exponentially, making the calculation of all hypotheses practical for only a small number of measurements and objects. A reduction of the possible assignment hypotheses can be achieved through gating processes, for example, which use the innovation covariance S to exclude highly improbable element assignment hypotheses. In addition, the probabilities of the element assignments e can be calculated in advance and stored in look-up tables as a means to improve the calculation efficiency.

In the following, the calculation rules are described for the five categories of element assignment hypotheses which use measurement value specific true positive probabilities in order to allow sensor-specific evidence to be considered for the existence of an object in the JIPDA filter. A true positive assignment node corresponds to the assignment of a track x_i to a measurement z_j. The probability of a true positive assignment node is given by:

$$p\left(e = \left\{x_i, z_j\right\}\right) = p_{k+1|k}(\exists x_i) p_{TP}\left(z_j\right) p_D(x_i) g\left(z_j | x_i\right), \tag{30}$$

where $p_{TP}(z_j)$ represents the measurement value specific true positive probability and $p_D(x_i)$ corresponds to the sensor specific detection probability for an object with state x_i. The likelihood for a measurement z_j for the object x_i corresponds to:

$$g\left(z_j | x_i\right) = \frac{1}{p_G} N\left(z, z_j - z_{i,k+1|k}, S_{ij}\right), \tag{31}$$

where p_G represents the gating probability, i.e., the probability that the true measurement of an object lies within the gating area. Obviously a high probability for a true positive assignment node requires a high predicted existence probability, a high true positive probability as well as a high likelihood.

A false positive assignment node corresponds to the assignment of a measurement to the false alarm source. The corresponding probability depends on the true positive probability $p_{TP}(z_j)$ as well as on the spatial false alarm probability $p_c(z_j)$:

$$p\left(e = \left\{©, z_j\right\}\right) = \left(1 - p_{TP}\left(z_j\right)\right) p_c\left(z_j\right). \tag{32}$$

The third assignment category represents the missed detections, i.e., the false negative assignment nodes whose probability is defined as follows:

$$p(e = \{x_i, \emptyset\}) = \left(1 - p_D(x_i)\right) p_{k+1|k}(\exists x_i). \tag{33}$$

The non-existence of an object is represented by the true negative assignment with probability:

$$p(e = \{x_i, \nexists\}) = 1 - p_{k+1|k}(\exists x_i), \tag{34}$$

which depends solely on the predicted existence probability of the object. The fifth assignment category assigns a measurement z_j at time $k + 1$ to a newly appearing object:

$$p\left(e = \left\{\Gamma, z_j\right\}\right) = p_{TP}\left(z_j\right) p_{\Gamma}\left(z_j\right). \tag{35}$$

The assignment to an appearing object was first introduced in Munz (2011) and allows the explicit modeling of the object initialization based on the hypothesis tree. Besides the sensor true positive probability, the spatial birth probability $p_{\Gamma}(z_j)$ is also required. The birth probability is modeled so that it is relatively low near already existing objects and is significantly greater at the periphery of the sensors' detection range.

The introduced calculation rules for the element assignment probabilities now allow the probabilities of all assignment hypotheses to be determined. Based on the set of all assignment hypotheses, the existence probability and the assignment weights β_{ij} can now be calculated. The a posteriori existence probability of an object x_i can be calculated by means of marginalization:

$$p_{k+1}(\exists x_i) = \frac{\sum_{E \in E_i^{\exists}} p(E)}{\sum_E p(E)}, \tag{36}$$

where the set $E_i^{\exists}$ represents all hypotheses in which track x_i exists:

$$E \in E_i^{\exists} \Leftrightarrow (x_i, \nexists) \cap E = \varnothing. \tag{37}$$

The calculation of the assignment weights:

$$\beta_{ij} = \frac{\sum_{E \in E_{ij}^{TP}} p(E)}{\sum_{E_i^{\exists}} p(E)}, \tag{38}$$

is also performed via marginalization. Here the set E_{ij}^{TP} represents all hypotheses which contain the assignment of measurement z_j to object x_i:

$$E \in E_{ij}^{TP} \Leftrightarrow \left(x_i, z_j\right) \in E. \tag{39}$$

The states of all objects are subsequently updated on the basis of Eqs. 23 and 24 and the assignment weights given in Eq. 38.

2.3.4 Random Finite Set (RFS) Approach, PHD and CPHD Filters

An alternative to multi-object tracking with multi-instance Kalman filters is given by the multi-object Bayes filter introduced in Mahler (2007) which models the multi-object states X as well as the measurements Z obtained at time $k + 1$ as random finite sets. A random finite set $X = \{x_1, \ldots, x_n\}$ in this case is a random variable in which both the number of objects n (including $n = 0$) as well as the individual states x_i are random. Further, the states of a random finite set are unordered and hence permutation invariant. The prediction and update equation of the multi-object Bayes filter correspond to Eqs. 4 and 6, however the state vectors x and the measurements z are replaced by the random finite sets X and Z.

In the prediction step, the multi-object Bayes filter uses a multi-object Markov density which along with the motion of objects also accounts for the appearance and disappearance of objects. The representation of objects by a random finite set additionally allows the modeling of dependencies between the objects in the prediction step, while the use of a multi-instance Kalman filter presupposes the statistical independence of the objects. In the context of vehicle environment perception, the dependencies between objects is not limited to the immediate spatial proximity since when a vehicle driving in front of another vehicle brakes, the trailing vehicle brakes as well.

The update step of the multi-object Bayes filter is based on the multi-object likelihood function, which avoids an explicit data association by averaging over all possible association hypotheses. The multi-object likelihood function can also be calculated using a hypothesis tree (Reuter et al. 2013a). In contrast to the hypothesis tree of the JIPDA algorithm shown in Fig. 8, the modeling of the non-existence of an object is not necessary since this is represented by a further realization of the random variable X. Furthermore, the nodes for appearing objects are not necessary because the appearance of objects is realized explicitly through the multi-object Markov density.

The multi-object Bayes filter can be implemented by means of sequential Monte Carlo (SMC) methods. However, due to the high-dimensional state space this is only applicable for a low number of objects. Based on the constant-gain approximation of the Bayes filter, an approximation of the a posteriori multi-object distribution by the first moment (PHD filter) is recommended in Mahler (2003) to reduce the computational effort. While the first moment of a probability density function is given by the mean, the first moment of the multi-object distribution is given by an intensity function where the integral over the intensity function represents the estimated number of objects in the respective area. A disadvantage of the PHD filter is the considerable fluctuation in the estimated number of objects due to the low memory of the filter and the approximation of the cardinality distribution by a Poisson distribution. In the case of the Kalman filter, an improved state estimation is obtained through the representation of the probability distribution through the first and second statistic moment. Since an approximation using the first and second statistic moment would be computationally intensive, the cardinalized probability hypothesis density (CPHD) filter, which propagates the intensity function and cardinality distribution over time rather than the multi-object distribution, is introduced in Mahler (2007). The CPHD filter thus represents a partial approximation of the multi-object distribution by the second statistical moment. Compared to the PHD filter, the CPHD filter distinguishes itself through the stable estimation of the number of objects but requires exact knowledge of the false alarm process beforehand which, however, is not available in the context of vehicle environment detection on account of the different environment situations.

A disadvantage of the PHD as well as the CPHD filter compared to the JIPDA filter introduced in Sect. 2.3.3 is the absence of the estimate of the existence probability for the tracked objects. The cardinality balanced multi-object multi-Bernoulli (CB-MeMBer) filter (Vo et al. 2009), a further approach to approximating the multi-object Bayes filter, approximates the multi-object probability density function by a multi-Bernoulli distribution and propagates its parameters over time. A multi-Bernoulli distribution represents M objects by M statistically independent Bernoulli distributions, where the Bernoulli distribution of each object with existence probability r is given by a singleton with spatial distribution $p(x)$ and the object does not exist with probability $1 - r$. An application of the CB-MeMBer filter in vehicle environment detection is investigated in Reuter et al. (2013b), Stiller et al. (2011).

The delta generalized labeled multi-Bernoulli (δ-GLMB) filter introduced in Vo and Vo (2013) enables an analytic implementation of the multi-object Bayes filter, however due to its complexity it is only suited as a reference implementation as well as a starting point for further approximation. Based on the δ-GLMB, the labeled multi-Bernoulli (LMB) filter was derived in Reuter et al. (2014). It provides significantly better tracking results than the previously applied approximations from the multi-object Bayes filter and at the same time allows a real-time capable implementation. The use of the LMB filter in the field of vehicle environment perception is the subject of current research.

2.4 Self-Localization and Inclusion of Digital Maps

Automatic detection for object recognition and object tracking, as described in the previous section, can be improved considerably through a priori information such as the number and width of driving lanes, lane branching, turn-offs, and intersection topologies as well as the position of traffic signs and traffic lights. Even though current commercially available maps contain very few such attributes, it is probable that automatic driving in more complex surroundings will not be possible without providing such support for the automatic detection. In addition, an environment evaluation profits from such map information since recognized objects can be evaluated in the context of the traffic area.

However, because digital maps are referenced absolutely, primarily in UTM or WGS84 coordinates (see ▶ Chap. 26, “Digital Maps for ADAS”), the use of attributes entered in them requires high-precision self-localization by the ego vehicle in the map. The precision of standard GNSS systems is not always sufficient. In addition, the reception in inhabited or wooded area is often severely impaired by multi-path effects and blocking-out of satellites. Because only the position on the map is crucial though, self-localization with sufficient precision is possible on the basis of automatically detected, recognizable landmarks which are also entered in the map. An example which is simple to realize is the determination of exact lateral position in a lane also contained in the map.

If this alignment is successful, all detected dynamic objects from the object-based representation can be assigned in the context of the map. It is common to create a layer structure in which, for example, the lowest level is the street topology, higher layers then contain lanes and other static attributes, and the detected dynamic objects can be found in an even higher layer. An overview focused on GNSS can be found in Skog and Handel (2009). Special aspects, especially regarding the use of different variations of digital maps are addressed in Levinson and Thrun (2010), Mattern et al. (2010), Schindler (2013), Konrad et al. (2012).

2.5 Time Aspects

Besides the correct spatial correlation, the determination of the correct time reference for all elements which flow into the vehicle environment representation poses a no less daunting challenge. The latter is after all expected to contain a timely consistent image of reality.

Because the sensors for automatic detection generally are not synchronized or cannot be synchronized and in addition possess different latencies due to the differences in the complexity of their pre-processing stages, at the very least a global system time is required to which all other measurements and incoming information is referenced. This ensures that the point in time of a measurement for a given sensor or the incoming external information can be collated by other measurements. For example, for the filtering processes based on the recursive Bayes filter, as described in Sect. 2.3, it is a prerequisite that measurements are introduced in chronological order. In order to make this possible, it may be necessary to wait for a slower sensor, i.e., one with a higher latency, before the measurements can be fed into the filter. This process is called buffering and has the disadvantage that the slowest sensor determines the overall latency of the vehicle environment representation. However, all variations have shown that such latency is principally unavoidable; the vehicle environment representation will always lag behind the real situation. Depending upon the configuration, this can be on the scale of up to several 100 ms. This latency must be considered in the situation prediction and action planning of an assistance function dependent upon it. Further information on the sources of the arising latency can be found, for example, in ▶ Chap. 23, "Data Fusion of Environment-Perception Sensors for ADAS".

3 Environment Representation with Grid Maps

3.1 Grid Map Concept

Grid maps partition the vehicle environment into cells. Each of these cells represents a location for which the respective cell contains information. Partitioning into cells is a spatial discretization of the vehicle environment. When the sensor data from different sensors allow inferences on the state of cells, then the mapping in grid maps corresponds to an indirect form of sensor data fusion.

It is possible to characterize grid maps based upon whether they depict space in two-dimensions or in three. The type of information stored in the individual cells also varies, as well as the size and shape of the cells. Grid maps can be referenced to a stationary, global coordinate system or to a local, vehicle-based coordinate system and can be generated online during operation or offline in a post-processing step.

Early publications on grid maps for the purpose of environment modeling originated in the field of robotics and described two-dimensional maps generated in real-time and whose cells indicated whether the space they represented was occupied or free (Elfes 1989; Thrun 2005). These maps are known as occupancy grid maps. Even though many different forms of grid maps have been introduced since then, occupancy grid maps remain an important foundation for many applications (Nuss et al. 2013; Petrovskaya et al. 2012).

The following sections describe how a grid map can be generated. It is necessary to know where the vehicle and its sensors are located with reference to the grid map. This can be determined through ego-motion estimation. The generation of occupancy grid maps is done primarily using distance measurement sensors such as laser or RADAR sensors. Furthermore, the generation of grid maps based on camera data is also being pursued. Moving objects play a special role in the construction of grid maps, especially if measurements are filtered over time. In addition, efficient memory management is an important issue with regard to the practical application.

3.2 Ego-Motion Estimation

3.2.1 Calculation of Vehicle Ego-Motion Through Dead Reckoning

In order to generate a consistent grid map, it is necessary to account for the motion of the vehicle. For the generation of a two-dimensional grid map, it is also necessary to estimate the pose (position and orientation) of the vehicle.

A simple possibility to estimate the absolute vehicle position results from the use of GNSS. The disadvantage of this method lies in the fact that reception is not guaranteed and that the precision fluctuates greatly. For many applications it is therefore advisable to select an arbitrary relation between the grid map and a global coordinate system and only to consider the relative motion of the vehicle. This can be done using dead reckoning, which is described in detail in the following section.

The vehicle pose at two points in time $(t_1,\ t_2),\ t_2 > t_1$ is shown in Fig. 9. At time t_1, let the vehicle pose $p_1 = [x_1\, y_1\, \psi_1]^T$ be known in the coordinate system of the grid map. Here, x and y correspond to the vehicle position and ψ to the orientation (yaw angle) of the vehicle. The goal is now to enter in the grid map a measurement taken at time t_2. Hence, pose p_2 at time t_2 must be determined. Vehicle velocity v and the yaw rate $\dot{\psi}$ are measurable. In addition, it is assumed that at every point in time, the vehicle is moving solely in the direction of the yaw angle ψ.

The change in vehicle pose is given by:

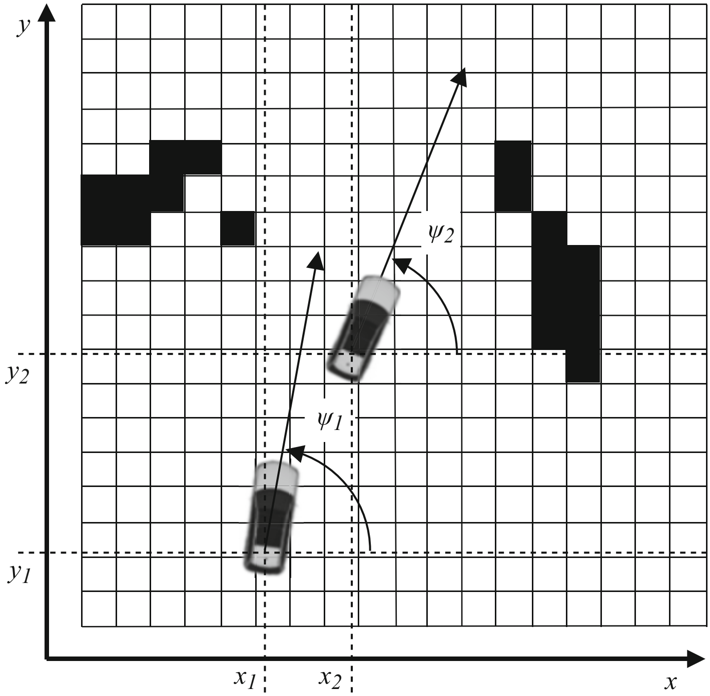


Fig. 9 Pose of the vehicle in the grid map measured at two different points in time

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\psi} \end{bmatrix} = \begin{bmatrix} v\cos(\psi) \\ v\sin(\psi) \\ \dot{\psi} \end{bmatrix}, \tag{40}$$

so that vehicle pose p_2 can be calculated from:

$$p_2 = p_1 + \int_{t_1}^{t_2} \begin{bmatrix} \dot{x}(t) \\ \dot{y}(t) \\ \dot{\psi}(t) \end{bmatrix} dt. \tag{41}$$

The assumption follows that the vehicle travels with constant velocity $\overline{v}$ and constant yaw rate $\overline{\dot{\psi}} \neq 0$ on a circular path during time interval $[t_1, t_2]$, with $\Delta t := t_2 - t_1$. In this case, the integral in Eq. 41 can be solved and

$$p_2 = p_1 + \begin{bmatrix} \frac{\overline{v}}{\overline{\dot{\psi}}}\left(\sin\left(\psi_1 + \overline{\dot{\psi}}\Delta t\right) - \sin(\psi_1)\right) \\ \frac{\overline{v}}{\overline{\dot{\psi}}}\left(-\cos\left(\psi_1 + \overline{\dot{\psi}}\Delta t\right) + \cos(\psi_1)\right) \\ \overline{\dot{\psi}}\Delta t \end{bmatrix} \tag{42}$$

results. When driving straight ahead with $\overline{\dot{\psi}} = 0$

$$p_2 = p_1 + \begin{bmatrix} \bar{v}\Delta t \cos(\psi_1) \\ \bar{v}\Delta t \sin(\psi_1) \\ 0 \end{bmatrix}. \tag{43}$$

The estimate of the velocity $\bar{v}$ and the yaw rate $\overline{\dot{\psi}}$ is generally made by measuring the wheel rates and via a yaw rate sensor, respectively.

The vehicle's direction is only approximated by the yaw angle. The actual direction of motion is given by the course angle v, which differs from the yaw angle ψ by the dynamic side slip angle $\beta : v = \psi - \beta$. If the side slip angle and side slip rate of the vehicle are known, then it is possible to forego the approximation of the course rate by the yaw rate. Optical measurement techniques allow the estimation of the side slip angle and the side slip rate (Scaramuzza and Fraundorfer 2011). Moreover, these parameters can be estimated with the aid of a physical vehicle model (Mayr 2001).

3.2.2 Alternatives to Using Dead Reckoning for Ego-Motion Estimation

Depending on the intended use of the grid map, various methods lend themselves to estimating ego- motion. Dead reckoning can be used to construct grid maps with high relative precision, as long as extreme driving maneuvers, which negatively affect the side slip angle estimation, do not occur. However, because errors in the ego-motion estimate are additive, the resulting grid maps exhibit a distortion relative to the actual vehicle environment which usually increases with the distance traveled. These grid maps are therefore suited primarily to the representation of the vehicle environment within a limited spatial area, e.g., a road segment of only a few 100 m.

Another application is the mapping of surroundings, for which the vehicle passes a point repeatedly, as for example in a parking lot. Distortion in the grid map can have an interfering impact when reentering a previously mapped area. Processes which additionally use high-precision GPS measurements to support on-board sensors in estimating the vehicle pose are suited to such cases. Furthermore, simultaneous localization and mapping (SLAM) algorithms provide the possibility to include the measurement results from environment detecting sensors in the ego-motion estimation. These algorithms are computationally more intense, although there are computationally efficient approximations as well (Thrun 2005).

3.3 Algorithms for Generating Occupancy Grid Maps

For many current and future driver assistance systems, the information on the drivable space in the vehicle environment is prerequisite. Drivable space here means the area where the vehicle can drive without a collision occurring. Occupancy grid maps provide the possibility to depict in detail what area of the vehicle environment is occupied by an obstacle. In doing so, classical occupancy grid maps are especially suited to depicting static or slowly changing surroundings.

3.3.1 Use of LIDAR or RADAR Data

Sensors which measure distance are best-suited for generating occupancy grid maps because they allow a direct inference on the occupancy status of individual cells. The basic concept can be explained based on the following simplified example. Figure 10 schematically illustrates the inference onto a two-dimensional occupancy grid map of a measurement by a multi-beam LIDAR sensor (see ▶ Chap 18, "Automotive LIDAR"). LIDAR beams are reflected by obstacles, allowing the occupancy of corresponding cells to be inferred (shown in black). Cells which lie between the sensor and the reflection points are assumed to be free (shown in white). This measurement does not provide information on further cells (shown in gray). Such models are called inverse sensor models which in general, however, are far more complex. An example of a probabilistic inverse sensor model is described in the following section.

Inverse Sensor Models

Sensor measurements are inherently afflicted with uncertainties. Based on a measurement, it is therefore impossible to make an unequivocal statement on the occupancy of a cell. Rather, the occupancy probability $p(o)$ is calculated for each cell. The event that the cell is occupied is described by o. The probability of the

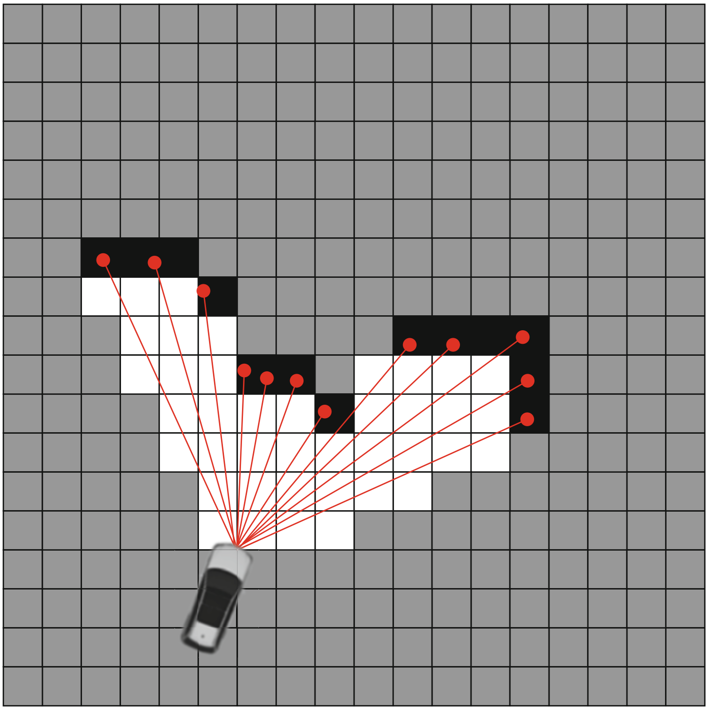


Fig. 10 Simplified occupancy grid map based on measurements by a multi-beam LIDAR sensor

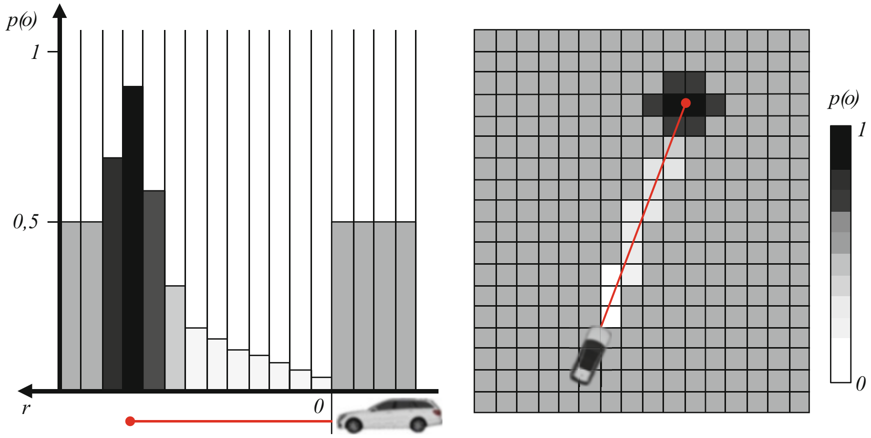


Fig. 11 Example of an inverse sensor model of a single LIDAR reflection showing the side view (*left*) and birds-eye view (*right*)

opposite event $f = \overline{o}$ (the cell is free) corresponds to the complementary probability $p(f) = 1 - p(o)$. This assumes that the occupancy state of a cell does not change over time; it is either occupied or empty.

A probabilistic inverse sensor model states which occupancy probability applies to a cell based on a sensor measurement. Figure 11 shows an example for a probabilistic inverse sensor model based on a single-beam distance measurement using LIDAR.

The left portion shows the occupancy probability $p(o)$ as a function of distance r from the sensor. The individual cells are modeled independently of one another. Cells in the area of the reflection point receive an occupancy probability above 0.5. In contrast, the occupancy probability of the cells between the sensor and the reflection point is lower than 0.5. This area is often referred to as free space and the corresponding part of the inverse sensor model as the free space function. The inverse sensor model accounts for the measurement accuracy of the sensor as well as the uncertainty of the vehicle pose estimate in the grid map. Therefore, besides the cell containing the reflection point, multiple other cells are assigned higher occupancy probabilities with $p(o) > 0.5$. In this example, the free space is assumed to be more certain the closer it lies to the sensor. This illustrates that the probability of erroneous measurements increases with distance from the sensor. Uncertainties which result from the two-dimensional modeling of the three-dimensional vehicle environment are also accounted for. For example, obstacles are only recognized if, on account of their size or the pitching of the vehicle, they are located within the LIDAR beams. This is a further reason for the decreasing safety of the free space. The modeling in the azimuth direction is performed analogously. The right hand portion of Fig. 11 shows the birds-eye view of the inverse sensor model.

In this example of an inverse sensor model each LIDAR beam is modeled individually, which is not inherently the case. This modeling impacts the mapping algorithm, which calculates the occupancy probability based on the complete LIDAR measurement consisting of many individual beams. The fundamental goal in the design of the inverse sensor model is to reconstruct the characteristics of the sensor as precisely as possible while keeping the complexity and computational effort low. Some approaches also apply machine learning algorithms to derive inverse sensor models (Thrun 2005).

Inverse sensor models for RADAR sensors differ in that it is possible for RADAR reflections to also pass through objects. Free space functions for RADAR sensors are therefore more complex. Uncertainties in the azimuthal component are generally greater for RADAR sensors than for LIDAR sensors. In addition, RADAR measurements often exist in the form of a grid map, which is referenced to the sensor and constructed in polar coordinates. On account of the different characteristics of different RADAR sensors (see ▶ Chap. 17, "Automotive RADAR"), the corresponding inverse sensor models also vary greatly. A simple sensor model can be created for which solely intensity measurements from the RADAR are entered as occupancies in the grid map and no free space function is used. However, this leads to the situation that occupancy probabilities can inherently only be increased. A forgetting factor can be implemented to counter this, as described in more detail in Sect. 3.4.

Combination Using Static Binary Bayes Filters

If no information is available yet on the occupancy of a cell, the occupancy probability is assumed to be $p(o) = 0.5$. This is the initial value for each cell. If more than one information source exists for a cell, then the information is combined. Depending on the inverse sensor model, this can occur if parts of a measurement are processed individually as with the individual LIDAR beams in the preceding example. Dependent on the application, it is often the case that multiple successive measurements are incorporated in an occupancy grid map.

Under certain assumptions, multiple measurements z_i can be combined using the static binary Bayes filter:

Given are two conditional probabilities $p(o|z_1), p(o|z_2) \epsilon (0, 1)$. Let measurements z_1, z_2 be independent. Assuming the identical a priori probability $p(o) = p(\overline{o}) = 0.5$, then the combined probability is (Thrun 2005):

$$p(o|z_1, z_2) = \frac{p(o|z_1)p(o|z_2)}{p(o|z_1)p(o|z_2) + (1 - p(o|z_1))(1 - p(o|z_2))}. \quad (44)$$

The combination rule, Eq. 44, has the following characteristics:

$$p(o|z_1, z_2) > p(o|z_1) \Leftrightarrow p(o|z_2) > 0.5, \quad (45)$$

$$p(o|z_1, z_2) = p(o|z_1) \Leftrightarrow p(o|z_2) = 0.5, \quad (46)$$

$$p(o|z_1, z_2) = p(o|z_2, z_1). \quad (47)$$

An occupancy probability $p(o|z_2) > 0.5$ increases the previous occupancy probability. An occupancy probability $p(o|z_2) = 0.5$ corresponds to a neutral element and the sequence of the combination has no effect.

Measurements from different sensors or points in time can be combined with one another using the combination rule, see Eq. 44. However, an important assumption for the practical realization of the grid map is the independence of individual cells. Dependencies between cells exist in reality and in principle can be accounted for, but they increase the computational effort required for the creation of the grid map enormously.

For practical reasons, it is common to store the occupancy probability in logarithmic form as the log-odds ratio:

$$l(o) := \log\left(\frac{p(o)}{1 - p(o)}\right). \tag{48}$$

The advantage lies therein, that in logarithmic form the combination rule, Eq. 44, can be performed as an addition:

$$l(o|z_1, z_2) = l(o|z_1) + l(o|z_2). \tag{49}$$

The same conditions apply as for Eq. 44. The reverse transformation is completed using:

$$p(o) = 1 - \frac{1}{1 + e^{l(o)}}. \tag{50}$$

In addition, the Dempster-Shafer theory of evidence (Shafer 1976; Zou et al. 2000; Nuss et al. 2013; Effertz 2008) is increasingly establishing itself in the field of grid mapping. It allows differentiation between uncertainty due to probability theory and uncertainty due to lack of information. Thus grid cells for which no information is available can be distinguished from cells whose state remains uncertain after the combination of multiple, possibly contradicting measurements.

Mapping Procedure

A sequence plan of the mapping algorithm for the presented inverse sensor model is shown in Fig. 12.

In this example a LIDAR measurement consists of multiple individual beams. If a measurement which took place at time t_i is entered, the sensor pose within the grid map at time t_i is first estimated. Subsequently, each beam z_h is incorporated in the grid map individually. The occupancy probability $p(o|z_h)$ for each affected cell is first calculated according to the inverse sensor model and then combined with the occupancy probability $p(o|z_{h-1}, \ldots, z_1)$ resulting from the previous LIDAR beams. For each cell, the occupancy probability from the preceding time step is used as the prior value.

Fig. 12 Sequence plan of a mapping algorithm

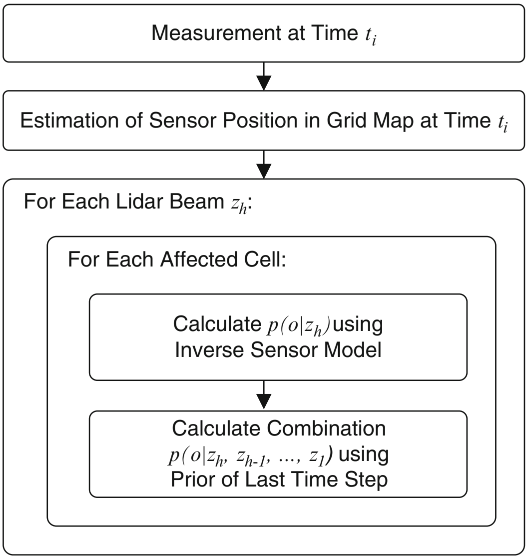


Fig. 13 Grid map of an actual vehicle environment

An occupancy grid map created on the basis of a similar inverse sensor model is shown in Fig. 13. The sensors used were multiple-beam LIDAR from Ibeo (Ibeo LUX3, 4 scan layers).

Practical Aspects of Occupancy Grid Map Generation

A fundamental problem in the generation of occupancy grid maps is the fact that the vehicle environment changes over time. This is addressed separately in Sect. 3.4.

Two-dimensional grid maps and two-dimensional inverse sensor models only describe a portion of the three-dimensional vehicle environment. This can result in contradictions, for example when the beams of a LIDAR sensor run underneath a traffic sign or a bridge at one point in time and then detect the object and are reflected by it at the subsequent point in time. The elevation profile of the surroundings can also influence the grid map.

The example for the inverse sensor model did not address how a beam which does not produce reflections is modeled. In such cases, inverse sensor models often assume free space for the cells crossed by the beam up to a certain maximum distance. However, the free space probability for such cells is lower than for those which lie between a reflection and the sensor because, especially in street environments, LIDAR beams can strike objects without a reflection being measured. This can happen with black, matte surfaces or on smooth surfaces such as guardrails which are struck at a sharp angle (Thrun 2005).

3.3.2 Use of Camera Data

Camera data can also be used to generate grid maps. Stereoscopic cameras provide closely-spaced distance measurements in the near detection range of the sensor. Thus these sensors are also suited for generating three-dimensional grid maps.

Monoscopic cameras do not provide any three-dimensional information. Each pixel of an image is assigned a beam in three-dimensional space according to a pinhole camera model. Under certain assumptions, however, it is possible to deduce a transformation between an image pixel and the coordinate of a two-dimensional grid map. The assumptions are that the driving surface on which the vehicle is located is flat, that the camera height above the surface is constant, and that the vehicle is not making any pitching or rolling movements. These assumptions taken together are the flat world assumptions. In order to calculate the transformation, the three-dimensional pose of the camera must be known in the grid map coordinates which have been augmented by the third spatial axis. The pose results from the vehicle pose and the extrinsic camera calibration. The intrinsic camera calibration then finally allows the transformation from a three-dimensional space relative to the camera to a pixel in the camera image (see ▶ Chap. 20, "Fundamentals of Machine Vision"). Altogether this process, called inverse perspective mapping (IPM), allows each point on the two-dimensional road surface to be assigned to an image pixel, as long as the former lies in the detection range of the camera.

IPM enables camera data to be introduced in grid maps. Examples are gray tones or objects classified in the video image. Figure 14 shows an example of IPM being applied to generate a grid map (Konrad et al. 2012). The occupancy probability of a cell here corresponds to the probability of the presence of road markings. A classification algorithm detects road markings, which are introduced into the grid map according to a suitable inverse sensor model.

3.4 Dealing with Moving Objects

A prerequisite for the application of the static binary Bayes filter is the assumption that a cell does not change its occupancy state; it remains either occupied or free. For the combination of individual LIDAR beams which were registered nearly

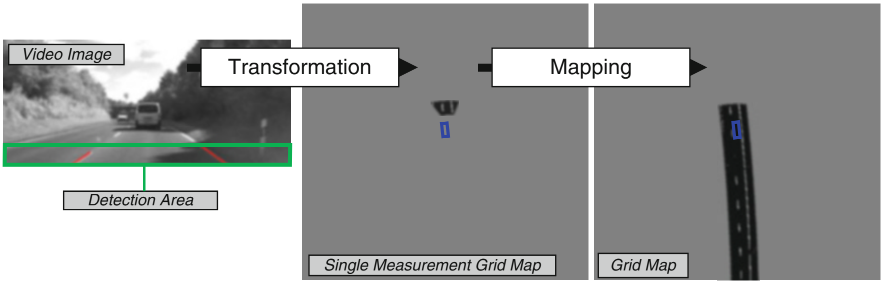


Fig. 14 Grid map representing road markings

simultaneously, this assumption holds approximately true since a cell does not change its occupancy state in such a short time interval.

In general, and with regard to longer time intervals, this assumption for the vehicle environment does not apply; it is violated by the presence of dynamic objects. In the literature, different approaches to deal with dynamic objects in grid maps are described. Essentially it is possible to differentiate between approaches in which dynamic objects should simply be filtered, and approaches in which the grid map is used to detect dynamic objects and thus discern them from the static surroundings.

A simple method of the former type is the introduction of a forgetting factor. In this case, the inference of the occupancy probability of a cell depends not only on spatial, but also on temporal conditions. The inference of the occupancy probability is therefore less certain the older the measurement. More recent measurements are weighted more strongly than older measurements when estimating the occupancy probability.

In many applications, grid maps are used explicitly to detect dynamic objects. This is generally done by determining the temporal consistency of the occupancy probability of cells. Cells which are consistently occupied or free are assigned to the static environment and cells for which the occupancy probability fluctuates greatly are marked as dynamically occupied areas. Especially occupancies occurring in areas previously detected as free are often treated as detections of dynamic objects. This often serves as a preprocessing step for tracking algorithms (Nuss et al. 2012, 2013; Petrovskaya et al. 2012). More elaborate system architectures allow for a stronger dependency between grid maps and tracking algorithms, as discussed in greater detail in Sect. 4.

3.5 Efficient Memory Management

The amount of data per area of the environment is fixed by the grid resolution and the amount of data per cell. The arrangement in cells additionally allows direct access to all information on a location through its coordinates. If in comparison measurement data were stored as raw data, the evaluation of a location would require searching for all measurement data which allow an inference to be made on the sought-for characteristic of that location.

In order to be able to access a grid cell especially quickly using its coordinates, it makes sense to store the grid map in memory as an array. A cell can then be accessed directly by means of its index. However, a disadvantage to this approach is that depending on the size of the map, memory will also be reserved for cells that will never lie within the detection range and thus will never contain information. One attempt at a solution for this problem is to divide the global grid map into tiles of which only those that are in the immediate vicinity of the vehicle are loaded into internal memory (Konrad et al. 2011).

Another approach makes use of octrees (Schmid et al. 2010). That data structure allows different cell sizes to be used for different areas. This makes it possible to have a greater degree of detail in important areas (near the vehicle) on account of a smaller cell size, whereas memory can be saved in other areas through larger cells.

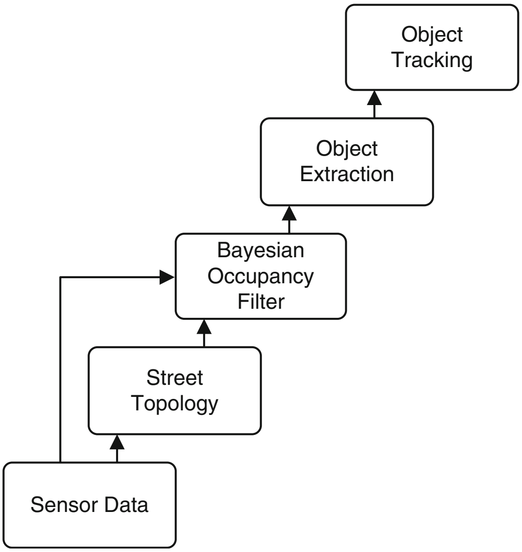
4 Architectures and Hybrid Representation Forms

Object-based representations and grid maps are depictions of the vehicle environment which always contain only a portion of the complete environment. Object-based representations depict, among other things, the position and velocity of individual objects and are based on process models which permit predictions for the near future. The representation of the vehicle environment therefore extends over a period of time. In contrast, the generation of classic occupancy grid maps assumes a stationary environment. Dynamic objects violate this assumption. Occupancy grid maps fuse and store information relative to location and not to objects and represent the environment in the complete detection range of the sensors. The road topology can be seen as the third component of the vehicle environment. It is however a static component of the environment. Depending on the application, road topology is detected while driving. This is possible especially in very structured environments such as highways. More complex road topologies which cannot be recognized online are often stored in digital maps, making the road topology available referenced to the vehicle. On account of their different characteristics, it is an interesting question how these different domains for representing the environment could be advantageously combined. This section will address a selection of approaches.

4.1 Bayesian Occupancy Filter (BOF)

The Bayesian occupancy filter divides the environment into individual cells, but rejects the assumption of a stationary environment (Coué et al. 2006). Rather, a four-dimensional space is applied consisting of two spatial coordinates and two coordinates for velocity along the spatial coordinates, respectively. A cell is therefore assigned a two-dimensional location and a two-dimensional velocity

Fig. 15 Bayesian occupancy filter using prior map knowledge



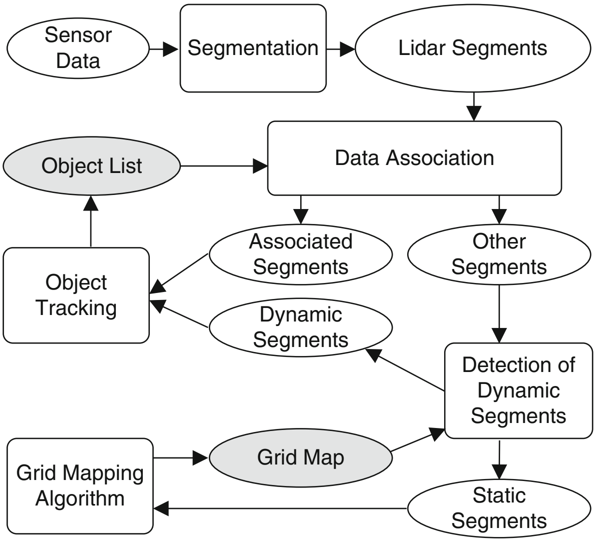
vector. This allows a Bayes filter consisting of prediction and innovation to be applied. The environment representation is also free of object assumptions, however, a Markov process which represents a state transition must be assumed for the prediction. This requires process assumptions such as a constant velocity. The degree of abstraction of the Bayesian occupancy filter is therefore higher than for classic occupancy grid maps and the computational effort is also greater. However, in return, dynamic events in the environment are also considered and recorded.

Enhancements exist in which highly dynamic spatial areas are grouped together as objects and are tracked with classic object tracking approaches (Baig et al. 2012). Additional enhancements account for road topology in order to influence the prediction step (Gindele et al. 2009). This should depict the fact that road topology influences the direction of motion of the traffic participants. The complete architecture of this approach for environment detection is shown in Fig. 15.

4.2 Simultaneous Localization and Mapping and Moving Object Tracking (SLAMMOT)

The SLAMMOT algorithm addresses the problem of simultaneously using LIDAR sensor data to estimate the ego-motion, generate a grid map, and track dynamic objects (Wang et al. 2007). The corresponding system architecture is shown in Fig. 16. The concept of this approach is to use the grid map to detect dynamic objects. LIDAR segments of dynamic objects are not entered into the grid map but are used as object detection for object tracking. The procedure of the algorithm is as follows: First the individual reflection points of the LIDAR measurement are segmented. The segmentation is carried out based on the spatial point density. In

Fig. 16 Simultaneous localization and mapping and moving object tracking: procedure



an association step, LIDAR segments are assigned to objects already being tracked. Segments which cannot be assigned are examined in the grid map to determine whether the corresponding reflections are from dynamic or stationary objects. Reflections from static objects are entered into the grid map while reflections from dynamic objects are used to create new objects for object tracking. In addition, reflections from static objects are used in order to estimate the ego-motion of the vehicle, but this is not addressed in any further detail here.

A related approach in which the object tracking and the grid map are even more dependent on one another is introduced in Bouzouraa and Hofmann (2010). Here, the reflections from dynamic objects are also entered in the grid map. In addition, a list of associated cells for every tracked object is stored in memory. This allows the motion of corresponding cells to be estimated based on their associated objects, which allows a prediction of the grid map for the future.

4.3 Hierarchical Modular Environment Perception (HMEP)

As mentioned at the beginning of this chapter, economic and practical aspects place demand on the architecture of environment perception. Especially modularity and generic interfaces between individual modules will be of increasing importance. This stands in contrast to a strong dependence between object-based environment models, grid maps, and the road topology. Hierarchical modular environment perception (Nuss et al. 2014) as shown in Fig. 17 provides a compromise.

This example shows a breakdown into four modules: the estimation of the ego-motion, the grid mapping, the generation of the road topology, and the object tracking. These modules have a hierarchical relationship. The results from these

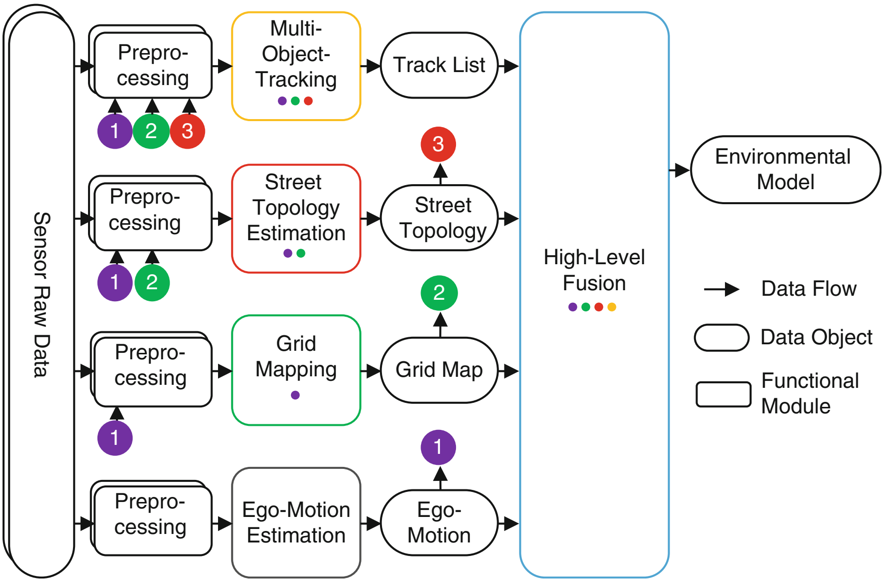


Fig. 17 Hierarchical modular environment perception (*HMEP*)

four modules comply with a specific standard, so that they are compatible. A superordinate module has access to every result from a subordinate module. The sequence of modules provides a model to which many of the architectures found in the literature adhere to. For example, all architectures introduced so far and many more are based on having a grid map as the basis for object tracking, usually to recognize moving objects. Likewise, many localization algorithms are also based on grid maps (Deusch et al. 2013). The hierarchy is reasonable in that it follows the degree of abstraction of the modules. Feedback is ruled out, which results in several limitations but is a prerequisite for the interchangeability of the modules. Finally, a consistent environment model which as a matter of consequence also complies with a specific standard (Nuss et al. 2014) is created from the individual modules. Standard architectures for a modular vehicle environment representation do not yet exist and are the subject of current research and development.

5 Conclusion

The fundamentals of object-based and grid-based representations of vehicle environments as well as the necessary basic algorithms they require are described within the framework of this chapter. Whereas every relevant object is described by its own dynamic state space model in object-based representations, the grid map representations are initially model free, i.e., they don't require any physical model hypotheses. They are therefore well suited for the representation of the static portion of the environment representation including freely drivable space or driving space limitations. In the construction of such grid maps, moving objects tend to behave more as interference which needs to be calculated out and must therefore be dealt with separately. Various approaches therefore recommend the enhancement of grid maps to include dynamic components, although no method for doing so has yet established itself. Grid map representations, however, have the additional advantage that sensor measurements – with the exception of the decision on whether they originate from dynamic or static objects – do not need to be evaluated and associated beforehand. In addition, the methodology allows the implicit fusion of multiple sensors through simple aggregation of the individual measurements in the cells of the grid map.

Object-based processes have the advantage that the individual dynamic modeling of each object makes it relatively simple to determine its semantic meaning. This chapter though did not elaborate on the classification algorithms necessary to do so. The detection of each object's state requires suitable filter algorithms which presently nearly without exception are based on approximations of the recursive Bayes filter. The Kalman filter is the best-known representative of these approximations.

Since in traffic scenarios there is generally more than a single object which needs to be dynamically detected and tracked, multi-instance approaches are applied as a solution, i.e., each single-model object is described by a single filter, for example, a Kalman filter. However, this approach requires the error-free assignment of the sensor measurements to the objects, termed the data association. In order to avoid data association errors which cannot be undone after filtering, probabilistic data association processes such as PDA and JPDA exist which apply probabilities to all possible associations and weight them before finally introducing them into the filter. Besides the dynamic object state itself, another important parameter for safety-relevant driver assistance systems is its existence probability, a measure of confidence that an object appearing in the vehicle environment representation really originated from an actual existing object. The JIPDA algorithm, for example, provides a probabilistic data association with integrated existence estimation.

Current work is also addressing approaches to integrate multi-object tracking in an algorithm. These approaches, which are based on the random finite set (RFS) theory, not only avoid data association errors prior to the filtering, but also allow probabilistic modeling of the dependency of the motion behavior of all detected objects in relation to each other, which of course is inherent between traffic participants. The relatively high computational effort of these processes, however, currently prevents their implementation being close to production.

Since both object-based and grid-based representations have specific advantages, the combination of both methods in hybrid architectures suggests itself. On account of the growing requirements and different functional variations, a modular architecture with as-generic-as-possible interfaces to the sensors and the situational evaluation or functions will become increasingly important.

References

Baig Q, Perrollaz M, Nascimento J, Laugier C (2012) Using fast classification of static and dynamic environment for improving Bayesian occupancy filter (BOF) and tracking. In: International conference on control automation robotics vision (ICARCV), Guangzhou, China, pp 656–661

Bar-Shalom Y, Fortmann T (1988) Tracking and data association. Academic, Boston

Bar-Shalom Y, Tse E (1975) Tracking in a cluttered environment with probabilistic data association. Automatica 11:451–460

Bouzouraa M, Hofmann U (2010) Fusion of occupancy grid mapping and model based object tracking for driver assistance systems using laser and radar sensors. In: IEEE intelligent vehicles symposium (IV), San Diego, CA, USA, pp 294–300

Coué C, Pradalier C, Laugier C, Fraichard T, Bessiere P (2006) Bayesian occupancy filtering for multitarget tracking: an automotive application. Int J Robot Res 25(1):19–30

Deusch H, Nuss D, Konrad P, Konrad M, Fritzsche M, Dietmayer K (2013) Improving localization in digital maps with grid maps. In: IEEE conference on intelligent transportation systems (ITSC), The Hague, Netherlands, pp1522–1527

Effertz J (2008) Sensor architecture and data fusion for robotic perception in urban environments at the 2007 DARPA urban challenge. In: Conference on robot vision (RobVis). Springer, Berlin/Heidelberg, pp 275–290

Elfes A (1989) Using occupancy grids for mobile robot perception and navigation. IEEE Comput 22(6):46–57

Gindele T, Brechtel S, Schroder J, Dillmann R (2009) Bayesian occupancy grid filter for dynamic environments using prior map knowledge. In: IEEE intelligent vehicles symposium (IV), Xi'an, China, pp 669–676

Julier S, Uhlmann J, Durrant-Whyte H (2000) A new method for the nonlinear transformation of means and covariances in filters and estimators. IEEE Trans Autom Control 45(3):477–482

Kalman R (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:35–45

Konrad M, Szczot M, Schüle F, Dietmayer K (2011) Generic grid mapping for road course estimation. In: IEEE intelligent vehicles symposium (IV), Baden-Baden, Germany, pp 851–856

Konrad M, Nuss D, Dietmayer K (2012) Localization in digital maps for road course estimation using grid maps. In: IEEE intelligent vehicles symposium (IV), Alcalá de Henares, Spain, pp 87–92

Levinson J, Thrun S (2010) Robust vehicle localization in urban environments using probabilistic maps. In: IEEE international conference on robotics and automation (ICRA), Anchorage, Alaska, pp 4372–4378

Mahler R (2003) Multitarget Bayes filtering via first-order multitarget moments. IEEE Trans Aerosp Electron Syst 39(4):1152–1178

Mahler R (2007) Statistical multisource-multitarget information fusion. Artech House, Boston

Mattern N, Schubert R, Wanielik G (2010) High-accurate vehicle localization using digital maps and coherency images. In: IEEE intelligent vehicles symposium (IV), San Diego, CA, USA, pp 462–469

Mayr R (2001) Regelungsstrategien für die automatische Fahrzeugführung (Control strategies for automated vehicles). Springer, Berlin/Heidelberg

Munz M (2011) Generisches Sensorfusionsframework zur gleichzeitigen Zustands- und Existenzschätzung für die Fahrzeugumfelderfassung (Generic sensor fusion framework for simultaneous estimation of state and existence uncertainty in vehicle environment perception). PhD thesis, Ulm University

Musicki D, Evans R (2004) Joint integrated probabilistic data association: JIPDA. IEEE Trans Aerosp Electron Syst 40(3):1093–1099

Nuss D, Reuter S, Konrad M, Munz M, Dietmayer K (2012) Using grid maps to reduce the number of false positive measurements in advanced driver assistance systems. In: IEEE conference on intelligent transportation systems (ITSC), Anchorage, Alaska, pp1509–1514

Nuss D, Wilking B, Wiest J, Deusch H, Reuter S, Dietmayer K (2013) Decision-free true positive estimation with grid maps for multi-object tracking. In: IEEE conference on intelligent transportation systems (ITSC), The Hague, Netherlands, pp 28–34

Nuss D, Stuebler M, Dietmayer K (2014) Consistent environmental modeling by use of occupancy grid maps, digital road maps, and multi-object tracking. In: IEEE intelligent vehicles symposium (IV), Dearborn, Michigan, USA

Petrovskaya A, Perrollaz M, Oliveira L, Spinello L, Triebel R, Makris A, Yoder J, Laugier C, Nunes U, Bessiere P (2012) Awareness of road scene participants for autonomous driving. In: Eskandarian A (ed) Handbook of intelligent vehicles. Springer, London, pp 1383–1432

Reuter S, Wilking B, Wiest J, Munz M, Dietmayer K (2013a) Real-time multi-object tracking using random finite sets. IEEE Trans Aerosp Electron Syst 49(4):2666–2678

Reuter S, Meissner D, Wilking B, Dietmayer K (2013b) Cardinality balanced multi-target multi-Bernoulli filtering using adaptive birth distributions. In: Proceedings of the 16th international conference on information fusion, Istanbul, Turkey, pp1608–1615

Reuter S, Vo B-T, Vo B-N, Dietmayer K (2014) The labeled multi-Bernoulli filter. IEEE Trans Signal Process 62:3246–3260

Scaramuzza D, Fraundorfer F (2011) Visual odometry: part I – the first 30 years and fundamentals. IEEE Robot Autom Mag 18:4

Schindler A (2013) Vehicle self-localization with high-precision digital maps. In: Intelligent vehicles symposium workshops (IV Workshops), Gold Coast, Australia, pp 134–139

Schmid M, Mählisch M, Dickmann J, Wünsche H (2010) Dynamic level of detail 3d occupancy grids for automotive use. In: IEEE intelligent vehicles symposium (IV), San Diego, CA, USA, pp 269–274

Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton

Skog I, Handel P (2009) In-car positioning and navigation technologies: a survey. IEEE Trans Intell Transp Syst 10(1):4–21

Stiller C, Puente León F, Kruse M (2011) Information fusion for automotive applications – an overview. Inf Fusion 12(4):244–252

Thrun S (2005) Probabilistic robotics. The MIT Press, Cambridge, MA

Vo B-T, Vo B-N (2013) Labeled random finite sets and multi-object conjugate priors. IEEE Trans Signal Process 61:3460–3475

Vo B-T, Vo B-N, Cantoni A (2009) The cardinality balanced multi-target multi-Bernoulli filter and its implementations. IEEE Trans Signal Process 57(2):409–423

Wang C, Thorpe C, Thrun S, Hebert M, Durrant-Whyte H (2007) Simultaneous localization, mapping and moving object tracking. Int J Robot Res 26(9):889–916

Zou Y, Yeong K, Chin S, Xiao W (2000) Multi-ultrasonic sensor fusion for autonomous mobile robots. In: Sensor fusion: architectures, algorithms and applications IV; proceedings of SPIE 4051, Orlando, FL, USA, pp 314–321

Data Fusion for Precise Localization

25

Nico Steinhardt and Stefan Leinen

Contents

N. Steinhardt (✉)
Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: nico@steinhardt.com.de

S. Leinen
Institut für Geodäsie, Fachgebiet Physikalische Geodäsie und Satellitengeodäsie, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: leinen@psg.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_26

Abstract

In current vehicles, redundant sensors with heterogeneous measurement principles are applied in increasing numbers. Taking the advantage of these already existing redundancies, the concept of a central virtual sensor for the estimation of kinematic vehicle properties is created, based on a set of close-to-series sensors, consisting of a MEMS inertial measurement unit, a GPS receiver, and odometry sensors. Furthermore, a real-time capable implementation of this architecture is realized, using a linearized Error-State-Space Kalman filter. This fusion filter is enhanced by a correction algorithm for measurement latencies of multiple sensors and a two-step plausibilization of raw measurement data. In addition, an integrated assessment of the data quality is implemented. It describes data consistency using an integrity measure and data accuracy with a virtual datasheet.

An increasing number of heterogeneous and often redundant sensors are being used in motor vehicles. In this chapter, which is essentially based on Steinhardt (2014), we outline the methodology for fusing heterogeneous sensor data for accurate localization and additionally for estimating vehicle dynamics. The aim of this fusion is to produce a consistent dataset with improved accuracy. We introduce classifications and ontologies for potential system architectures and fusion filters and derive special filter enhancements for use with heterogeneous sensors similar to those used in series production. This results in the concept of a virtual sensor as a new layer between sensors and applications: We also derive the associated requirements to describe data quality, especially for use in safety-critical applications. This includes descriptors for both integrity and accuracy; an example application is presented for a given set of sensors. We also consider the special case of a vehicle on a moving surface (e.g., a ferry) and propose potential approaches for solving the problems associated with this. Furthermore, the results of the applied fusion filter are presented and discussed. Our analysis concludes with a look at potential enhancements and the integration of additional sensors.

1 Requirements of Data Fusion

Nowadays an increasing number of driver assistance systems, which rely on sensor measurements for their function, are being used in all classes of vehicle. Generally these involve the installation of several sensors, which, despite of very different measuring principles and characteristics, supply redundant information. This redundancy, which is mostly utilized via models, is to some extent used locally – i.e., in the control units for specific applications. This always takes place with specific reference to the device and associated software, e.g., for self-diagnostics. Especially in safety-critical systems, a lot of effort goes into detecting measurement errors and sensor defects. Because of growing user demand for cooperative, complex functions, it is becoming necessary to jointly evaluate the existing sensor data from different application functions in a networked environment. However, because so far data have only been used and application-specifically processed, inconsistencies are to be expected in the data.

As far as hardware is concerned, because of the improved performance of processors and bus systems, an environment already exists for centralized processing that takes account of the heterogeneous characteristics of all installed sensors. Building on this, centralized evaluation of signal quality offers the potential to relieve the applications of a considerable amount of fault recognition. Moreover, the cross-checking of redundant measurement data offers an additional opportunity for detecting errors and faults, over and above the existing principles. The aim is therefore to construct an open, extendable architecture for central sensor data fusion and signal quality description. The essential requirements of such a central, *virtual sensor* are:

- Real-time capability and causal filter behavior, i.e., processing data in their actual chronological order
- Short, known, and constant as possible latency and group delay
- Robustness as well as detection and compensation of faults
- Improved accuracy and availability or an optimum compromise between the two
- Generation and output of consistent fusion data for all applications
- Evaluation of the integrity and output of quality descriptors

A basic principle for the fusion of measured data is the existence of redundancy (Pourret et al. 2008) by measuring the same physical values via multichannel measurement of the same quantity (parallel redundancy) or conversion of another measured quantity into the one required (analytical redundancy). By their very nature, heterogeneous measuring principles require differences in the measured data to be considered by the fusion filter:

- Synchronous or asynchronous measurement compared to other signal sources
- Different resolutions
- Different, possibly variable, sampling rates
- Latency times between measurement and data availability
- Time-variant availability of information sources
- Dependency upon environmental conditions
- Dynamically varying measurement accuracy during operation

Regardless of their cause, sensor errors (Niebuhr and Lindner 2002) can be divided into systematic components, quasi-stationary components that are constant over several measurements such as an offset or a scale factor error, and stochastic components that are random from measurement to measurement, for example, signal noise. While, in principle, the random components cannot be corrected deterministically, known systematic errors can be corrected on the basis of a model and quasi-stationary errors by measurements, provided that they are observable. If uncorrectable but detectable errors occur, it is at least possible to prevent any negative impact on the fusion result. The virtual sensor must therefore be robust against random disturbances and detect and compensate for deterministic errors. Likewise, temporal errors in measured data must be corrected and temporary sensor faults or unavailability must be bridged.

Once redundant data have been processed to give a consistent fusion result, errors in individual measurements can no longer be clearly identified or attributed to a cause. However, because of the self-diagnostic capability of signal sources that is usually available and, in particular, model-based plausibility control methods within data fusion, the validation against redundant signals is possible. This means that error detection is improved relative to the current decentralized use of measured data. Consequently, it is necessary to run a validation and transmit its results to user functions.

Since unrestricted ongoing use of existing functions is required, the output fusion result must guarantee that the necessary data rate and resolution of the most demanding function in the system are fulfilled or that a compromise is found in case of conflicting requirements.

2 Basic Principles

Below we briefly outline the various coordinate systems that are required to model the sensor data, to fuse them, and to provide accurate localization and for appropriate output of the fusion results for the different applications. The “Coordinate Systems” section is based on the convention used in Wendel (2007). This is followed by a description of the main localization sensors and their characteristics.

2.1 Coordinate Systems

All the following systems are three-dimensional, right-handed Cartesian coordinate systems.

Inertial coordinate system. Vectors in this coordinate system are identified by the index $_i$. This celestial Cartesian coordinate system is defined such that the origin is in the center of the Earth and the z-axis runs along the Earth’s rotation axis. The x- and y-axes are perpendicular to each other in the equatorial plane and are aligned using fixed stars. An inertial measuring unit measures accelerations and angular rates relative to this inertial system.

Terrestrial coordinate system according to WGS84. Vectors in this coordinate system are identified by the index $_e$. This Earth-fixed coordinate system rotates with the Earth and is defined such that the origin is in the center of the Earth and the z-axis lies along the Earth’s rotation axis. The x-axis is the line intersecting the equator and the prime meridian (Greenwich), the y-axis completes the x- and z-axes to a right-hand Cartesian system. Positions are often given as ellipsoidal coordinates instead of Cartesian coordinates: latitude and longitude in the unit rad and height above the Earth ellipsoid in the unit m (positive up).

Navigation coordinate system. Vectors in this coordinate system are identified by the index $_n$. The origin of this Cartesian coordinate system is in the vehicle and usually coincides with the origin of the vehicle-fixed coordinate system. In contrast, the axes are aligned Earth fixed and point toward East and North and upward along the gravity vector.

Vehicle-fixed coordinate system in accordance with DIN 70000 (Deutsches Institut für Normung 1994), also known as the body coordinate system. Vectors in this coordinate system are identified by the index $_b$. The system is oriented such that the x-axis points in vehicle forward direction, the y-axis is perpendicular to it to the left, and the z-axis points upward perpendicular to the x-y-plane. The origin must be explicitly defined in relation to the vehicle. If the origin is selected as in the navigation coordinate system, then the transformation of vectors between the two systems is described solely by a rotation around the attitude angles yaw Ψ, pitch Θ, and roll φ.

Wheel coordinate system. Vectors in this coordinate system are identified by the index $_w$. The origin is in the center of the wheel; the plane of the x- and y-axis is oriented parallel to the x-y-plane of the body coordinate system and rotated through the wheel steer angle δ_L around the z-axis. The z-axis is oriented parallel to the z-axis of the vehicle-fixed system. For simplicity it is assumed that camber angle, caster angle, and spread angle are negligible, that δ_L already takes account of the toe angle, and that, when the steer angle is changed, rotation only takes place around the z-axis.

2.2 Localization Sensors and Their Characteristics

The main sensors used for accurately localizing vehicles are inertial measurement units (IMUs); receivers for global satellite navigation systems (GNSS), the most common of these being the Global Positioning System (GPS); and wheel sensors for odometry – that is, calculating the distance traveled by counting wheel revolutions.

An IMU records the absolute, three-dimensional values of *acceleration* and *angular rate* of the vehicle in the inertial coordinate system – inertial sensors are therefore accelerometers and gyroscopes. However, localization and navigation applications usually use the terrestrial coordinate system. A strapdown algorithm (Titterton and Weston 2004) is used to continuously compute attitude, velocity, and position of the vehicle from the inertial measurements. Wendel (2007) defines the strapdown algorithm as (...) *a calculation rule that specifies how the navigation solution for the current time step is computed from the navigation solution of the previous time step using measured accelerations and angular rates. The strapdown computation can be roughly divided into three steps*: *propagation of the attitude by integration of the angular rates*, *propagation of velocity by integration of the accelerations and propagation of the position by integration of velocity* (*translated from German*).

The strapdown algorithm is a recursive and therefore real-time capable computation method. Since an IMU measures motion quantities relative to an inertial system, the strapdown algorithm compensates for the acceleration due to gravity, Earth's rotation, Coriolis acceleration, and transport rate (angular rate of the vehicle due to movement along the Earth's ellipsoid surface), which act as disturbance variables in determining motion in Earth-fixed and vehicle-fixed coordinate systems.

Typical errors of inertial sensors are measurement noise, offsets, and scale factor errors: Often these errors vary over time and depend on external impacts, such as temperature. One advantageous feature of IMUs is that they are independent of external disturbances and therefore, as long as there are no defects, permanently available.

GNSS receivers (e.g., GPS) measure the distances between the phase centers of the satellite and receiver antennae by measuring the signal travel time. Since these distance measurements still contain significant errors, in particular those of the clocks involved, they are referred to as *pseudoranges*. Moreover, the relative velocity in the line of sight to the satellite can be calculated by time differentiation of the *pseudoranges*; these measurements are referred to as *deltaranges*. Accuracy of the velocity estimate can be improved relative to calculation from signal travel time measurements by using *carrier phase measurements*.

The absolute position is calculated from pseudorange measurements in Earth-fixed coordinates. Errors due to the Earth's rotation during the signal travel time are compensated based on a model. The velocity in Earth-fixed coordinates is calculated from deltarange measurements. Typical errors in pseudorange measurements are receiver clock errors, satellite clock errors, ephemeris errors (deviation of the real satellite orbit from the calculated orbit), and ionospheric and tropospheric errors. The receiver clock error is estimated and compensated for in the position calculation. Satellite clock and ephemeris errors are estimated by the system operator and corresponding correction parameters are sent in the navigation message. Ionospheric and tropospheric errors can be partially corrected based on a model. If a dual-frequency receiver is used, it is possible to eliminate ionospheric errors from the measurements. Since these error values only change slowly, compared to the typical sampling rate of a GNSS receiver, they can be neglected for computing the deltaranges (time differences). For these measured quantities, the only significant error is the drift of the receiver clock error resulting from the frequency error of the receiver oscillator.

Compared to IMU navigation, GNSS positioning is characterized by its long-term stable absolute accuracy, as there is no growing sensor error over time. On the other hand, errors caused by environmental conditions, which affect all measurements of a GNSS receiver, are problematic: They are caused by diffraction and reflection of the electromagnetic waves on the Earth's surface, mountains, or buildings and result in multipath reception and consequently in measurement errors. Furthermore, signal reception can be partially or completely blocked – for example, when driving through a tunnel. The availability of GNSS measurements therefore depends upon the environment and is often reduced, even if the receiver is itself operating perfectly.

Odometry measurements (Reif 2007) are based on wheel ticks (wheel rotation impulse measurements) by active or passive magnetic field sensors: These measure the angular rate of a wheel described around its *y*-axis in wheel-fixed coordinates. The steer angles required for the transformation into vehicle-fixed coordinates can either be modeled, for example, as constant zero on the un-steered rear axle, or calculated on the basis of a model by measuring the wheel steer angle, often also by magnetic field sensors. If the wheel roll radius is known or can be estimated, the speed of the vehicle body at the respective tire contact patches – related to the surface under the vehicle – can be estimated by odometry measurements. Various models of different complexities, such as single-track and two-track models, can be used to estimate vehicle motion, especially horizontal speed, yaw, and sideslip rate; based on this, cf. Bevley and Cobb (2010). Various models are also available for modeling tire effects, for example, a linear tire model or the "magic formula" model (Pacejka and Bakker 1993). Typical errors in odometry measurements are angle noise and angle quantization noise caused by the finite number of ticks per revolution. Angle noise is caused by random variations in the angle increment from which a new wheel tick is counted. There is also environmental influence caused by high slip, for example, due to a low friction coefficient, high accelerations, or driving over potholes.

3 Classification and Ontologies for Sensor Data Fusion Filters

As also described in ▶ Chap. 23, "Data Fusion of Environment-Perception Sensors for ADAS," the term sensor data fusion is normally used in two different contexts relating to motor vehicle applications:

- Merging of measurements with a (largely) different coverage, with the aim of producing a dataset that combines all coverage areas (complementary sensors)
- Merging of measurements with a (largely) similar coverage, with the aim of improving measurement quality (redundant sensors)

In this chapter, the term "fusion" is used in the sense of the second context, i.e., for the combined processing of measurements from redundant sensors (with the same coverage). However, as described in Sect. 2.2, the sensors have heterogeneous properties, and this therefore helps to improve the quality of the fusion result. Below we outline the essential classes of filters for fusing redundant data.

We then provide an overview of fusion filters, concepts, and essential characteristics. These are divided according to the ontology of coupling of the filter with the measurements, the estimation methods, and standard filter types.

3.1 Classification of Coupling of Sensors to the Filter

The sensor measurements can be integrated into a fusion filter at different levels (Groves 2008), a distinction essentially being made between the following variants:

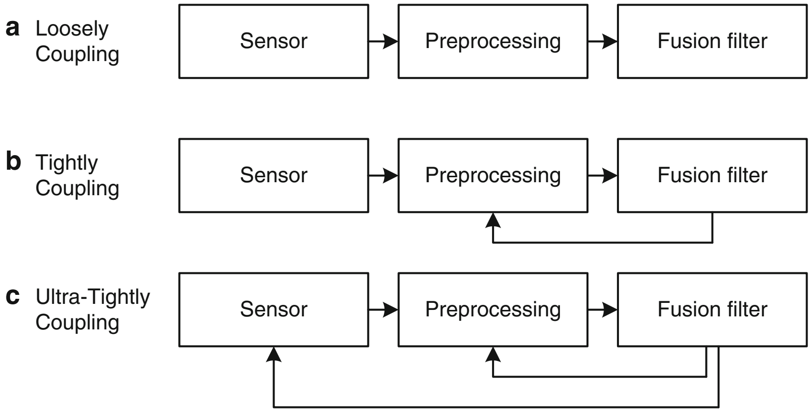


Fig. 1 Different modes of coupling sensors to the filter

Loose Coupling

As illustrated in Fig. 1a, the redundant data are preprocessed by the sensor and then fed into the filter in a purely feed-forward architecture. This structure is simple and intuitive and requires little computation effort. However, the sensor and its measurement errors are modeled as a black box, so that the quality of the fused data is directly dependent on the quality of the sensor and the associated model. Using the example of a GNSS, the fusion filter is supplied with preprocessed position estimates rather than raw data: fusion is therefore dependent upon the receiver's accuracy in estimating position, and at least four (Hofmann-Wellenhof et al. 2003) satellites have to be received simultaneously for the system to become operational.

Tightly Coupling

As illustrated in Fig. 1b, the redundant sensor data are fed into the filter as raw data, so that the preprocessing model is coupled to the fusion filter. This structure is more complex and is designed as a feedback architecture with feedback of correction data from the fusion filter into data preprocessing. This means that the quality of the measurements is solely dependent upon the sensors, and the quality of the solution estimated from these measurements is primarily dependent upon the fusion algorithm. Using the example of GNSS, the processing of raw data (pseudoranges, deltaranges, possibly carrier phases) and integrated error estimation allows that, at least for a limited period, position estimation may be continued with less than four satellites using this architecture.

Ultra-tight Coupling/Deep Integration

The structure is based on that of Tightly Coupling, except that correction and control data are fed back into the sensors, as illustrated in Fig. 1c. The measurement process itself is influenced, as measurement parameters are adjusted by the filter result. This requires both modeling of the internal sensor data processing in the filter and also availability of corresponding sensor interfaces. Consequently, much more hardware and computation time are needed than with the other methods. The quality of the measurements and of the solution depends upon the measurement hardware and the algorithm as well. So, using the example of GNSS, it is also possible to control the bandwidth of the receiving filter and the control loops for signal acquisition (tracking loops) inside the receiver.

3.2 Classification of Filter States

There are many different methods of fusing the measured data from different sensors to create a consistent, uniform dataset. Below we outline two main architectures w.r.t. the estimated quantities, i.e., states:

Full-State Filter

In filter fusion with estimation of full-state variables (e.g., position, speed, etc.), measured data are fed into the filter from all sensors in the same way and independently of each other, as illustrated in Fig. 2a. The full-state variables of the fusion are computed and output by the filter itself. The advantages of this method are the intuitive structure of the filter and the fact that the filter continues to be reliable even if an individual sensor should fail, as long as enough redundant measurements are still available. The shortcomings of this concept are the characteristics of the output data, which are directly dependent upon the dynamics of the input data and variable over time, and estimation errors

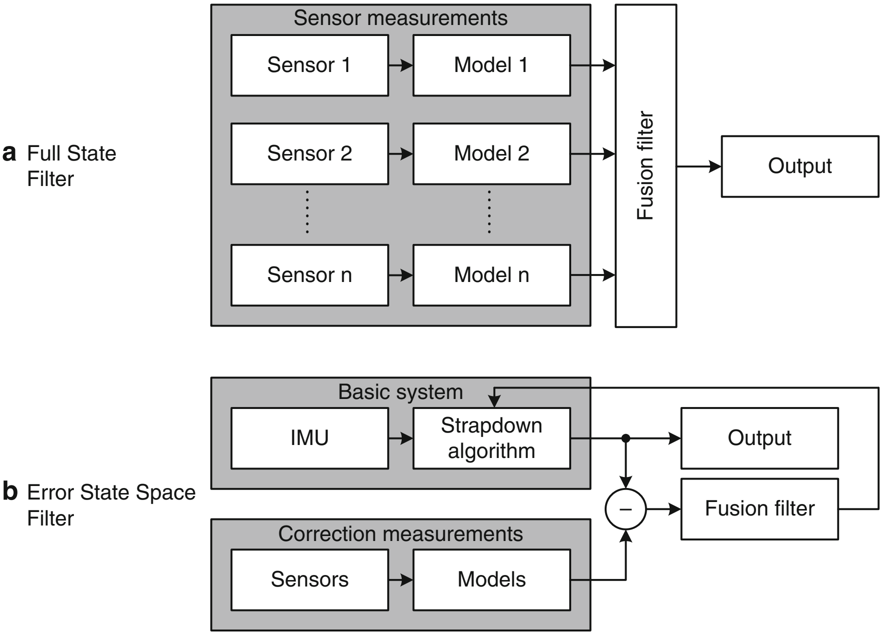


Fig. 2 Different types of state estimation

dependent on the current operating point, if a linearized filter is used, as is often the case. A typical filter of this type is an (extended) Kalman filter.

Error-State-Space Filter

In a fusion filter with estimation of error values (Wendel 2007), the difference between the output variables of a basic system and additional correction measurements is used as observable in the filter to estimate error states (e.g., position error instead of full position). This is illustrated in Fig. 2b using the example of an integrated navigation system: Here the basic system consists of IMU and strapdown algorithm, while GNSS and odometry, inter alia, are used for correction. The fusion filter estimates the error from the correction measurements relative to the basic system, and the navigation system output variables are then calculated in the corrected basic system. The temporal and value range dynamics of the error are usually small relative to the system dynamics. The resulting advantages of this method are therefore mostly constant output data characteristics, which are largely uncoupled from the dynamics, and negligibly small linearization errors. A weakness of this concept is its dependence upon the availability of the sensor supplying the basic system with data. In the case of vehicle navigation, this is usually an IMU, which is distinguished by its high availability. A typical filter of this type is an Error-State-Space Kalman filter.

3.3 Classification of Different Filter Types

Table 1 provides an overview of the different types of fusion filters with their essential characteristics. The list is based on descriptions given in Wendel (2007) and Thrun et al. (2006). In addition, the relevant technical literature presents and discusses many different variants and adaptations of these filters.

4 Enhancements for Fusion Filters

The following section describes enhancements to the listed fusion filter concepts, which offer advantages particularly for automobile sector applications.

4.1 Integration of Odometry Measurements

The use of IMU and GNSS is a typical variant of an integrated navigation system. In the motor vehicle sector, odometry signals are available from the wheel revolution sensors that have been part of standard equipment for a long time now. Odometry can serve to bridge any short-term unavailability of GNSS. Odometry measurements of speed are acquired in wheel coordinates and can be transformed into body coordinates via measured wheel steer angle and processed in the fusion filter.

The main systematic error that generally occurs, independently of an optional preceding tire model, is the dynamic tire roll radius error. This constitutes a scale

Table 1 Fusion filter types

Bayes filter	Linear parametric	Only of limited application in digital systems No continuous output variables
Kalman filter (KF)	Linear parametric	Assumption: system and measurements normally distributed (Gaussian white noise) Uncertainty propagation (variance and mean) Optimal estimator with Gaussian white noise (unbiased and minimal variance) Low computational effort Variances have squared unit of the state variables
Information filter (IF)	Linear parametric	Like Kalman filter, but propagation of the inverse variance-covariance matrix State of infinite (initial) uncertainty can be numerically represented Significantly more computational effort than KF, if state variances are required as an output
Extended KF/IF	Linearized parametric	Like KF or IF, but: Nonlinear relationships can also be modeled by linearization Loss of optimality (unbiased and minimal variance) due to linearization Moderately higher computational effort
(Extended) Error-State-Space KF	Linearized parametric	Like (Extended) KF, but: Estimation of error increments rather than full states Prediction step by basic system measurements Expected value of error increments is zero, therefore small linearization error Slightly less computational effort than (Extended) KF
Unscented KF	Nonlinearized parametric	Unscented transformation: nonlinearized propagation of points taken from the normal distribution, average, and variance and then calculated from the transformation result Advantage over EKF if there is strong nonlinearity Similar computational effort to EKF
Histogram filter	Nonparametric	Does not require normal distribution of input and output for optimality State space segmented into a finite number of regions (static or dynamic) Nonlinear propagation of points High computational effort is critical for real-time applications
Particle filter	Nonparametric	Random sampling points of the input distribution projected nonlinearly onto output Simple algorithm that is flexible in terms of the nonlinear system equation Number, density, and variance of the random sampling points can only be determined experimentally High computational effort is critical for real-time applications

factor error (Bevley and Cobb 2010) in the conversion of wheel revolution rates into speeds. The obvious solution is therefore to estimate the size of this error using the fusion filter and correct it. If we assume that, in normal traffic on public roads, the majority of driving situations involve a planar acceleration in the range $\leq 5\ \mathrm{m/s^2}$

(Hackenberg and Heißing 1982), modeling of slip and sideslip that is linearly dependent on acceleration is assumed to be suitable for supporting the navigation filter by means of odometry measurements. Odometry measurements are discarded, while the acceleration is outside this range. Under this boundary condition, the following systematic errors, which cannot be observed by measurements, are modeled:

- Velocity errors due to slip – correction via a linear tire model with assumed constant slip stiffness and acceleration measurement
- Velocity errors due to sideslip – correction via a linear tire model with assumed constant sideslip stiffness and velocity and acceleration measurement

The usual odometry models use measurements of an undriven and un-steered axis to estimate the velocity and yaw rate of the vehicle. Depending upon the system model of the fusion filter, these variables, but also individual wheel speeds, can be used as measured data for data fusion.

4.2 Compensating for Delayed Measurement Availability

Depending upon the structure and measuring principle of a sensor, there is a time lag between the time assigned to the measurement and delivery of the measured data to the fusion filter. Further lags arise due to different sampling rates and times of the basic systems and the various correction measurements, as well as filter delay. If this correspondingly outdated data is processed in the fusion filter, errors can occur, depending upon the operating point and dynamics of the measured variable. Thus, e.g., for a delay of 100 ms, which is typical for many GNSS receivers, and at a vehicle speed of 30 m/s, we obtain a position error of 3 m.

For the special case of an Error-State-Space filter, a method is outlined (Dziubek 2013a), whereby several sensor measurements with different amounts of delay are processed with a sufficiently low computational effort for real-time use. The following assumptions are made for this:

1. The changes in the errors of the states $\vec{X}$ in the basic system are considered to be negligible and independent of the measured values within a time span. This time span τ corresponds to the longest delay of measured data to be expected in the system that still fulfills the stated assumption. Depending upon the sampling rate f_{Basis} of the basic system, n measured values are held in the memory (n rounded up to integer value) for the time span τ:

$$n = \tau \cdot f_{Basis} \tag{1}$$

2. If assumption (1) is valid, it is permissible within τ to divide the n stored historical data, which already is corrected by the currently known error $\vec{X}_n$,

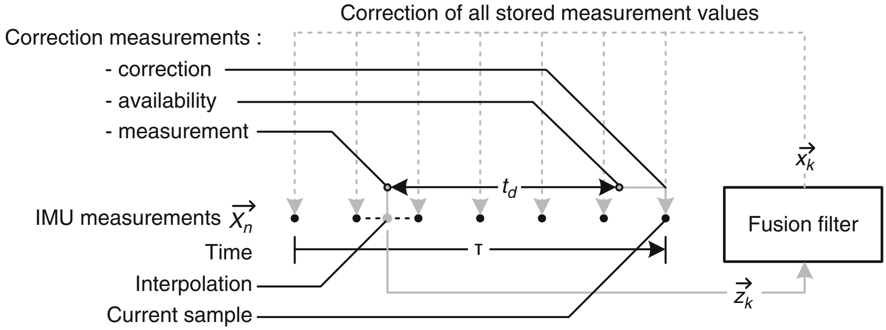


Fig. 3 Functional principle of compensating for delayed availability

into real operating points $\vec{V}_n$ and independent residual errors $\vec{\varepsilon}$. Since the error is assumed to be constant over τ, it is identical to the error of the current measuring epoch and can therefore also be correctly described by the fusion filter's stochastic model:

$$\vec{X}_n = \vec{V}_n + \vec{\varepsilon} \tag{2}$$

3. All changes in measured data between two subsequent filter processing steps can be described as approximately proportional to the duration, so that errors associated with a linear interpolation can be regarded as negligible.
4. The latency t_d between the current sampling time of the basic system of the Error-State-Space filter and the correction measurements is generally known or determinable.

As can be seen in Fig. 3, in the Error-State-Space Kalman filter, the observation vector $\vec{z}_k$, which is the filter input variable, is obtained from the difference between correction measurements and the measurement of the basic system. If assumption (1) is valid, it is permissible that an error which was determined at t_d in the past is applied in the current measuring epoch without loss of accuracy, so long as $t_d \leq \tau$. Thus, the storage of the data of the basic system used to calculate $\vec{z}_k$ within time span τ is sufficient for the purpose of virtual past measurements. During the lag time, corrections using other sensor data with different delay times are also generally taking place. In order to maintain assumption (2), on correction of the current measured data by the error increments $\vec{x}_k$ calculated by the filter, the error $\vec{\varepsilon}$ that applies to all stored measured data $\vec{X}_n$ is also corrected. This guarantees that, irrespective of the time lag, the current errors are always being corrected even for the stored states. Since the value $\vec{X}_m$, calculated by the basic system at the mth sampling interval, already contains the totaled correction increments $\vec{\varepsilon}_0$ from the past, the following applies for updating the corresponding correction $\vec{\varepsilon}$:

$$\vec{\varepsilon} = \vec{\varepsilon}_0 + \sum_{(s=0)}^{(s=m)} \vec{x}_s \tag{3}$$

where $m = k$ for the current measurement epoch. This allows a computing time-efficient, recursive implementation, by summation of the corrections $\vec{x}_k$ calculated by the filter in a measuring epoch, to all stored $\vec{X}_n$ values.

Since synchronous sampling of basic system and correction measurements is generally not assumed, under assumption (3), linear interpolation of the measured dataset $\vec{X}_k$ is used for error computation. Under assumption (4), the lag time t_d is used to select the two measurements of the basic system $\vec{X}_i$ and $\vec{X}_j$ closest to this time point and therefore include it. Here $t_i < t_d \leq t_j$. Linear interpolation between the measured values takes place as follows:

$$\vec{X}_k = \vec{X}_i + \left(\vec{X}_j - \vec{X}_i\right) \cdot \frac{t_d - t_i}{t_j - t_i} \tag{4}$$

The values $\vec{X}_k$ corrected by the currently known error and interpolated to the correction measurement time are subtracted from the correction measurements to give $\vec{z}_k$ and used to correct the states.

Figure 3 illustrates the functional principle of compensating for delayed availability. We can see how the previously stored values $\vec{X}_n$ of the basic system are used as valid corrections for the next filtering step after correction measurements are available. The resulting error corrections $\vec{x}_k$ are applied to all stored values. This compensates for general time lags between several sensors.

4.3 Plausibility Check

Fusion filters are normally based on a stochastic model, which allows measured and estimated data to be weighted. This involves describing uncertainties of the measured and fused data by means of stochastic characteristics. In the example of the Kalman filter, these are the measurement and estimation uncertainties, which are modeled in the form of variance-covariance matrices. As long as measurement errors are random in nature and correspond to the assumed statistical distribution in terms of their size and frequency, the optimality of the estimation algorithm is not impaired. However, this cannot be guaranteed in reality, since sensors and measurements are also subject to the influences of systematic errors, an example of this being multipath interference in GNSS. It is therefore necessary to check the plausibility of the measured data, and a concept for this is illustrated below. The aim of a plausibility check is to identify and remove measurements which are not compliant with their stochastic model. The concept fulfills the following requirements:

- Using the total redundancy of all measured data for error detection
- Compatibility with the fusion filter concept and its requirement of the processing of heterogeneous measurements
- Checkability of the measured data, irrespective of times of measurement and sampling rate of the sensors
- Removal of measurements only when they deviate significantly from their stochastic model by rigor tests parameterized with defined significance level
- Retention of the largest possible number of unaltered measurements for the fusion filter
- Consideration of all measurement and estimation uncertainty models to achieve a self-regulating behavior adapted to the current uncertainty of the filter

The application of plausibility checking is illustrated using the example of an integrated navigation system with an Error-State-Space Kalman filter (see Sect. 6). Here, the basic system consists of an inertial measurement unit and a strapdown algorithm, and the plausibility check is performed, for example, for GNSS pseudorange measurements, in a tightly coupled architecture. Plausibility is checked both by comparing measurements of the basic system with the pseudoranges and also by comparing the individual pseudoranges of a measurement epoch with each other.

The fusion filter uses Gaussian white noise as a stochastic model, with a standard deviation σ_{Meas} of a normal distribution being modeled for each measured value.

For the stochastic method *plausibility check of individual measurements against the basic system*, the total uncertainty σ is determined from the uncertainty of measurement σ_{Meas} and estimation σ_{Basis}, and a $n{\cdot}\sigma$ environment is used to calculate the error detection threshold ξ. For the method *plausibility check of individual measurements*, *in this case pseudoranges*, *against each other*, which is largely independent of the filter, the individual uncertainties of measurement σ_{Meas} are used to calculate the $n{\cdot}\sigma$ environment. In both cases, the parameter n determines the significance level and is freely selectable. Cross-checking of pseudoranges benefits from as few faulty measurements as possible, while checking against the basic system is not dependent upon the number of outliers. It is therefore advantageous to first run the check with the basic system.

4.3.1 Checking the Plausibility of an Individual Measurement Against the Basic System

The observable of the fusion filter for pseudoranges (code measurements) is the reduced observation δ_{PSR}, i.e., the difference between the measured pseudorange and that calculated from the available data (user position, satellite position, etc.):

$$\delta_{PSR} = z_{PSR} - \check{z}_{PSR} \tag{5}$$

where z_{PSR} is the pseudorange obtained by the receiver from the travel time measurement and $\check{z}_{PSR}$ is the distance calculated from the current position estimate of the strapdown algorithm and the satellite position known from the ephemeris data. The standard deviation of the pseudorange measurement $\sigma_{PSR,Meas}$ is obtained from the "receiver clock error converted into distance" filter state computed by Tightly Coupling with standard deviation σ_{Clk} and from the measurement noise σ_{PSR} of the pseudorange measurement:

$$\sigma^2_{PSR,Meas} = \sigma^2_{Clk} + \sigma^2_{PSR} \tag{6}$$

A three-dimensional projection of the filter state "position error" with standard deviation σ_{Pos} into the line of sight to the satellite is not carried out, to avoid coupling of the vehicle attitude with the threshold values. The position uncertainty σ_{Pos} is therefore calculated as a worst-case estimate from the geometric sum of coordinate uncertainties, in this case given in navigation coordinates:

$$\sigma^2_{Pos} = \sigma^2_e + \sigma^2_n + \sigma^2_u \tag{7}$$

The $n{\cdot}\sigma$ environment and therefore the threshold $\xi_{IMU,Code}$ are then obtained as

$$\xi_{IMU,Code} = n \cdot \sqrt{\sigma^2_{PSR,Meas} + \sigma^2_{Pos}} \tag{8}$$

In the fusion filter, the parameter n establishing the significance level for measurement error detection is a trade-off between high availability of measurements (n larger) on the one hand and increased error elimination (n smaller) on the other. The test rates a measurement as invalid if

$$|\delta_{PSR}| > \xi_{IMU,Code} \tag{9}$$

4.3.2 Plausibility Check of Individual Measurements, in This Case Pseudoranges, Against Each Other

The plausibility check is based on the geometry of the spatial triangle spanned by the user GNSS antenna and two observed satellites p and q (Dziubek 2013b). Assuming that the receiver position is approximately and the satellite positions are precisely given (so they can be regarded as error-free for the purpose of the plausibility check), the distance of the satellites l_{Eph} is calculated as a reference value:

$$l_{Eph} = \left\| \vec{r}_{p,Eph} - \vec{r}_{q,Eph} \right\| \tag{10}$$

This satellite distance can also be calculated as l_{Meas} from the measured pseudorange z_p and z_q according to the cosine theorem:

$$l^2_{Meas} = z^2_p + z^2_q - 2 \cdot z_p \cdot z_q \cdot \cos(\alpha) \tag{11}$$

Assuming that the receiver position error of the strapdown algorithm is negligibly small relative to the overall length of a pseudorange (on average approximately 22,000 km), the angle α to the two satellites in view from the receiver position can be computed "error-free."

The difference in length Δl of the satellite distance from the two calculations serves as a parameter for evaluating the pseudorange error:

$$\Delta l = \left| l_{Meas} - l_{Eph} \right| \tag{12}$$

From the Gaussian error propagation of the measurement noise (variances) of the pseudorange σ_p^2 and σ_q^2 and the uncertainty of the ephemeris data σ_{Eph}^2, we obtain for the standard deviation $\sigma_{\Delta l}$ of Δl:

$$\sigma_{\Delta l} = \sqrt{\left(\frac{z_p - z_q \cdot \left(\vec{e}_p \cdot \vec{e}_q\right)}{l_{Meas}}\right)^2 \cdot \sigma_p^2 + \left(\frac{z_q - z_p \cdot \left(\vec{e}_p \cdot \vec{e}_q\right)}{l_{Meas}}\right)^2 \cdot \sigma_q^2 + \sigma_{Eph}^2} \tag{13}$$

The ephemeris error can be assumed to be negligible, so that for l_{Eph} we obtain a variance of $\sigma_{Eph}^2 = 0$. The threshold value $\xi_{Code,Code}$ is calculated as

$$\xi_{Code,Code} = n \cdot \sigma_{\Delta l} \tag{14}$$

It is concluded that there is an error in one of the measurements z_p or z_q, if the condition

$$\Delta l > \xi_{Code,Code} \tag{15}$$

is fulfilled.

All p, q $(p \neq q)$ combinations are tested to isolate the faulty measurements. If there are r observed satellites, the number s of pair comparisons is calculated, provided that $r \geq 2$, as required for checking, using the Gaussian summation formula:

$$s = \frac{(r-1) \cdot r}{2} \tag{16}$$

The check is performed for all s pairs of satellites. For each individual satellite, a numerator F_r is incremented for each pair with detected inconsistency. The numerator is then increased for both satellites involved.

A threshold F_{Max} for rejecting a measurement should be chosen such that data are only rejected when an error is clearly identified. Particularly for the cases:

- $r = 2$: If an error is detected during the pair comparison, it cannot be clearly assigned to one of the two measurements. Therefore, to safely avoid an error, both measurements must be rejected on the basis of the geometric comparison.

- $r \geq 3$: Assuming that faults occur randomly or due to geometry to a varying size in all measurements, those measurements should be rejected where faults are detected in a minimum number of pair comparisons.

These conditions are achieved by setting the minimum number of checks with fault detection equal to the number of available pairs, but at least equal to 1:

$$F_{Max} = r - 1 | F_{Max} \geq 1 \tag{17}$$

If the sum of detected errors of a measurement is equal to or greater than F_{Max}, this measurement is rejected. This completes the plausibility check for the GNSS pseudoranges of a measurement epoch.

The principle of checking against the basic system, as well as against other measurements from a measurement epoch, can be transferred accordingly to other measured quantities such as deltarange and odometer measurements.

5 Data Quality Description

This section outlines methods for describing data quality, beyond the quality of the states calculated by the fusion filter in the form of a variance-covariance matrix. First of all we explain the quality measure *integrity*, which aims to rate the consistency of redundant data. We present a selection of the many methods for integrity assessment. We then present a concept for calculating an accuracy measure, which provides an integral quality assessment of the data output of the fusion filter to application functions.

5.1 Integrity

According to Strang et al. (2008), the term *integrity* in navigation and positioning is mostly described as (quotation translated from German),

> "(. . .) the correctness of the position information provided by the localization component (. . .)"

and this definition can in principle be extended to cover all quantities to be estimated by the fusion filter. A statement regarding correctness is given in the form of an integrity measure.

5.1.1 Requirements

Even in the absence of disturbances, data and measurements have a defined variance and an expectation value. Deviations between measurements are therefore inevitable and permissible within the specified measurement precision without this indicating an error. As long as the measurements lie within their specified dispersion range, it is assumed that they are consistent, i.e., the measured data are free from contradictions within the range of their uncertainties. Assessment of data integrity based on these assumptions requires that at least two redundant datasets are available, so that they can be cross-checked for consistency, i.e., the absence of contradictions in the sense of the assumed stochastic model.

Functions using the fused data therefore require both the result of the consistency check and also characteristic values regarding the testing power and availability of this check as integrity information for assessing the data. This gives rise to the following general requirements of an integrity measure:

- Cross-checking of all available redundant measurements for consistency
- Detection of errors and inconsistencies with the shortest possible detection time and/or time to alarm and defined significance level
- Issuing the result of the check in the form of a statement on usability of the data
- Issuing a confidence measure to describe the test power and to take account of uncertainties and availabilities

This gives rise to the integrity measure defined here as a combination of checking the available data for consistency, assessing test power on the basis of the uncertainty of the data and overlapping of the confidence intervals.

5.1.2 Algorithms for Assessing Integrity

According to Strang et al. (2008), the definition of integrity in navigation and positioning can be extended as follows (quotation translated from German):

> (. . .) the correctness of the position information provided by the localization component (. . .). This is described by two values: threshold value and time-to-alarm.
>
> The threshold value specifies the position error that can still be tolerated for a particular application. It is also known as the Protection Level and is usually given separately in the horizontal plane (Horizontal Protection Level, HPL) and in the vertical axis (Vertical Protection Level, VPL).
>
> The time-to-alarm (ToA) describes the time span allowed between occurrence of the event triggering the alarm and its capture at the output from the localization component. (. . .)

The term *integrity* can be concretized (Pullen 2008) by defining the following sub-parameters: integrity risk, alarm limit, and time to alarm. In this context *integrity risk* means the probability of an unacceptable system error occurring without a prompt warning being given. The *alarm limit* defines the threshold value of the still acceptable error, where the integrity alarm is triggered when the error exceeds the threshold. The *time to alarm* is described as the time between occurrence of an unacceptable error in the navigation solution and triggering of the alarm.

In a general sense the terms *correctness* and *accuracy* describe the compliance with or the definition of a confidence interval that includes all data uncertainties.

A basic statistical method of quality control for measured and estimated data and suitable for assessing integrity is the *global test* (or *overall model test*) (Leinen 2009), in which a Gauss-Markov model is checked for compliance with an assumed χ_n^2 distribution within a defined significance level. In order to assess the optimality of a parameter estimation method, two criteria (Bar-Shalom et al. 2001) are used to rate consistency. These are the expectation value of the estimate, which should coincide with the true value, and the minimal variance of the estimate. In this context, the mean square error is a common indicator of both criteria. A suitable method for the global test is the *normalized innovation squared* (*NIS*) *test*.

The checkability of data is used as a basis for sensor validation and for detection of significant errors. The following approaches are suitable for this (Pourret et al. 2008):

- Hardware redundancy: cross-checking of information from several identical sensors
- Analytical redundancy: cross-checking of model-based information linked with other sensors
- Temporal redundancy: checking of several runs of the same experiment, therefore not real-time capable
- Knowledge-based methods: modeling of process knowledge/human knowledge that can be used to identify inconsistencies between signals

Redundancies are used, for example, in sensor validation using Bayesian networks (Pourret et al. 2008). Each sensor is assigned a validity probability via a combination of conditional probabilities. Apart from the previously mentioned NIS test (Agogino et al. 1988), which is based on redundancies, other validity-checking methods are the parity space method (Abdelghani and Friswell 2001) and the mathematically similar principal component analysis (Ding et al. 2004). Both of these are based on dividing observations into statistically independent components and then checking them. A knowledge-based method is fuzzy logic (Goebel and Agogino 1999), which is frequently used for sensor validation in power plants.

Receiver autonomous integrity monitoring (RAIM), which comes from the GNSS sector, is an umbrella term covering several different methods. In particular, these use redundancy-based methods of integrity assessment in geodesy, navigation, and positioning.

The common purpose of RAIM algorithms is onboard measurement error detection with the shortest possible time to alarm and defined significance level. Moreover, an estimate of the maximum effect of an undetected error is calculated as an assessment parameter for the current system state.

The usual RAIM algorithms used in geodesy are based on the detection, identification, and adaption (DIA) method (Kuusniemi 2005), which uses a global test to detect inconsistencies that differ significantly from the stochastic model and, if necessary, identifies outliers via a local test. Adaption (Bhatti 2007) to the outliers can be achieved by replacing the erroneous measurement or by adapting the null hypothesis to the outlier. In RAIM, the null hypothesis $\boldsymbol{H_0}$ is that the deviations

Table 2 Test scenarios of statistical hypothesis tests

Facts	$\boldsymbol{H_0}$ is accepted	$\boldsymbol{H_\alpha}$ is accepted
$\boldsymbol{H_0}$ is true	Correct decision Confidence level $1-\alpha$	False alarm (type 1 error) Error probability = significance level α
$\boldsymbol{H_\alpha}$ is true	False alarm (type 2 error) Probability β	Correct decision Test power $1-\beta$

(residuals) of the measurements behave like normally distributed random variables, in accordance with their stochastic model. In contrast, the alternative hypothesis $\boldsymbol{H_a}$ assumes an error: if this hypothesis is accepted, an integrity alarm is triggered. When the hypotheses are checked by the global test (Kuusniemi 2005), with stochastic parameters α and β, we obtain the scenarios shown in Table 2:

RAIM applications in navigation are based on the following three basic fault detection methods:

- Range domain: consistency checks of pseudorange measurements.
  - Generalized: consistency check on the (raw) measured data level
- Position domain: test statistic of position fix is derived from subsystems.
  - Generalized: consistency check on the level of the fused result
- Time domain: consistency check based on the plausibility of the course of measured data over time.

In addition, w.r.t. forming the test statistic in the time domain, the following methods can be distinguished: snapshot methods, which only use the data from the current measurement epoch, and sequential methods, which also use stored historic data in the calculation.

For the four basic RAIM methods belonging to the snapshot methods in the range domain, least squares residuals, parity space, range comparison (Brown 1992), and normalized solution separation (Young and McGraw 2003), mathematical equivalence is proven.

The multiple solution separation (Bhatti 2007) snapshot method in the position domain is based on the assumption that only one measurement per measurement epoch is faulty. In N observations, in addition to the overall solution, position solutions with $N-1$ observations are formed, excluding one observation in each case, so that, in accordance with the assumption, at least one correct solution exists. The test statistic is formed by evaluating the differences from the overall solution; because it involves the calculation of several solutions, this method requires a high computation effort.

The sequential range domain method of Autonomous Integrity Monitoring by Extrapolation (AIME) (Bhatti 2007) is certified for use in the Airbus family. It determines the test statistic via the innovations of the fusion filter over the periods 150 s, 10 min, and 30 min, so that, in principle, slowly growing errors can be detected. However, this also means a corresponding increase in the time to alarm. This process is particularly suited to tightly coupled methods, because, by involving the inertial navigation system, it is able to check redundant data, even if there are fewer than four satellites observable.

A weakness of the outlined methods, which are based on a global χ^2 test, is that slowly growing errors, caused, for example, by changes in the ionospheric delay of GNSS signals or IMU offset drifts, cannot be detected or – as with AIME – can only be detected over a long period. A potential solution that has been described for this is the Rate Detector algorithm (Wu et al. 2008), which observes the rate of change of the test statistic via a separate Kalman filter to detect continuous deviations from the expectation value that are, however, too small to trigger an integrity alarm.

One approach that addresses the known shortcomings of RAIM algorithms – assumption of only one simultaneous error and inability to detect slowly growing errors – is piggyback architecture (Bhatti and Ochieng 2009). This converts the inertial measurements into virtual pseudorange measurements and runs the AIME test statistic, fault detection, and isolation after solution separation (Bhatti 2007) – i.e., the calculation of several navigation solutions, excluding individual observations and the calculation of the NIORAIM protection level (Hwang and Brown 2005). Because it calculates virtual pseudoranges and uses solution separation, this method requires a high computation effort.

The concept of Vehicle Autonomous Integrity Monitoring (VAIM) (Feng and Ochieng 2007) has been introduced especially for road traffic applications. For using sensors as used in series production, the high-integrity IMU/GPS navigation loop method is described in Sukkarieh et al. (1999). Both methods use greatly simplified assumptions and algorithms and have the disadvantages of RAIM algorithms already outlined above.

A time domain method, which, unlike the other methods, does not observe states and measurements but the trajectory traveled, is trajectory monitoring (Le Marchand et al. 2009). However, this method has also been shown to have weaknesses at low speeds.

There are also algorithms based on the use of several models: for example, interactive multiple model filtering (Toledo-Moreo et al. 2007) uses two different models for fault detection and switches between them depending upon the driving situation. Multiple-model adaptive estimation (MMAE) (Abuhashim et al. 2010) addresses the known shortcomings of RAIM algorithms and their disadvantage of only being capable to work with correctly parameterized fusion filters, by having a bank of several independent filters with different error hypotheses, which assign probabilities to the modeled errors. The multiple calculations that are required mean that both methods require greater computation effort. Also, these approaches target predefined error types and cannot respond appropriately to un-modeled error types.

5.2 Accuracy

5.2.1 Requirements

Data processing or control in user functions requires detailed information comprising several signal properties. Information about the total uncertainty (Mansfeld

2010) of the filter states is not sufficient on its own to establish control or regulation with dynamic quality. Therefore, a real-time description is given of typical characteristics of measured and also processed data in the form of accuracy measures for different classes of characteristics. The *virtual sensor* is therefore enhanced to provide the applications with a dynamic *virtual datasheet* corresponding to the current availability and accuracy of the sensors. This contains all the information about the signals fused in the current measurement epoch required for processing in the applications, but it is so far abstracted that there is no direct dependency upon individual sensors. The signal sources and the user functions are consequently decoupled, not only on the data level but also on the description level.

This leads to the following requirements for the accuracy measure:

- Abstraction of the descriptive level by the virtual sensor
- Self-description of the virtual sensor specific to the application via a real-time datasheet
- No impact upon the fusion result
- Description of all fusion filter states
- Consistency with the existing fusion filter

5.2.2 Existing Measures

For the purpose of describing the essential terms (Pullen 2008) used for assessing the performance of a navigation system, *accuracy* is defined as a statistical quantity of the difference between the estimated position and the unknown true position. Depending upon the particular application and the assumed distribution function, various measures are used to describe an uncertainty interval, such as “circular error probable” (CEP). Also, the concept of *continuity* is defined as the reliability of the position issued by a navigation system: the *continuity risk* describes the probability of the system ceasing to deliver the specified output data. In the strictest sense, the concept of *availability* is defined as the simultaneous compliance with the requirements for accuracy, integrity, and continuity. However, it is worth mentioning that only partial fulfillment of these criteria often suffices in practical applications. Therefore, Pullen (2008) suggests the more practical definition of availability as the fulfillment of the system requirements at a specific point in time.

A generally valid quality description that is applicable to measured or estimated data is proposed in Wegener and Schnieder (2013). Measurement quality is used as the overarching concept, which is made up of the following components:

- Measurement uncertainty: quantitative description of the doubt about the measured result in the form of an overlapping interval, within which the true value of the measurand is to be found with a defined degree of confidence
- Measurement accuracy: qualitative, theoretical measure for the approximation of a measured value to the true, unknown value
- Consistency: description of the consistency of various measured values
- Latency (delay time): time between measurement and provision of measured data

- Availability: provision of data at a specific point in time (“point availability”)
- Reliability: probability of the availability of measured data beyond a defined period

Even though this definition is oriented toward real descriptors that are relevant for practical application, the classification of measurement uncertainty and measurement accuracy is still restricted to the standard description of total error used to date. In the following section we therefore use the term of an accuracy measure in the sense of measurement accuracy, but enhanced by classification into different types of errors.

5.2.3 Concept of an Accuracy Measure

In order to describe the characteristics of measured data, they are classified into different types of errors. This divides the total error into individual errors. The accuracies assigned to these individual types of errors are referred to below as descriptors. The computation and forwarding of descriptors to user functions allows the current signal characteristics to be individually assessed for each function, even if there is no longer any direct link to the sensors. Classification into descriptors provides additional information; the summing up of individual errors according to the error propagation law gives the total uncertainty in turn.

As a rule, measured data is processed in steps but always based on fundamental operations. From this it is possible to subdivide the signal processing that has taken place into separate sections modeled as black boxes, which always have the descriptors as input and output vector. Within these encapsulated systems, the output values of the descriptors are computed in the form of an error propagation, whereby known interdependencies of descriptors are taken into account by using an error propagation law. Otherwise, for simplicity, the descriptors are considered to be independent and without impact upon each other. Optionally, additional parameters can be used for computing the descriptors, for example, via corrections from the fusion filter.

An example implementation is shown in Fig. 4 in the form of a block diagram. The modeled operations cover the correction of offsets and scale factor errors of an acceleration measurement $\vec{a}_b^{IMU}$ via the fusion filter, its rotation in navigation

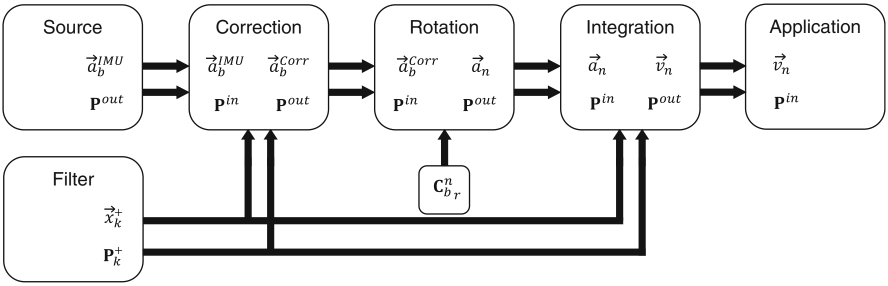


Fig. 4 Structure of the accuracy calculation

coordinates by the rotation matrix $\hat{\mathbf{C}}_{b_r}^{n}$, and their integration into a velocity $\vec{v}_n$ with simultaneous correction of the absolute value by the fusion filter. The notation ^{in} is used for input values and ^{out} for output values.

Modeling of the signal path starts with the sensors as source. Start values are used for the descriptors in accordance with the sensor specifications in their real datasheets. In this way, the signal characteristics are always specified at each signal processing step corresponding to the current operating state. In terms of fulfilling this specification, the continuity risk of an Error-State-Space fusion filter in accordance with the above definition is obtained from the continuity risk of the basic system, here the IMU, since its availability and specification compliance represents the minimum basis necessary for operation of the fusion filter.

The descriptors are determined on the basis of the user function requirements. For the calculation method, a specific error propagation law is selected for each characteristic. In principle, the error propagation calculation can be performed with user-defined distribution functions individual to the descriptors.

6 Example of an Implementation

6.1 Architecture

Below we give an example application of a fusion filter for high-precision positioning. For this we have selected an Error-State-Space approach with an extended Kalman filter, based on the sensors outlined in Sect. 2.2. A MEMS inertial measurement unit with 6° of freedom is used in combination with a strapdown algorithm as a basic system. Corrections are provided by GPS code and carrier phase measurements and odometer measurements in the form of Tightly Coupling integration. The filter is extended to include the compensation of delayed availability outlined in Sect. 4.2, and the plausibility check described in Sect. 4.3 is integrated for correction measurements.

The fusion filter (Dziubek and Martin 2013) consisting of the described blocks is illustrated in the structural diagram in Fig. 5. As a central element of data fusion, the basic system includes the correction of IMU sensor errors, the strapdown algorithm, and the fusion filter (Error-State-Space KF).

The Tightly Coupling loop is made up of the preprocessing and measurement models for the raw measured data from GPS pseudorange and deltarange measurements and from odometry. The preprocessed and corrected data from these blocks are inputs for the Error-State-Space Kalman filter, and the corrections for Tightly Coupling are fed back in a closed control loop as output variables from the filter. The relative latency of the GPS and odometer measurements is measured and compensated for by correction for delayed availability. The plausibility check detects and eliminates correction measurements, which, due to external disturbances, no longer correspond to their stochastic model. This is done both as integrated checking using measurements and uncertainties from the fusion filter

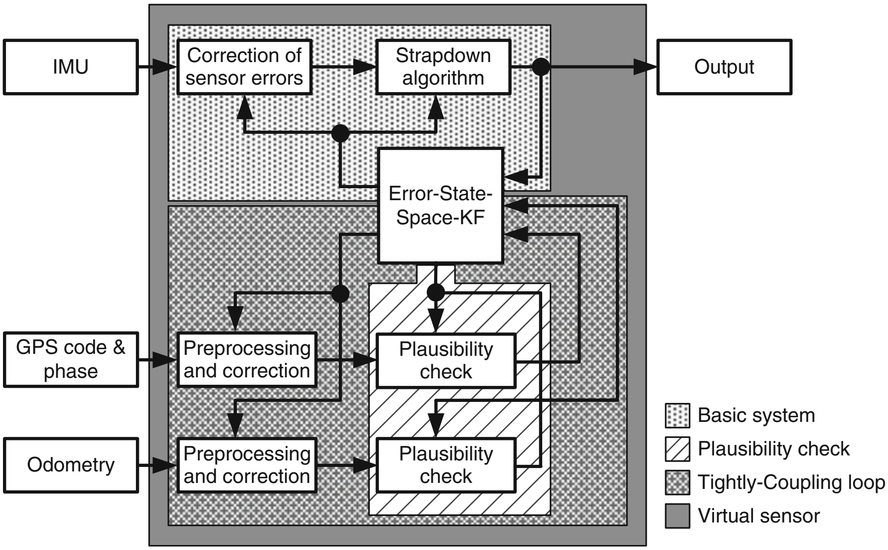


Fig. 5 Block structure of the fusion filter

and the correction measurements and also via the model-based check of redundant correction measurements against each other independently of the filter.

This modular filter structure allows additional correction measurements to be added without changing the existing measurement models. The minimum requirement for integrating an additional correction measurement is parallel or analytical redundancy for at least one of the filter states, which also allows plausibility checking with measured data from the basic system. If a correction dataset contains several measurements with mutual parallel redundancy, the abovementioned check against each other, which is independent of the fusion filter, can also be performed via a model description of known relationships. Furthermore, in order to correct the time delay of a measurement, this delay must be measurable or must be known. If the correction measurement is affected by systematic errors, which are observable by the measurement itself or via other measurements, they can be included in a Tightly Coupling control loop.

In this case, the structure of the fusion filter fulfills the system architecture requirements detailed in Sect. 1. Sensors, IMU, and correction measurements are decoupled from the output of the data fusion. Therefore, the filter acts as a virtual sensor and calculates a consistent set of output data independently of the number of correction measurements available. Since the strapdown algorithm is time invariant with deterministic process steps, a constant group delay time is achieved for the signals. Due to the triggered computation of the basic system, the output rate is identical to the IMU measurement rate; this is the highest of all sensors used. Moreover, IMU availability is not restricted by external influences but is solely determined by the hardware reliability of the sensors. In all the other sensors used

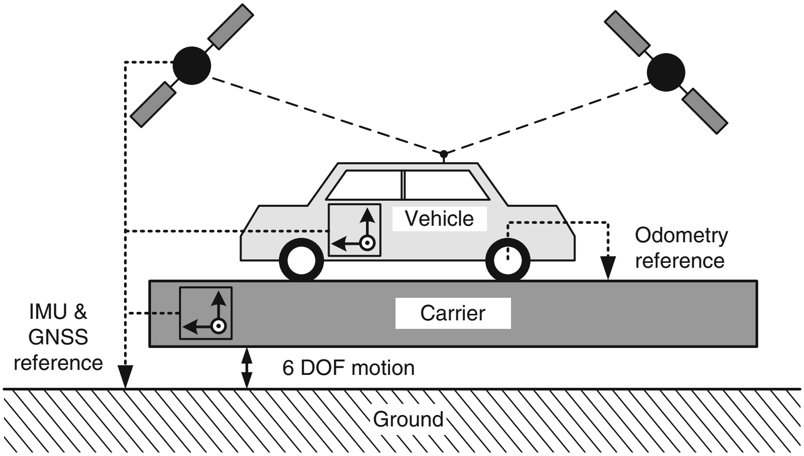


Fig. 6 Vehicle on carrier platform

for correction measurements, dynamic changes in availability as well as different and nonconstant sampling rates are compensated for.

The fusion filter that has been described and its enhancements essentially fulfill the requirements for specific application in the automotive sector. The system architecture of the fusion filter enables a real-time signal processing adapted to the characteristics of heterogeneous sensors. Thus, as a central virtual sensor, the fusion filter is able to generate a consistent dataset with better accuracy than the individual sensors.

6.2 Moving Reference Systems/“Carrier Platform”

A special case of a moving vehicle that is encountered in practice is when the vehicle is transported on a carrier platform, which itself moves w.r.t. the Earth, as shown in Fig. 6. Examples are transport on a car ferry, a turntable in a car park, or transport on a trailer or car train. In this instance, the carrier platform performs movements generally with 6° of freedom: because of the heterogeneous measuring principles of the fused sensors, this leads to the following inconsistencies between measurements in the vehicle and therefore to incorrect estimation of the vehicle dynamics states by the fusion filter.

Existing systems for correcting or compensating of sensor errors, for example, use model assumptions about straight travel and the offset error as low frequency effect (Gross-Bölting and Kolkmann 2002), threshold values for detecting situations that are suitable for a reconciliation (Keller et al. 2002a), and the formation of a regression line incorporating wheel revolution measurements (Keller et al. 2002b) to reconcile sensors – in this case the yaw rate sensor. If such a reconciliation takes place on a moving carrier, a wrong error value is determined due to the superimposed motion. Additional boundary conditions are therefore defined to prevent this type of false reconciliation.

The existing systems that have been outlined above are therefore solutions to specific technical problems that occur in practical application. What they have in common is that they only allow reconciliation in special situations clearly identified by logical linkages and that, if they represent any remedy at all, it is only to the symptoms that arise due to a carrier motion. In contrast, the aim of the approach outlined below is, by modeling known physical relationships and sensor measuring principles, to model the cause of the error, i.e., the carrier motion, as a consistent part of the vehicle fusion filter with its sensors.

As indicated in Sect. 2.1, the inconsistency in the measurements due to the motion of the carrier is the result of the sensor reference points being located in different coordinate systems. The sensors used in the fusion filter are affected by heterogeneous reference points as follows:

- IMU: absolute measurement of the dynamic quantity acceleration and angular rate in the inertial coordinate system; as described in Sect. 2.2, these measurements are converted into the terrestrial coordinate system via the strapdown algorithm. Movement of the vehicle due to carrier motion can be measured by the IMU and is correctly processed via terrestrial referencing.
- GNSS: absolute measurement of position and velocity of the receiver antenna; the measurements are performed in terrestrial coordinates. A movement of the vehicle due to carrier motion can be measured by a GNSS and is correctly processed via terrestrial referencing.
- Odometry: absolute measurement of vehicle speed relative to the carrier, which in this case is not the Earth's surface; the carrier motion is therefore not measurable, and inconsistencies occur in the data on conversion into the vehicle-fixed Cartesian coordinate system, as defined in DIN 70000 (Deutsches Institut für Normung 1994).

When processing inconsistencies in the measured data due to carrier motion, it must be remembered that it is not expedient to model the carrier dynamics simply by increasing the odometer measurement noise, since the carrier dynamics do not occur randomly in the measured data and are therefore not zero mean values. Therefore, increasing the measurement noise (Bar-Shalom et al. 2001) only serves to slow down but not to prevent the accumulation of errors in the fusion filter.

If carrier motion data that are independent of the fusion filter are available, the inconsistencies can be corrected, for example, by a car ferry transferring the dynamic data of its own navigation system and the corresponding measurement uncertainties to the vehicle. If the position and attitude of the vehicle on the ferry are also known, this can be used to make a deterministic correction to eliminate the inconsistency in the measured data by superimposing carrier motion data and odometer data and computing the corresponding superimposed measurement noise of both measurands by error propagation.

If no carrier dynamic data are available, the carrier status of the vehicle is described by the two hypotheses:

(a) "Vehicle is definitely not on a carrier."
(b) "Not known whether the vehicle is on a carrier or not."

This is based on the following model assumptions:

1. The vehicle and the carrier do not move simultaneously, i.e., relative to each other. In the example of a ferry, this means that the ferry does not set out until all vehicles have been parked and secured and that the latter only start to move again after the ferry has docked.
2. A minimum time always elapses between parking of the vehicle and the start of carrier movement.
3. The dynamics of the carrier are always limited; the maximum dynamic can be modeled.

It follows from assumption 1 that the movement of a moving vehicle is not overlain by carrier dynamics. Thus, odometry measurement of wheel revolution rates >0 is a criterion for ruling out carrier movement. As an additional exclusion criterion, assumption 2 provides a standstill time below a typical threshold value. This limit should be set higher than typical standstills – for example, at traffic lights – but also be small enough to allow hypothesis "a" to be rejected quickly – if the vehicle actually is on a carrier. If hypothesis "b" is assumed to be valid, the odometry measurements are modeled as being potentially faulty.

This uncertainty modeling is intended to take account of the nonzero mean measured values and to influence the stochastic model of the fusion filter, in order to still be able to plausibly describe the errors that arise. For this purpose, the superimposed carrier movement is still assumed to be zero, and a variably adapted uncertainty is introduced via the stochastic model. Although this does not prevent inconsistent measurements resulting in errors in the fusion filter, it does prevent a drop in the variances to a level not corresponding to the resulting errors. It also prevents an implausible continuous rise in the uncertainties over time – as is the case where odometry measurements are completely rejected. These errors, which are therefore correctly modeled in the system model, are optimally corrected at the start of a drive – and therefore the changeover to hypothesis "a" – in terms of the Kalman filter.

The process of superimposing known carrier dynamic data and modeling unknown carrier dynamics can be used irrespectively of the availability of carrier motion data, without having to make changes at the fusion filter itself. Modeling is performed exclusively during preprocessing of odometry sensor measurements and their noise model. These values are applied as observations in the filter, so that the previously standard procedure of processing carrier dynamics as a special case with changeover of filter operating mode is no longer necessary.

6.3 Implementation of Integrity Measure

To assess the integrity of the data of the fusion filter outlined in Sect. 6.1 in one measurement epoch k, the innovation $\vec{i}_k$ calculated in the Kalman filter, that is to say, the unweighted state correction vector, is used. It describes the difference between correction measurements and the measurements of the basic system transformed by the measurement model $\mathbf{H}_k$ into the units of the correction measurement. The algorithm used for computing the integrity measure, which is optimized especially for Tightly Coupling as described in Sect. 5, is "Autonomous Integrity Monitoring by Extrapolation" (AIME). However, in the interests of time to alarm and computation effort, it is not the averaged innovation over defined periods that is used in this context, and AIME is therefore used as a snapshot method. Using the innovations $\vec{i}_k$ of the current measurement epoch, this method checks the null hypothesis H_0 of normally distributed input data by means of a chi-squared test. Thus, the method supplies the required clear integrity statement within one sampling interval of the basic system. The testing power, and therefore the probability of false detection or no detection, can be adjusted to the requirements of the user function in accordance with Table 2. Below, this selected method is implemented for the fusion filter. A test statistic (Bhatti 2007) TS_k is generated from the innovation in the form of the normalized innovation squared (NIS), which mathematically represents the sum of squares of statistically independent, standard normally distributed residuals:

$$TS_k = \vec{i}_k^{\,T} \cdot \mathbf{S}_k^{-1} \cdot \vec{i}_k \tag{18}$$

Standardization is done by means of the associated variance-covariance matrix $\mathbf{S}_k$ of the innovation. This consists of the filter internal variables usually used for Kalman filters, namely, "measurement model $\mathbf{H}_k$," "a priori variance-covariance matrix P_k^-," and "variance-covariance matrix of measurement $\mathbf{R}_k$":

$$\mathbf{S}_k = \mathbf{H}_k \cdot \mathbf{P}_k^- \cdot \mathbf{H}_k^T + \mathbf{R}_{\boldsymbol{k}} \tag{19}$$

If the null hypothesis H_0 is valid, TS_k is chi-square distributed. The expected value of TS_k corresponds to the number of available measurements N_k in the current measurement epoch k:

$$TS_k \sim \chi^2_{N_k} \tag{20}$$

$$E\{TS_k\} = N_k \tag{21}$$

Integrity is verified by checking the null hypothesis H_0 with the significance level adapted to the selected false alarm rate α. H_0 assumes that the system is error-free:

$$H_0 : TS_k \leq \chi^2_{N_k,\,1-\alpha} \tag{22}$$

If the inequality is satisfied, H_0 is accepted; otherwise, H_0 is rejected, and an integrity alarm is triggered.

A confidence measure coupled to the error threshold value is defined as a confidence interval to describe the power of the hypothesis test. This is referred to as the protection level PL_k (Diesel and Luu 1995) and is made up of the two components: system uncertainty PL_k^S and measurement uncertainty PL_M^K.

In this case the system uncertainty PL_k^S is synonymous with the weighted standard deviation of one or more (m) states relevant for the considered protection limit and is therefore calculated from the main diagonal of the a posteriori variance-covariance matrix $\mathbf{P}_k^+$ of the Kalman filter. The weighting by means of the statistical uncertainty limit is done using the parameter n in accordance with the requirements of the user function:

$$PL_k^S = n \cdot \sqrt{\mathbf{P}_k^+(1.1) + \mathbf{P}_k^+(\mathrm{i,i}) + \ldots + \mathbf{P}_k^+(\mathrm{m,m})} \tag{23}$$

In order to calculate the measurement uncertainty PL_K^M, an estimation is made of the impact of the maximum still-not-detectable error upon the fusion result – on the assumption that only one such error occurs at any one time during a measurement epoch.

The linearized calculation of the sensitivity of the fusion result $d\mathbf{R}_k$ to errors $d\mathbf{r}_k$ in the measured data is done by determining the slope (Diesel and Luu 1995) $\mathfrak{v}_k$:

$$\mathfrak{v}_k = \frac{d\mathbf{R}_k}{d\mathbf{r}_k} = \mathbf{K}_k \cdot \left(\mathbf{D}_k^{-\frac{1}{2}} \cdot \mathbf{L}_k^T\right)^{-1} \tag{24}$$

An eigenvalue decomposition of the matrix $\mathbf{S}_k$ is performed for this, whereby $\mathbf{D}k$ is obtained as the diagonal matrix of the eigenvalues and $\mathbf{L}_k$ as the modal matrix of the eigenvectors. $\mathbf{K}_k$ is the Kalman gain of the current measurement epoch.

As a boundary condition, it is assumed that only one measurement per measurement epoch is faulty (Bhatti 2007). Since the estimate of the maximum error is calculated, the largest common slope of the m respective states in a measurement epoch determines this:

$$\tilde{\text{o}}_{Max,\,k} = \max_m(\mathfrak{v}_k) \tag{25}$$

Since a statistical founded error detection probability is estimated at this point, an alternative hypothesis H_a is set up in accordance with Table 2. From the calculation of probability β for a type 2 error during checking of the test statistic, the maximum error that is still just undetectable is calculated:

$$\varrho_k = \sqrt{\lambda_{\beta,k}} \tag{26}$$

Here $\lambda_{\beta,k}$ is the non-centrality parameter of the $\chi^2_{N_k,\lambda}$ distribution assumed for the distribution of innovations at the selected values for significance level α and type 2 error . For the protection level, this gives a confidence interval of probability $1-\beta$.

This confidence interval is projected over the maximum slope $\upsilon_{Max,k}$ into the result domain, so that PL_K^M is calculated as follows:

$$PL_k^M = \upsilon_{Max,k} \cdot \varrho_k \tag{27}$$

The overall protection level PL_k is obtained from the two described components:

$$PL_k = \sqrt{PL_k^{S^2} + PL_k^{M^2}} \tag{28}$$

Using the example of horizontal position information, the protection level describes a confidence range of the estimated position; this is valid if no integrity alarm is given. The calculated integrity measure therefore consists of the results of the consistency check coupled via the stochastic assumptions and the protection level. As a snapshot method, the algorithm detects errors in the measured data within the current measurement epoch. The result of the test statistic check provides a clear result with defined significance level, while the protection level describes a confidence interval for the fusion result. Both measures take account of the uncertainty and availability of redundant measurement data.

6.3.1 Simultaneous Errors

In calculating the protection level, a maximum of one significantly faulty measurement per measurement epoch is assumed for Eq. 25. However, in practice, several simultaneous errors are to be expected in GNSS and odometry application in motor vehicles, for example, multipath interference of GNSS signals and odometry faults due to uneven ground surface. The probability of several such errors occurring simultaneously, and having mutually consistent effects, is assumed to be sufficiently small that this case can be ignored in real use. This type of error can therefore be detected by the plausibility check as described in Sect. 4.2. Since the plausibility check discards such faulty measurements, they are no longer used in the integrity assessment. The only case that the plausibility check does not cover is where an alarm is not triggered because there are not enough checkable redundant measurements. However, this does not principally contradict the assumption of a single error per measurement epoch, because, in this case, there are anyway only a few measurements available. Appropriate validation tests must be performed before the practical application of such an algorithm.

6.3.2 Slowly Growing Errors

A known shortcoming of RAIM and AIME outlined in Sect. 5 is the inability to detect slowly growing errors (SGE). In current practice these algorithms are therefore only used to a limited extent, and other methods, which are more costly in terms of computation time and memory requirement, such as the MMAE algorithm, are used instead. However, a disadvantage of those methods is that they only recognize errors reliably which are described in their models.

A typical instance of SGE is the time-variant error in a pseudorange measurement caused by changes in the ionosphere; this takes place so slowly (Bhatti 2007) that it brings about a wrong correction in the estimated position of the filter. However, the amount of change from one measurement epoch to the next is not big enough to detect the error using a snapshot method. In the case of the fusion filter illustrated here, SGE are to be expected for the utilized sensors in the following cases:

- IMU: drift of offset or scale factor errors caused by defects or external influences – such as ambient temperature.
- GNSS: pseudorange measurements due to ionospheric influences and multipath reception; deltarange measurements, on the other hand, are not affected due to time differentiation of the measured values.
- Odometry: slowly growing errors of measured speed due to changes in the roll radius, for example, due to gradual loss of tire pressure or a change in tire temperature.

The IMU and odometry quantities potentially affected by SGE are already implemented as an error model in the fusion filter, and in each measurement epoch, the raw measurements are corrected by the known, continuously reestimated errors. Thus, the slow growth of these errors does not lead to significant errors in the fused data, so long as the fusion filter corrects these quickly enough using redundant measurements. On the other hand, problematic disturbances to correction measurements – because they go far beyond the filter dynamics – can be avoided with defined detection probability and threshold by means of the plausibility check described in Sect. 4.2 and can be identified using the hypothesis test outlined in Eq. 22. Moreover, by checking the totaled absolute values of the error corrections with defined maximum values, it is possible to detect errors outside the range specified for the respective sensor.

Irrespective of the fusion filter, slowly growing errors of individual pseudorange measurements, which are not modeled as errors in the fusion filter, lead to contradictions in the geometric comparison of the plausibility check described in Eqs. 11, 12, 13, 14, 15, 16, and 17 and can therefore be identified with a defined detection threshold.

6.4 Accuracy Measure

In order to show an example implementation of an accuracy measure in the fusion filter that fulfills the criteria outlined in Sect. 5.2, we have selected the descriptors: measurement noise, offset, and scale factor error. In the application outlined here, it is also assumed that the descriptors are normally distributed; this simplifies a joint application with the stochastic model of the fusion filter. The general variance propagation law with fully occupied variance-covariance matrix should be used for correlated descriptors, but for uncorrelated descriptors, this is simplified to scalar propagation of the variances.

The method is implemented for the basic system operations outlined in Fig. 4: namely, the correction of offsets and scale factor errors of an acceleration measurement $\vec{a}_b^{IMU}$ by the fusion filter, its rotation in navigation coordinates using a rotation matrix $\hat{\mathbf{C}}_{b_r}^n$, and their summation to yield velocity $\vec{v}_n$ with simultaneous correction of its absolute value by the fusion filter.

For simplicity, let us assume that the errors of the rotation matrix $\hat{\mathbf{C}}_{b_r}^n$ are negligible: For the computation $\mathbf{P}$ generally represents the variance-covariance matrix of the respective descriptors, while $\mathbf{P}_k^+$ is the a posteriori variance-covariance matrix of the fusion filter in the current measurement epoch. A double arrow over a variable $\overset{\rightrightarrows}{u}$ means that the vector $\vec{u}$ is used as a main diagonal in a quadratic matrix otherwise filled with zeros.

6.4.1 Data Source

At the start of the accuracy calculation, the acceleration measurements are given as data: $\vec{a}_b^{IMU}$ is the measurement vector of acceleration, $\overset{\rightrightarrows}{\varsigma a}_b$ the scale factor error estimated by the filter in the main diagonal form, $\overrightarrow{\delta a}_b$ the offset error estimated by the filter, $\mathbf{P}_{ra}^{in}$ the variance-covariance matrix of measurement noise, $\mathbf{P}_{offs}^{in}$ the variance-covariance matrix of the offset error, and $\mathbf{P}_{scale}^{in}$ the variance-covariance matrix of the scale factor error. The variance-covariance values are modeled from characteristic values in the sensor datasheet and by modeling known physical relationships.

6.4.2 Correction Step

The acceleration measurement is corrected for offset errors and scale factor errors by

$$\vec{a}_b^{Corr} = \left(\left(\mathbf{I} - \underset{b}{\overset{\rightrightarrows}{\varsigma a}}\right) \cdot \vec{a}_b^{IMU}\right) - \overrightarrow{\delta a}_b \tag{29}$$

The corresponding output values $\mathbf{P}^{out}$ are obtained for $\mathbf{p}_{ra}^{in}$ by variance propagation:

$$\mathbf{P}_{ra}^{out} = \left(\mathbf{I} - \underset{b}{\overset{\rightrightarrows}{\varsigma a}}\right) \cdot \mathbf{P}_{ra}^{in} \cdot \left(\mathbf{I} - \underset{b}{\overset{\rightrightarrows}{\varsigma a}}\right)^T \tag{30}$$

Conversely, $\mathbf{P}_{offs,a}^{+}$ and $\mathbf{P}_{\mathrm{scale,a}}^{+}$ are overwritten by the corresponding variances of the fusion filter, to keep them consistent with the correction of the offset error and scale factor error in the basic system.

6.4.3 Transformation Step (Rotation)

The output values of the correction are now transferred to a different coordinate system by rotation using a rotation matrix $\hat{\mathbf{C}}_{b_r}^{n}$ whereby, as mentioned above, $\hat{\mathbf{C}}_{b_r}^{n}$ is assumed to be error-free. The transformation equation of rotation is

$$\vec{a}_n = \hat{\mathbf{C}}_{b_r}^{n} \cdot \vec{a}_b^{Corr} \tag{31}$$

All variance-covariance matrices $\mathbf{P}_{ra}^{out}$, $\mathbf{P}_{offs}^{out}$, and $\mathbf{P}_{scale}^{out}$ follow by application of the variance propagation law:

$$\mathbf{P}^{out} = \hat{\mathbf{C}}_{b_r}^{n} \cdot \mathbf{P}^{in} \cdot \hat{\mathbf{C}}_{b_r}^{n^T} \tag{32}$$

6.4.4 Integration Step

Summation of the accelerations to give velocity $\vec{v}_n$ is done with the sampling interval Δt that is assumed to be error-free. Here $\vec{v}_{n_r}$ is the value of the velocity at the last sampling step:

$$\vec{v}_n = \vec{v}_{n_r} + \vec{a}_n \cdot \Delta t + \hat{\mathbf{C}}_{b_r}^{n} \cdot \Delta \vec{v}_b \tag{33}$$

The calculation of the variances is based on the simplified model assumption that, if the measurement errors are symmetrically distributed around zero, a scale factor error does not cause a shift of the mean value. Gaussian error propagation yields $\mathbf{P}_{ra}^{out}$ and $\mathbf{P}_{scale}^{out}$ with the calculation in accordance with

$$\mathbf{P}^{out} = \mathbf{P}^{in} \cdot \Delta t^2 \tag{34}$$

On the other hand, the offset error of velocity is corrected by the fusion filter, and therefore, $\mathbf{P}_{offs}^{out}$ is overwritten by the corresponding variance of the fusion filter.

6.4.5 Total Uncertainty

The structure of the accuracy computation is such that, after each encapsulated step, it is possible to access the processed measurements and also their virtual datasheet and make these data available to the user functions. The block structure can be altered in a modular way to give a cross-system architecture and the option of expansion to include subsequent user functions. The total uncertainty of a descriptor can be calculated from the individual uncertainties, in the illustrated example of normally distributed values, by summation of the variances. Although each of the descriptors relates to different signal properties, they are available in the same unit. The total uncertainty of the data can therefore be calculated by summation of the individual uncertainties:

$$\mathbf{P}_{tot}^{out} = \mathbf{P}_{ra}^{out} + \mathbf{P}_{offs}^{out} + \mathbf{P}_{scale}^{out} \tag{35}$$

6.5 Example Results

In order to assess the performance of the fusion filter, the difference w.r.t. a reference measurement system is calculated. Within the specified uncertainty limits, this difference is used as a measurement for the error of the filter. The following measures are selected for this:

- Standard deviation σ as a measure for the noise of the measured data
- Mean μ as a measure for the average error of the measured data
- Median ε_{Q50} as a measure for the average error of the measured data cleaned of outliers
- Root-mean-square error ε_{RMS} as a measure for the total error of a measurement
- Maximum error ε_{Max} as a measure for maximum error and outlier size

The measures used for verification of the fusion filter are selected according to the following criteria:

- They are quantities for which the fusion filter has estimated an error; the associated variance is therefore available and verifiable in the matrix $\mathbf{P}_{\kappa}^{+}$.
- They are directly affected by corrections due to observations and not only via the system model; the effects of disturbed correction measurements are therefore clearly identifiable.
- They are involved in the plausibility check described in Sect. 4.2 as quantities for determining the threshold ξ, so that the effects of and upon the plausibility check are identifiable.
- Via the system model they are dependent upon as many other estimated quantities corrected in the filter as possible, so that they also represent the errors and uncertainties of these quantities and therefore allow a statement to be made about overall performance.

These characteristics relate to the variables:

- Velocity $\hat{\vec{v}}_b$
- Position $\hat{\vec{\Phi}}_n$

in the fusion filter. Because of its particular relevance to automotive applications, the following discussion is restricted to horizontal plane quantities. They contain totaled IMU measurements and corrections from the fusion filter. Here it is assumed that errors in all other estimated quantities over the duration of a test result in identifiable velocity and position errors.

By way of an example, the results from a real test drive (total drive **G**; distance, 15.7 km; duration, 1,000 s; extract from representative route with overland and urban segments) are compared with the following legs with different characteristics, and the statistic measures are given in Table 3:

Table 3 Results of the fusion filter

Drive/leg	Horizontal position difference					Horizontal speed difference				
	$\frac{\sigma}{\mathrm{m}}$	$\frac{\mu}{\mathrm{m}}$	$\frac{\varepsilon_{Q50}}{\mathrm{m}}$	$\frac{\varepsilon_{RMS}}{\mathrm{m}}$	$\frac{\varepsilon_{Max}}{\mathrm{m}}$	$\frac{\sigma}{\mathrm{m}/\mathrm{s}}$	$\frac{\mu}{\mathrm{m}/\mathrm{s}}$	$\frac{\varepsilon_{Q50}}{\mathrm{m}/\mathrm{s}}$	$\frac{\varepsilon_{RMS}}{\mathrm{m}/\mathrm{s}}$	$\frac{\varepsilon_{Max}}{\mathrm{m}/\mathrm{s}}$
G	2.89	4.45	3.85	5.31	17.0	0.51	0.28	0.15	0.58	4.72
A	2.27	3.89	3.87	4.5	16.77	0.12	0.19	0.17	0.23	0.7
B	2.42	4.57	4.44	5.17	14.66	1.29	1.09	0.34	1.69	4.72
C	2.5	11.81	13.29	12.07	13.68	0.15	0.1	0.02	0.17	0.71

- Leg **A** (distance, 5.1 km; duration, 250 s; mostly overland, slight disruptions)
- Leg **B** (distance, 1.6 km; duration, 100 s; urban area, significant disruptions of the algorithm)
- Drive **C** (distance, 0.6 km; duration, 80 s; driving through a tunnel with subsequent standstill, route outside the total journey)

The measures demonstrate the dependence of accuracy, both of absolute position and also absolute velocity, upon the environmental conditions of GPS reception. Table 4 gives the results of the integrity assessment for the said experiments, given as the number of integrity alarms. In addition, the plausibility check for the correction measurements is selectively deactivated so as to illustrate its influence upon the results. The results clearly show the need to include a plausibility check of the correction measurements preceding the data fusion. It only makes sense to perform an integrity calculation based on plausibility-checked data.

The example results outlined above were obtained with parameterization of the fusion filter which were primarily optimized to position accuracy and determined under typical everyday boundary conditions. This includes that odometry measurements are almost permanently available and are noisy according to normal road characteristics and GPS is affected by the normal disturbances for overland and urban journeys. Unavailability of GPS is only short term, for example, when driving through a city tunnel.

The parameterization of the fusion filter is always a compromise between different, possibly conflicting, requirements.

In general, the following steps are recommended for optimizing parameterization:

1. Establishing the optimization goals, e.g., optimal position and velocity accuracy, or a compromise between several estimated accuracies
2. Establishing test cases and the desired behavior required for them
3. Parameterizing the stochastic model
4. Parameterizing the plausibility checking thresholds
5. Parameterizing the stochastic parameters of the integrity measure

In particular, it is necessary to ensure that the plausibility check removes no or very few measurements in slightly disturbed scenarios and that, in badly disturbed scenarios, a compromise is found between test power, speed of compensation of undetected errors, and the stability of the fusion filter.

Table 4 Results of the integrity assessment

Algorithm/ drive	Complete plausibility check	GPS plausibility check inactive	Odometer plausibility check inactive	Plausibility check of GPS and odometry inactive
G	1	8	0	5
A	0	0	0	0
B	1	7	0	4
C	1	35	1	55

Before practical application, it is necessary to perform investigations and falsification tests, in particular for the assumptions made for the integrity measure. This can be done, for example, by a software-in-the-loop test for error detection as described in Sect. 4.3, whereby the reaction to several simultaneous errors and the detection of slowly growing IMU, odometry, and GNSS errors are tested under controlled, reproducible conditions.

It should be noted that, when using an Error-State-Space filter, the influences of the fusion filter on the output data (full states) are small and that the sensor noise also corresponds to the noise of the output acceleration and angular rates. The smaller the drift of the estimated IMU error, the smaller the influence of the fusion filter upon the result. This is a basic consideration in selecting sensors for a practical fusion filter application.

7 Outlook and Conclusion

The data fusion described in this section, and illustrated using an example, essentially uses sensors that are already installed in current motor vehicles. Only the access to raw GNSS data that is required in the Tightly Coupling method is still unusual.

In the classification proposed in Ch. ID#0500, the filter is classed as fusion of raw data. The stochastic model is based on pure variance propagation and does not require any classification uncertainty or object hypotheses. The same applies for the enhancements to include plausibility checking and compensation of delayed availability, which represent an extension of the filter based on physical models that use the same assumptions – only slowly changing, normally distributed errors.

At the time of publication of this chapter, a comparable fusion filter going under the name of M2XPro (**M**otion information **2 X Pro**vider) is being developed by Continental Teves AG & Co. oHG (Continental 2014).

A number of existing sensors and technologies offer potential for further improving data fusion for localization in the future:

- Dual-frequency GNSS receivers offer the option of eliminating ionospheric delays, which is a big contributor to absolute position errors (Mansfeld 2010).

- Multi-GNSS receivers, which not only perform GPS measurements but also GLONASS, Galileo, etc., measurements, provide improved availability of navigation satellites – especially under constrained reception conditions (Mansfeld 2010).
- Deep integration GNSS receivers with feedback of dynamic quantities from the fusion filter into the control loop inside the receiver help to improve satellite tracking and to reduce receiver noise.
- It is possible to support multidimensional vehicle attitude, possibly even at standstill, by using several GNSS antennae – connected to one or more receivers.
- Vehicle attitude at standstill can be estimated via the Earth's magnetic field using magnetometer/compass sensors.
- Barometer measurements can be used to support the altitude component of the position by measuring air pressure.
- Absolute localization can be improved by using digital maps and/or known or learned landmarks (Lategahn 2013).
- (Stereo) cameras offer the possibility of slip-free velocity and angular rate measurement relative to the environment (Dang et al. 2005).
- RADAR and LIDAR sensors (Winner at al. 2009) support the:
  - Estimation of speed relative to other vehicles
  - Estimation of speed relative to fixed objects
  - Feedback of fused data to support sensor object hypotheses
  - Coupling with grid mapping algorithms, possibly also on the stochastic level (Grewe et al. 2012)
- Spring travel sensors on the wheel suspensions allow estimation of yaw and pitch angle of the vehicle, irrespective of the slope of the vehicle's tire contact patch.

Compared with the current state of the art, the carrier platform model from Sect. 6.2 for moving reference systems represents a particular development for the automotive sector. As regards correction of the symptoms that occur as the result of reconciliation of sensors on a moving carrier, so far there have only been a few isolated solutions in certain applications – such as ESC or ACC.

The system architecture presented here has been kept general and allows the modular integration of other components without the need to make modifications to existing integration concepts. Equally, the filter is transferable to different sensor environments so that it can also be used for vehicles other than road vehicles, for example, in shipping and aerospace.

References

Abdelghani M, Friswell M (2001) A parity space approach to sensor validation. IMAC XIX – 19th International Modal Analysis Conference, Orlando, USA

Abuhashim T, Abdel-Hafez M, Al-Jarrah M (2010) Building a robust integrity monitoring algorithm for a low cost GPS-aided-INS system. Int J Control Autom Syst 8:1108–1122, Bucheon, Korea

Agogino A, Chao S, Goebel K, Alag S, Cammon B, Wang J (1988) Intelligent diagnosis based on validated and fused data for reliability and safety enhancement of automated vehicles in an IVHS. Institute of Transportation Studies, University of California, Berkeley, California PATH Research Report UCB-ITS-PRR-98-17

Bar-Shalom Y, Rong Li X, Kirubarajan T (2001) Estimation with applications to tracking and navigation. Wiley, New York

Bevley D, Cobb S (2010) GNSS for vehicle control. Artech House, Norwood

Bhatti U (2007) Improved integrity algorithms for integrated GPS/INS systems in the presence of slowly growing errors. Dissertation, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London

Bhatti U, Ochieng W (2009) Detecting multiple failures in GPS/INS integrated system: a novel architecture for integrity monitoring. J Glob Positioning Syst 8(1):26–42, Calgary, Alberta, Canada

Brown R (1992) A baseline RAIM scheme and a note on the equivalence of three RAIM methods. Proceedings of the 1992 National Technical Meeting of The Institute of Navigation, San Diego, p. 127–137

Continental (2014) homepage (M2xPro): http://www.conti-online.com/www/automotive_de_de/themes/passenger_cars/chassis_safety/passive_safety_sensorics/sensor_sensorsystem_de/m2xpro_de.html (English version: http://www.conti-online.com/www/automotive_de_en/themes/passenger_cars/chassis_safety/passive_safety_sensorics/sensor_sensor_system_en/m2xpro_en.html), Accessed Feb 2015

Dang T, Hoffmann C, Stiller C (2005) Visuelle mobile Wahrnehmung durch Fusion von Disparität und Verschiebung (Visual mobile perception by fusion of disparity and displacement). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung (Driver assistance systems with machine perception). Springer, Berlin

Deutsches Institut für Normung (1994) DIN70000 – Straßenfahrzeuge; Fahrzeugdynamik und Fahrverhalten (German Institute for Standardization (1994) DIN70000 – Road vehicles, vehicle dynamics and driving behaviour). Beuth Verlag, Berlin, Section 1.2

Diesel J, Luu S (1995) Calculation of thresholds and protection radius using chi-square methods. Proceedings of the 8th International Technical Meeting of the Satellite Division of The Institute of Navigation, Palm Springs

Ding J, Wesley Hines J, Rasmussen B (2004) Independent component analysis for redundant sensor validation. The University of Tennessee, Nuclear Engineering Department, Knoxville

Dziubek N (2013) Zeitkorrigiertes Sensorsystem (Time-corrected sensor system). Continental Teves AG & Co. oHG, Patent WO 2013/037850 A1

Dziubek N (2013) Verfahren zum Auswählen eines Satelliten (Method for selecting a satellite). Continental Teves AG & Co. oHG, Patent WO 2013/037844 A3

Dziubek N, Martin J (2013) Sensorsystem umfassend ein Fusionsfilter zur gemeinsamen Signalverarbeitung (Sensor system comprising a fusion filter for common signal processing). Continental Teves AG & Co. oHG, Patent WO 2013/037854 A1

Feng S, Ochieng W (2007) Integrity of navigation system for road transport. Centre for Transport Studies, Imperial College London, London

Goebel K, Agogino A (1999) Fuzzy sensor fusion for gas turbine power plants. Proceedings of SPIE, Sensor Fusion: Architecture, Algorithms, and Applications III, Vol. 3719, Orlando, pp 52–61

Grewe R, Hohm A, Hegemann S, Lueke S, Winner H (2012) Towards a generic and efficient environment model for ADAS. Intelligent Vehicles Symposium, Alcalá de Henares

Gross-Bölting M, Kolkmann D (2002) Verfahren zum Abgleich eines Systems zum Messen der Gierrate eines Kraftfahrzeuges sowie ein solches System (Compensation method for a vehicle yaw rate measuring system as well as said system). Hella KGaA Hueck & Co., Patent EP 1 264 749 B1

Groves P (2008) Principles of GNSS, inertial and multisensor integrated navigation systems. Artech House, Boston, Sections 12.1.3, 12.1.5

Hackenberg U, Heißing B (1982) Die fahrdynamischen Leistungen des Fahrer-Fahrzeug-Systems im Straßenverkehr (Driving-dynamic performances of the vehicle-driver-system in public road traffic) Automobiltechnische Zeitung (ATZ), Nr. 84, Wiesbaden, Table 1

Hofmann-Wellenhof B, Legat K, Wieser M (2003) Principles of positioning and guidance. Springer, Vienna, Section 9.3.1

Hwang P, Brown R (2005) NIORAIM integrity monitoring performance in simultaneous two-fault satellite scenarios. ION GNSS 18th International Meeting of the Satellite Division, Long Beach

Keller F, Lüder J, Urban W, Winner H (2002) Verfahren und Vorrichtung für die Bestimmung von Offsetwerten durch ein Histogrammverfahren (Method and apparatus for determining offset-values using a histogram). Robert Bosch GmbH, Patent DE 000010205971 A1

Keller F, Lüder J, Urban W, Winner H (2002) Verfahren und Vorrichtung für die Bestimmung von Offsetwerten durch ein Regressionsverfahren (Method and device for determining an offset value using a regression method). Robert Bosch GmbH, Patent DE 000010206016 A1

Kuusniemi H (2005) User-Level Reliability and Quality Monitoring in Satellite-Based Personal Navigation. Dissertation, Tampere University of Technology, Tampere

Lategahn H (2013) Mapping and Localization in Urban Environments Using Cameras. Dissertation, Institut für Mess- und Regelungstechnik, Karlsruher Institut für Technologie (KIT), Karlsruhe

Le Marchand O, Bonnifait P, Ibañez-Guzmán J, Bétaille D (2009) Vehicle localization integrity based on trajectory monitoring. Intelligent Robots and Systems, St. Louis

Leinen S (2009) Parameterschätzung I Vorlesungsskript (Parameter estimation I lecture notes). Institut für Physikalische Geodäsie der TU Darmstadt (now: Fachgebiet Physikalische Geodäsie und Satellitengeodäsie – Institute of Physical and Satellite Geodesy), Darmstadt, Section 5.1.2

Mansfeld W (2010) Satellitenortung und Navigation (Satellite-based localization and navigation), 3rd edn. Vieweg + Teubner, Wiesbaden

Niebuhr J, Lindner G (2002) Physikalische Messtechnik mit Sensoren (Physical measurement technology with sensors), 5th edn. Oldenbourg Industrieverlag, Munich, Section 1.2.2

Pacejka H, Bakker E (1993) The magic formula tyre model. Tyre models for vehicle dynamics analysis. Delft, The Netherlands

Pourret O, Naim P, Marcot B (2008) Bayesian networks: a practical guide to applications. Wiley, West Sussex

Pullen S (2008) Quantifying the performance of navigation systems and standards for assisted-GNSS. InsideGNSS, Oregon, Sept/Oct 2008

Reif K (2007) Automobilelektronik (Automotive electronics), 2nd edn. Vieweg Verlag, Wiesbaden, Sections 4.6.4, 4.6.5, 4.6.8

Steinhardt N (2014) Eine Architektur zur Schätzung kinematischer Fahrzeuggrößen mit integrierter Qualitätsbewertung durch Sensordatenfusion (An Architecture for the Estimation of Kinematic Vehicle Properties with Integrated Quality Assessment by Sensor Data Fusion). Thesis submitted to Fachbereich Maschinenbau, Technischen Universität Darmstadt, 01.04.2014

Strang T, Schubert F, Thölert S, Oberweis R et al (2008) Lokalisierungsverfahren (Localization methods). Deutsches Zentrum für Luft- und Raumfahrt e. V. (DLR). Institut für Kommunikation und Navigation, Oberpfaffenhofen

Sukkarieh S, Nebot E, Durrant-Whyte H (1999) A high integrity IMU/GPS navigation loop for autonomous land vehicle applications. IEEE Trans Robot Autom 15(3):572–578, New York

Thrun S, Burgard W, Fox D (2006) Probabilistic robotics. The MIT Press, Massachusetts Institute of Technology, Cambridge

Titterton D, Weston J (2004) Strapdown inertial navigation technology, 2nd edn. The Institution of Electrical Engineers, Stevenage, Section 3.5

Toledo-Moreo R, Zamora-Izquierdo M, Úbeda-Miñarro B, Gómez-Skarmeta A (2007) High-integrity IMM-EKF-based road vehicle navigation with low-cost GPS/SBAS/INS. IEEE Trans Intell Transp Syst 8(3):491–511, New York

Wegener M, Schnieder E (2013) Definition der Messqualität und ihre quantitative Bestimmung (Definition of measurement quality and its quantitative determination). Deutsche Gesellschaft für Ortung und Navigation (DGON), POSNAV ITS, Berlin

Wendel J (2007) Integrierte Navigationssysteme. Sensordatenfusion, GPS und Inertiale Navigation (Integrated navigation systems. Sensor data fusion, GPS and inertial navigation). Oldenbourg Wissenschaftsverlag, Munich

Winner H et al (2009) Handbuch Fahrerassistenzsysteme (Handbook of Driver Assistance Systems). Vieweg + Teubner, Wiesbaden, Sections 12 and 13

Wu X, Wu S, Wang J (2008) A fast integrity algorithm for the ultra-tight coupled GPS/INS System. 9th International Conference on Signal Processing, Beijing

Young R, McGraw G (2003) Fault detection and exclusion using normalized solution separation and residual monitoring methods. NAVIGATION. J Inst Navig 50(3):151–170, Manassas, USA

Digital Maps for ADAS

26

Thomas Kleine-Besten, Ralph Behrens, Werner Pöchmüller,
and Andreas Engelsberg

Contents

T. Kleine-Besten (✉) • R. Behrens • W. Pöchmüller
Robert Bosch Car Multimedia GmbH, Hildesheim, Germany
e-mail: thomas.kleine-besten@de.bosch.com; ralph.behrens@de.bosch.com

A. Engelsberg
Bosch SoftTec GmbH, Hildesheim, Germany
e-mail: Andreas.Engelsberg@bosch-softtec.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_27

Abstract

Navigation software cannot work without a digitized map. The question is, what is a digital map exactly? What kind of data is part of a digital map? How is the data structured? This chapter gives an overview of the digital map in the format of NDS. The documentation is based on the NDS specification and looks first to the navigation database from feature perspective with a more and more deeper look into the building blocks of the database. A short outlook into the future of the Navigation Data Standard will finish this chapter.

Digital map data for navigation systems is not generated automatically when the data is collected. There are firms that specialize in geodata acquisition, such as HERE and TomTom. Besides such commercial organizations, official institutions like the Land Registry Offices in Germany likewise collect geodata. Last but not least, geodata is also collected by countless volunteers and is made freely available to everyone – OpenStreetMap plays an important role in this regard. To some extent, the geodata in OpenStreetMap is more up to date and accurate than the commercial data. Scientific studies have been carried out – such as by the Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS – in which OpenStreetMap data has been compared with other map data (Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS 2011).

The data collected during geodata acquisition is not used by navigation systems directly. It is first compiled before use, that is to say, the data is converted to a format that navigation systems can process. This conversion involves reduction of the data but also enrichment of the geodata with additional information. In the past, each navigation system manufacturer developed its own data format; a reutilization of the compiled navigation data did not in effect take place.

Standardization of the map data for navigation systems has been an important issue in the automotive industry for many years now, particularly since the costs of producing navigation-compatible data are very high. Already in 2004, a group of companies formed an alliance to drive forward map data standardization. A registered association called Navigation Data Standard (NDS) e.V. subsequently emerged from this initiative in 2009. NDS e.V. has set itself the goal of standardizing map data for navigation systems: the standardization process includes defining the requirements, specifications, and official versions of the NDS map format. The licenses are available to the association members free of charge (Müller 2010).

The association comprises vehicle manufacturers, automotive industry suppliers, map data suppliers, and telematics service providers. As of March 2014, 25 firms are listed as members (NDS Association 2014a). The firms in the consortium are based in Europe, USA, and Asia. This distribution of members shows they are working on global standardization.

1 Standardization Objectives

The requirement for a map data standard arose almost of its own accord due to increasing demand around the world for vehicles designed to be factory fitted with a navigation system as standard. Initially, numerous local navigation solutions emerged, but the development and integration of regional solutions is time consuming and costly.

Economic reasons and the expectation that navigation data should always be as up to date as possible have resulted in the following NDS standardization objectives (NDS Association 2014b):

1. Development of a globally accepted map format
   To date, there is still no universally applicable standard for navigation-compatible map data; many navigation solution manufacturers utilize their own proprietary map format. The aim of standardization is to create a navigation map format that can be used throughout the world.
2. Separation of the application and data
   When updating the navigation data, many navigation app developers at the same time install a new version of their software. The NDS map format aims at clear separation in this regard: that is to say, the ability to update the navigation data independently of the navigation application.
3. Data compatibility and interoperability
   It is currently not possible for navigation data to be exchanged between two vehicle manufacturers. Often, navigation data cannot even be exchanged between two different model ranges made by the same vehicle manufacturer. Standardization aims at making the data backward compatible and interchangeable. This will mean a vehicle manufacturer will then only need one NDS storage device that is supported by several different navigation applications. This simplifies the logistics and makes handling various navigation systems easier for users and auto repair shops.
4. Specification of a data update method
   Currently, navigation system manufacturers each have their own method for performing data updates; this leads to uncertainty and frustration among users, since every device behaves differently. The consortium aims to clearly regulate the procedure used for updating navigation data.
5. Data compactness and application efficiency
   Navigation systems that have been ordered and supplied as original equipment with a new vehicle must be able to outlast a service life of around 10 years. This means the systems must not only be highly robust but must also be capable of working with modified data and varying volumes of data. Standardization aims at achieving a data format design that ensures the navigation system application will function at an acceptable level of performance even after multiple updates.

6. Independence of the data format from the storage and transmission medium
   This objective defines the concept that the map data format must also be capable of being used with various media. It must be possible for the data to be not only transferred by means of storage devices but also via air interfaces.
7. Expandability of the navigation data
   The map format must support the expansion and integration of proprietary content. Every vehicle manufacturer wants to ensure its vehicles sell particularly well – one way of achieving this is by providing special functions. This objective therefore seeks to ensure that a manufacturer-specific expansion of the data is possible without undermining the goal of compatibility and interoperability.
8. Support for copy protection
   The map format must protect the data against piracy and misuse; the standard must therefore support digital rights management (DRM).

2 Features of the NDS Standard

A feature of the NDS map format is the organization of the data into so-called building blocks. The term “building block” is used because it is comparable to a “brick” that possesses standardized interfaces. Each “brick” can be replaced individually, and when doing so, the option exists to vary its “color” and “size,” so to speak. What this means in real terms is that NDS supports the option to perform incremental updates and that extended features can at the same time be integrated in a building block. Such extensions can also be used to implement customizations of the data according to customer requirements.

The data of the individual building blocks are linked to one another by means of particular attributes, which must follow a standardized format across all building blocks. The most important standardized attributes are coordinates and names. All the various building blocks of the NDS format support, for instance, coordinates for georeferencing; this therefore makes it possible to search for information in all parts of the database with this coordinate as search parameter.

The preferred data storage method is a database based on SQLite. The database provides an efficient basis for incremental updates and for searching within the various building blocks by means of standardized parameters.

Right at the outset of the standardization activities, attention was paid to providing extension options for new functions. Of particular note here are the map display extensions for 3D objects and extended 3D city models as well as the addition of a special ADAS building block with a flexible method for optimally supporting advanced driver assistance systems.

In addition to this, the standardization efforts also include the development of an instrument for verifying the quality of the database.

3 Growth in Data Volume Due to New Features

The data required to display maps in 3D must be evaluated critically from different viewpoints: on the one hand, the visual portrayal is the distinguishing stylistic element that is experienced by the viewer immediately and directly; on the other hand, in order to provide the user with a more detailed portrayal, more content needs to be added to the database, which in turn must be loaded and rendered quickly.

Below are several items of data embodied in the NDS standard:

- Aerial photographs and satellite images
- Digital terrain model
- High-resolution images for visualizing intersections and highway exits
- 3D city model
- Data for travel guide apps
- Supplementary data for ADAS (advanced driver assistance systems)

4 Structure of the Data Within an NDS Database

To start with, the data in the NDS database is divided into "regions" (see Fig. 1). Within the regions, the data is organized in "components," which in turn contain the "building blocks." The building blocks contain the smallest item known as "tiles."

This structure can be illustrated by means of the following example: An NDS database contains the data for a particular market (database coverage), e.g., Europe. This market is divided into regions, such as France, Spain and Portugal, Benelux, Germany, etc. These regions can then be updated independently of one another. The way these regions are divided up is not standardized but is redefined for each project.

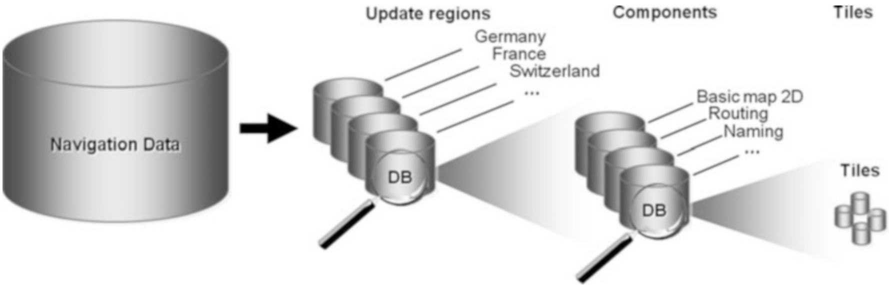


Fig. 1 Definition of the update regions (Used with the kind permission of the NDS Association)

5 NDS Building Blocks

This section presents the various building blocks of the NDS database (see Fig. 2) (NDS Association 2013) as well as their most important constituent parts and functions.

5.1 Overall Building Block

The overall building block is used for storing the data that is common to all building blocks. Every NDS database must contain an overall building block. The contents serve to provide the application with information about the variable elements and attributes of the specific database. This is essentially metadata and regionally specific information.

5.2 Routing Building Block

The routing building block contains the road network. It serves as a basis for various applications:

- Route calculation
  The route calculation function reads the topology of the road network and its attributes. These attributes include such things as the length of a road segment, the permitted speed for a road segment, and the classification of a road segment (freeway, highway, country road, etc.). With the help of the attributes, a route with the desired parameters (e.g., "short route," "fast route," "avoid ferries," etc.) is calculated based on various cost functions.
- Map matching
  Position fixing comprises several stages. First of all, the information from various sensors (i.e., GPS, gyroscope, wheel speed sensors, and accelerometers) is used in a certain method known as "dead reckoning" to calculate the precise position.

  In the subsequent map-matching procedure, the road geometry from the routing building block is used to depict the position obtained through dead reckoning on the road network.
- Route guidance
  The route guidance function uses the topology and geometry of the road network to compare the current position with the road network and to work out instructions for the driving maneuvers.
- ADAS
  With the specific purpose of providing support for advanced driver assistance systems, a layer was added to the routing building block to hold details like curvature and road width.

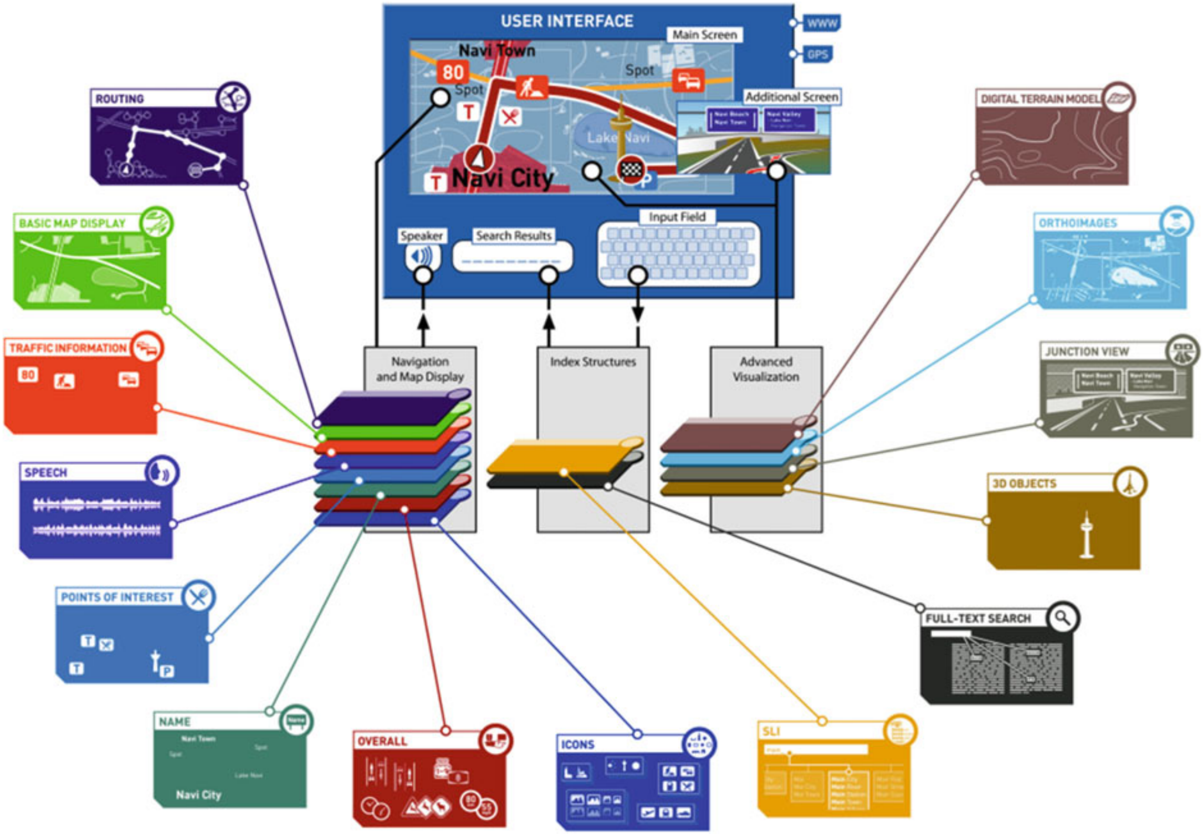


Fig. 2 Overview of NDS building blocks (Used with the kind permission of the NDS Association)

5.3 SQLite Index (SLI)

The SLI building block is used for inputting destinations. The destination input function is nowadays equipped with a wide variety of options. Besides the conventional hierarchical destination input method, in which the user enters one letter after another, the SLI also supports complex destination input options. This includes, among other things, First Letter Input (FLI), which is a mode specially designed for the Asian market; in FLI mode, users only need to enter the first letter of a syllable. A further input mode called One-Shot-Destination-Entry allows users to enter the entire address in one go before initiating the search – similar to using an Internet search engine.

The major advantage with SLI-based destination input is that the user can perform the search not only based on a single criterion (such as city, street, zip code, or house number) but based on several criteria simultaneously.

5.4 POI Building Block

Points of interest (POIs) are distinctive places or ones that may be of interest to the user; they are displayed on the navigation map, and the user must be able to select them as navigation destinations (Wikipedia 2014). Besides location information, the POIs can contain additional information, such as telephone numbers and opening times.

The POIs can be presented and used in various ways:

- POIs can be displayed as an icon on the navigation map.
- POIs can be displayed in list form, sorted according to their category, location, or other criteria.
- Additional information can be displayed when a POI is selected, for instance, pictures, descriptive texts, payment information, etc.

Besides the point-based POIs, there are also line and area-related POIs: A line POI could, for instance, be a road of particular interest to tourists, and an area-related POI could be a large zoo.

A further differentiating factor is the period of validity of the POIs. Some POIs are only valid for a limited time: for example, sports stadia may be called by a different name or may serve a different purpose during an Olympics or a soccer World Cup. For this reason, a distinction must also be made in terms of the update frequency of the various POI types.

There are two building blocks pertaining to the POIs:

- Integrated POI Building Block
  The integrated POI building block may contain direct references to other NDS building blocks; such references point to route segments or intersections in the

routing building block, or they point to names in the naming building block. These direct references serve to improve search performance.

- Nonintegrated POI Building Block
  The nonintegrated POI building block is interlinked with the other building blocks solely via the geocoordinates. This means it is only possible to use geocoordinates to search and ascertain whether POI information belongs to a specific map section or not.

Both POI building blocks have the same internal data structure.

5.5 Naming Building Block

The naming building block contains the names of the roads in the road network out of the routing building block and the names out of the basic map display building block. This ensures that the names in a route list and on a map display are used identically.

The POI names and traffic information names are deliberately not a constituent of the naming building block. This is due to the need to update this specific data more frequently. The corresponding names are therefore a constituent of the POI building block and traffic building block respectively.

5.6 Full Text Search Building Block

In order to implement a full text search, the content must first have been indexed in advance. The POI building block, the naming building block, and all other building blocks that contain names serve as the source material for generating the index. An index is thus built using this information. This index is then filtered with the search string; the subsequent result of the search is a list of links to the files that contain one or more of the words in the search string and that point directly to the source data. With the help of this information, it is possible to implement an input function that is comparable to the metasearches of well-known Internet platforms.

5.7 Phonetic/Speech Building Block

The speech building block contains two different types of data.

The first part contains a database of phonetic transcriptions that is used for voice recognition and for text-to-speech conversion. It is important that the database is compatible with the phonetic transcription of the languages that are available on the navigation system. The phonetic transcription to be used must be configured for the correct language before use. This is no trivial task, since, particularly in multicultural regions, there are situations in which two languages are used in parallel.

The second part contains a database of professionally recorded voice output phrases – during navigation, these are pieced together for route guidance and are transferred to the audio output at the right moment.

5.8 Traffic Information Building Block

The traffic information building block (TI) enables the navigation applications to use traffic data of various standards. The location tables for the Traffic Message Channel (TMC) are stored in the TI; they allow the traffic reports to be assigned to the various languages. With regard to traffic reports that are compliant with the Transport Protocol Expert Group (TPEG), the traffic information building block provides the location tables so that the event information (e.g., congestion, disruption, road closure, etc.) of a TPEG message can be interpreted. The event information is transmitted as a number; with the help of this number, it is then possible to load the corresponding text in the respective language from the location table.

5.9 Basic Map Display Building Block

The basic map display building block (BMD) is used to group the essential data for a map display; the basic map display building block obtains the names for the display from the naming building block.

The content of the BMD is points (e.g., town centers, mountain peaks); lines (e.g., roads, waterways, boundary lines, railroad lines); areas (e.g., forests, bodies of water, building footprints, land use); icons, drawing styles, and drawing rules (e.g., details about at which points roads, etc. can be labeled).

Additionally, the data for displaying the 2.5D city models is already contained in the BMD. Every building is defined by various parameters, such as its height, the color of the exterior walls, and the type and color of the roof.

The data contained in the BMD is sufficient to ensure that a complete view of the navigation map can be provided in 2D and 2.5D mode.

5.10 Advanced Map Display

The term advanced map display (AMD) refers to the extended content required for the map display in order to portray the map as realistically as possible.

The extended content of the AMD comprises the following building blocks:

- Digital terrain model building block
- Orthoimages building block
- 3D objects building block

5.10.1 Digital Terrain Model Building Block

This building block contains the elevation model, i.e., the topography of the Earth's surface (digital terrain model – DTM). Besides the elevation data for each area tile, it also contains the textures for displaying the Earth's surface in cases in which no satellite or aerial images are available. To ensure the Earth's surface can be displayed without discontinuous transitions, the polygon mesh (Batched Dynamic Adaptive Meshes – BDAM) is also stored here.

5.10.2 Orthoimages Building Block

The orthoimages building block contains satellite and aerial images of the Earth's surface. The satellite and aerial images must be capable of being selected with respect to the viewing position over the map. The images for the map display are selected according to the map scale and perspective and are utilized as a texture file for the BDAM.

5.10.3 3D Objects Building Block

The 3D objects building block is used for storing the 3D objects that are needed to portray the digital map as realistically as possible.

The 3D objects are first of all assigned to the update regions. Within these regions, the 3D objects are combined into a space known as a BoundingBox. Several objects of the same type can be placed in this space. The BoundingBox can therefore encompass, for example, a block of buildings. The BoundingBox can be rotated and positioned freely in space. All the relevant buildings can thus be described within the BoundingBox.

The BoundingBox is organized in such a way that 3D objects are stored systematically in feature classes. Feature classes are, for instance, residential buildings, industrial buildings, bridges, trees, streetlights, and many other types of objects.

An individual 3D object is described using geometric information, a material description, textures, and names. Some of the 3D objects are depicted in great detail: these are 3D landmarks. They constitute very well-known and distinctive elements, such as the Eiffel Tower in Paris or Big Ben in London.

5.11 Junction View Building Block

Images from the junction view building block can be used to display complex intersection situations (see Fig. 3), freeway exits, and, for example, the roads joining a traffic circle. In this case, it is possible to generate the images from the road data at runtime, or prepared images can be used. The prepared images are stored in the junction view building block. The correct image is identified using such parameters as the image type, time of day, weather, color depth, etc. With these parameters and a link from the route list pointing to the image information, it is possible to select the correct image and adapt it.

Fig. 3 Junction View (Used with the kind permission of the NDS Association)

6 NDS Database Structure and Generalization

Some building blocks contain highly granular data. It is therefore necessary to group the data and store it partitioned in the database; this grouping is called "generalization" (see Fig. 4). A map view for displaying the whole of Germany on a vehicle's 10-inch screen or on a tablet, for instance, has no use for data like road geometries for the smallest categories of road in residential areas. In a map overview display, often only the higher-category roads are loaded, e.g., freeways. To ensure good performance when loading this data, the data is stored in a suitable form right from the outset. "Good performance" in this context means the user only considers the navigation map to be visually acceptable and responsive when the refresh rate is greater than 25 frames per second.

7 Structure of the NDS Database

The NDS database mainly uses an adapted SQLite engine for data processing. The NDS Committee requested, for instance, that the SQLite engine be extended with the "multiplexing" function, which has thus been implemented. Using this function, it is possible to send several queries to the database from different applications simultaneously.

The database itself is made up of functions, attributes, and metadata. The data structure is described with the help of DataScript files. A DataScript compiler translates these files into the appropriate programming language for the respective

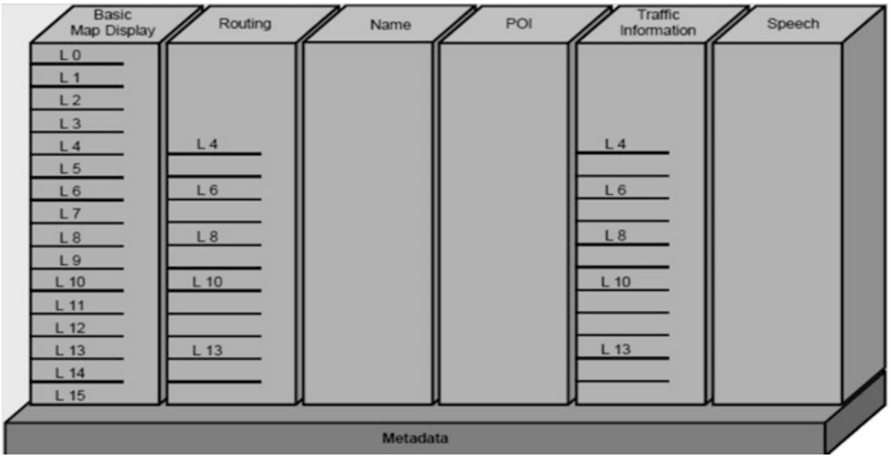


Fig. 4 Generalization (Used with the kind permission of the NDS Association)

target system, such as C++ or Java. The app developer then only needs to know the generated access classes in order to access the data; the direct SQLite commands are invisible to the app developer behind the access classes.

- Functions (features) in the NDS database
  All real, existing objects are represented in the NDS database by means of one or more features; a road segment between two intersections is represented as a link feature in the routing building block.
- Attributes in the NDS database
  The NDS database makes a distinction between fixed and flexible attributes. The attributes describe the specific properties of the NDS features. A fixed attribute is, for instance, the length information for a section of road; this attribute is always a constituent of the road segment. An optional item of information for a road segment is, for example, the days of the week on which a road is open or passable. This information is referred to as a flexible attribute of a road segment.
- Metadata in the NDS database
  The metadata contains all the information necessary for describing variable content and database attributes. The metadata refers to a specific part of the NDS database, to a building block, or to the entire database. For instance, the static metadata identifies whether distance data is stored in metric form or not; furthermore, the ISO country code is also stored in the static metadata.
- The content of the database itself is stored as a BLOB (Binary Large Object); it is possible to perform a search in the database by means of the functions, attributes, and metadata – these values are stored in the database in plain text. The situation is different, however, when it comes to the detailed map data, which is stored in binary form as a BLOB; before the individual data can be used, a BLOB must be read out of the database and interpreted. The BLOB

structure was introduced so that the binary size of an NDS database corresponds to a size that can still be handled well on an embedded device.

7.1 DataScript and RDS

The entire format of the NDS database is described using the formal description language DataScript. DataScript has two parts: firstly, there is the formal description of the database, and secondly, there is the language binding, that is to say, the representation of the data interface in a dedicated programming language. DataScript was originally developed by Godmar Back (2003). Binary formats, bitstreams, and file formats can be described with the help of DataScript. It allows the data formats needed in NDS to be described unambiguously. A reference implementation of the DataScript compiler also exists for the programming language Java. This led to the NDS consortium selecting DataScript after having evaluated various database description methods. A dialect of DataScript emerged within the NDS consortium and is referred to as Relational Data Script (RDS) (Wellmann 2013).

7.2 NDS Format Extension

The NDS data format lets app developers implement extensions with the aim of later standardization and allows them to implement adaptations and proprietary extensions. This level of freedom is necessary so that each derivate of NDS can also offer additional functions. An example of this would be the proprietary addition of parking lots (as navigation destinations) in skiing areas and the portrayal of ski lifts and pistes on the navigation map. Not everyone needs this feature, but some users may find it a useful addition.

The companies deciding in favor of an adaptation or extension of the NDS functions must comply with a strict system of rules: the NDS database must always fully comply with the standard, every modification must support the standard-compliant mechanism for a database update, and every modified NDS database must at all times completely satisfy the requirements concerning interoperability with pure NDSdatabases.

7.3 NDS Database Tools

The NDS standard supports the development efforts through several tools, which can be used for validating the database, for examining the content of an NDS database, and for testing the utilized NDS format. Developers therefore have at their disposal a validation suite, a rudimentary map viewer, the RDS compilers, an adapted and optimized SQLite engine, and various drivers.

8 The Future of the NDS Standard

Development related to the NDS standard continues to be very active. An increasing number of navigation products will become available in the coming years – with the focus primarily on a set of new functions for end users. Special emphasis will be on standardization of the incremental update process and increasing interoperability of the systems. Just beginning now are expansion efforts relating to connected services and the extension of the standard to provide data support to applications in the domain of highly automated driving.

References

Back G (2003) DataScript. http://datascript.sourceforge.net. Accessed 2003

Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS (2011) http://www.iais.fraunhofer.de/5368.html. Accessed 2009 and 2011

Müller TM (2010) Navigation data standard (NDS): Bald Industriestandard? (Navigation Data Standard (NDS): Soon an industry standard?), Automobil-Elektronik, December 2010, p 30 et seq. http://www.automobil-elektronik.de/wp-content/uploads/sites/7/2011/04/ael_6_10_internet.pdf

NDS Association (2013) (Preliminary Release) NDS Version 2.1.1.10.9 Format Specification, 2013, p 83 et seq

NDS Association (2014a) Partners. http://www.nds-association.org/partners/. Accessed 2014

NDS Association (2014b) Objectives. http://www.nds-association.org/the-nds-standard/. Accessed 2014

Wellmann H (2013) DataScript tools. http://sourceforge.net/projects/dstools. Accessed 2013

Wikipedia (2014) Point of interest. http://en.wikipedia.org/wiki/Point_of_interest. Accessed 2014

Vehicle-2-X

27

Hendrik Fuchs, Frank Hofmann, Hans Löhr, and Gunther Schaaf

Contents

H. Fuchs (✉) • F. Hofmann
Corporate Research, Digital Communication and C2XC Systems, Robert Bosch GmbH,
Hildesheim, Germany
e-mail: hendrik.fuchs@de.bosch.com; frank.hofmann2@de.bosch.com

H. Löhr
Corporate Research, Security and IT Systems, Robert Bosch GmbH, Renningen, Germany
e-mail: gunther.schaaf@de.bosch.com

G. Schaaf
Corporate Research, Driver Assistance Systems and Automated Driving, Robert Bosch GmbH,
Renningen, Germany
e-mail: gunther.schaaf@de.bosch.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_28

Abstract

Connectivity among vehicles and with infrastructure is becoming increasingly important. It is the technological basis for future "cooperative intelligent transport systems." The ability of a vehicle to communicate with its immediate surroundings (other vehicles, the road transport infrastructure, or traffic control centers) – commonly referred to as vehicle-to-X communication – enables a great number of new or improved functions for driver information and assistance. Such functions can lead to increased road safety, improved traffic efficiency, and greater personal comfort and convenience for the driver. After a short introduction, the aspects of the underlying data communication are explained. This comprises especially the radio channel and transmission system, the frequency allocation, as well as the required standardization activities. Next, a system overview is presented explaining the different subsystems, their structure and functionality, as well as the interaction with one another. Then, the important issues of data security and privacy protection are addressed, explaining the protection objectives and challenges followed by the description of potential solutions and their mechanisms. After that, the vehicle-to-X applications are described. The explanation of the basic principle and the message types is then illustrated by two practical examples, followed by some results on performance and user acceptance from the German field trial sim^{TD}. Finally, an economic assessment and possible introduction scenarios are discussed.

1 Motivation and Introduction

Connectivity among vehicles and with infrastructure is becoming increasingly important. It is the technological basis for future "cooperative intelligent transport systems" (C-ITS). The ability of a vehicle to communicate with its immediate surroundings – that is to say, with other vehicles and with the road transport infrastructure (e.g., traffic beacons, road sign gantries, traffic lights, etc.) as well as with traffic control centers – makes it possible to provide drivers with a great number of new or improved functions. Such functions can, for instance, lead to increased road safety, improved traffic efficiency, and greater personal comfort and convenience for the driver. In this context, it has become customary to use the term vehicle-to-X communication, where "X" refers to the respective communication partner. Generally the expressions “vehicle-to-X” or “vehicle-2-X” (V2X) and “car-to-X” or “car-2-X” (C2X) are used, though C2X tends to be favored in Germany and V2X is more commonly used internationally. Throughout this chapter, “V2X” (or “V2XC” for “V2X communication”) will therefore be used.

V2X has been under investigation in various German, European, and international R&D projects since as far back as the 1990s. In the first phase, they centered on the development and testing of the underlying technologies – such as in the projects “FleetNet” (Franz et al. 2005), “Network on Wheels” (Festag et al. 2008), and “PReVENT” (PReVENT 2008). The second phase included system testing using individual demonstrators in vehicles and in the infrastructure. As a result, it was possible to demonstrate the basic feasibility, e.g., in the EU projects “CVIS” (CVIS 2010), “SAFESPOT” (SAFESPOT 2010), and “COOPERS” (COOPERS 2010). In the third phase, proof of suitability for practical use was demonstrated in large-scale field trials in real traffic (“sim^{TD}” (sim^{TD} 2013a) in Germany and “Safety Pilot” (Safety Pilot 2014) in the USA), and proof of interoperability was established – that is to say, the ability of various systems developed according to the same standard to work together, such as in the “DRIVE C2X” project (DRIVE C2X 2014). Current activities are focusing on basic approaches involving hybrid communications (mobile communications and WLAN) as well as devising a V2X systems network for creating solutions to overcoming barriers to introduction, e.g., in the “CONVERGE” project (CONVERGE 2015). These activities have been supported by the Car-2-Car Communication Consortium (C2C-CC), which was initiated by the vehicle manufacturers and has meanwhile gained the support of the entire automotive industry. The C2C-CC has set itself the goal of further improving safety and efficiency in road traffic through cooperative ITS systems (C2C-CC 2015).

The inclusion of communication as a vehicle sensor makes it possible to extend the horizon of what is perceivable past the range of the driver’s vision as well as past the range of the vehicle’s onboard sensors (e.g., RADAR, LIDAR, and video). The coverage of autonomous onboard systems is curtailed by specific sensor properties, such as a sensor’s requirement for a line-of-sight connection or its limited range. Compared with this,V2X-based advanced driver assistance systems (ADAS) enable considerably more comprehensive coverage: in particular, they provide a view around bends and through obstacles, like buildings, terrain, and other vehicles. This firstly enables the effectiveness of existing driver assistance systems to be improved considerably and secondly allows many completely new functions to be realized. A disadvantage of the technology, however, is the dependency – to a greater or lesser extent – on the equipment installed in the participating vehicles and on the infrastructure components. The equipment penetration rate has a particularly strong influence on the functions that can potentially be realized, and this must therefore be factored in when considering introduction scenarios. Both aspects will be examined more closely in later sections.

2 Data Communication

2.1 Radio Channel and Transmission System

Radio-based data exchange between vehicles among themselves and between vehicles and infrastructure components places considerable demands on the transmission system. Multipath propagation due to reflections, Doppler shift as a result of high vehicle speeds, and shadowing effects of the direct signal path must be taken into account both in the system design and in transmitter and receiver development. Besides testing in the field, simulations and laboratory tests are primarily used for this. In order to be able to take into account the described propagation effects, various statistical channel models have been developed that simulate real conditions.

The radio technology used for C-ITS is based on the conventional Wi-Fi standard with special extensions. This utilizes the OFDM (Orthogonal Frequency Division Multiplexing) transmission method in which the data stream within the available frequency range is divided into multiple narrow carrier frequencies. This method has already proven its robustness in mobile environments in other systems, such as Digital Audio Broadcasting (DAB).

The radio range is mainly impaired by shadowing effects due to obstacles blocking the direct signal path between the communication partners. Such obstacles can, for instance, be other vehicles – particularly trucks – and building facades in cities, especially in the vicinity of intersections. To ensure sufficient radio range in these kinds of difficult situations, the system must have sufficient transmitting power and possess suitable receiver algorithms.

Simulations and field tests have shown that even without a direct signal path, ranges of 250–500 m can be achieved on freeways and ranges of 40–80 m at narrow inner-city intersections (Skupin 2014). The investigations performed within the scope of the sim^{TD} project have demonstrated that these ranges are sufficient for the intended functions in the respective context. In the case of direct signal paths, the ranges are significantly higher. Furthermore, the communication range can be increased considerably by relaying messages using the multi-hop method.

2.2 Frequency Allocation

To facilitate cooperative ITS services, 30 MHz bandwidth in the frequency range of 5.9 GHz was allocated in Europe under the designation ITS-G5A, with the option to extend it in the future. The individual radio channels use a bandwidth of 10 MHz. The diagram in Fig. 1 provides an overview of the frequency range and the nomenclature used (the control channel is labeled CCH; the service channels are labeled SCH).

In the USA, the frequency allocation is done by the FCC (Federal Communications Commission); see Fig. 2. The 75 MHz bandwidth is divided into seven channels with 10 MHz for each. Utilization of the same frequency range in Europe and the USA makes it possible for the hardware in vehicles to be standardized.

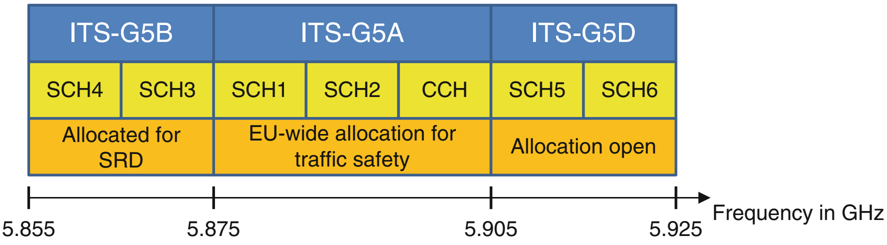


Fig. 1 Frequency allocation for ITS services in Europe (Source: Bosch)

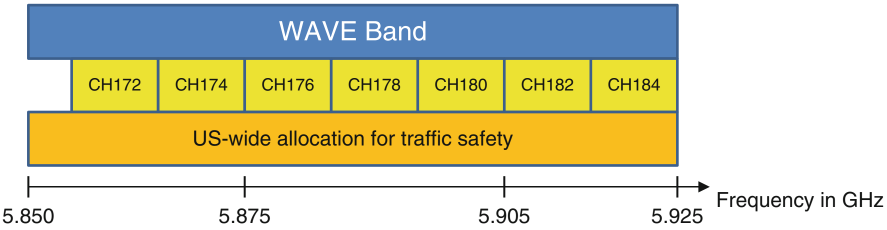


Fig. 2 Frequency allocation for ITS services in the USA (Source: Bosch)

2.3 Standardization

The standardization of C-ITS plays a particularly important role in ensuring a standardized exchange of data between vehicles and between vehicles and infrastructure components. In Europe, Mandate M/453 of the European Union (European Commission 2009) thus exists to instruct the organizations CEN, CENELEC, and ETSI to draw up a set of standards and specifications for ensuring the interoperability of C-ITS in Europe (ETSI ITS 2015; ETSI C-ITS 2015; CEN/TC 278 2015). “C-ITS Release 1” was adopted in early 2014 and specifies all the standards relevant to the first phase of introduction (ETSI ITS 2013a). In the USA, this task is incumbent upon the IEEE and SAE; in Japan, it is the responsibility of the ARIB. The standardization activities are supported by the C2C-CC.

3 System Overview

A C-ITS system comprises several subsystems that interact with one another. These subsystems are referred to as ITS stations (ETSI ITS 2010). A distinction is made between infrastructure subsystems and mobile (e.g., vehicle) subsystems:

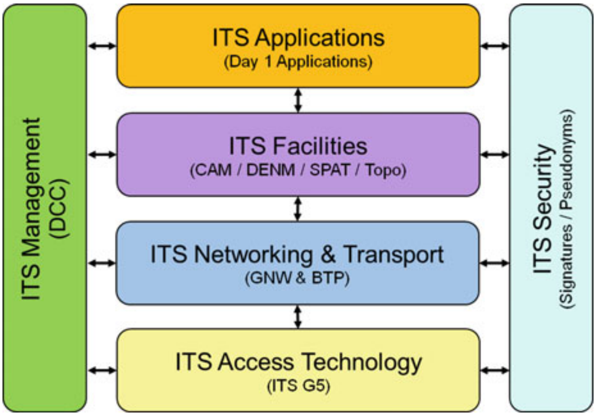


Fig. 3 Extended layer model of the ITS-G5 transmission system (Source: ETSI)

- Infrastructure subsystems
  Roadside ITS station (e.g., road sign gantries, traffic lights, etc.)
  Central ITS station (e.g., traffic control centers)
- Mobile subsystems
  Vehicle ITS station (i.e., a unit built into the vehicle with access to the vehicle sensor systems)
  Personal ITS station (e.g., a smartphone, PDA, etc.)

In the following, the basic architecture of an ITS station will be explained – it applies to all the aforementioned subsystems. The analysis below relates mainly to the vehicle subsystem, i.e., the vehicle ITS station.

3.1 ITS Station

An OSI (Open Systems Interconnection) model with additional cross-layer functionalities (as shown in Fig. 3) can be used to structure and illustrate the architecture of an ITS station.

The access component comprises layer 1 and layer 2 of the ISO OSI model and is based on the IEEE 802.11p Wi-Fi standard. This addition was specially developed for the automotive environment, firstly, in order to meet the challenges of the mobile transmission channel and, secondly, to enable ad hoc communication between vehicles with a very low latency.

The networking and transport component comprises layers 3 and 4 of the ISO OSI model. This includes, for instance, geo-networking, which allows messages to be relayed from vehicle to vehicle in order to achieve significantly greater communication ranges.

The facilities component comprises layers 5, 6, and 7 of the ISO OSI model and covers an extensive range of tasks. These tasks include, among other things, generating messages for sending to other vehicles and infrastructure components (see CAM and DENM in Sect. 5), computing the vehicle's own position, determining the precise time, and managing a local dynamic map.

The applications component addresses three categories of services: road safety, traffic efficiency, and miscellaneous services. Details on this are provided in Sect. 5.

The core task of the management component is overload monitoring of the transmission channel. Since all vehicles and infrastructure components use the same frequency range, attention must be paid to ensuring no transmission collisions occur and that all the important messages are transmitted almost without delay. Thus, before using the channel, a transmitting station must not only make sure it is free but must also adjust its transmitting power and, if necessary, the repetition rate of messages to suit the channel load. This ensures equitable utilization among all participants. To achieve this, an interaction with several layers of the transmission system is necessary (depicted in Fig. 3 by a vertical block spanning the layers).

The security component also has several cross-layer tasks and is therefore likewise depicted as a vertical block. Its purpose is to ensure the integrity, authenticity, and anonymity of messages (as explained in more detail in Sect. 4). Of importance here are the signing of messages to be sent and verification of the signatures of received messages.

4 Data Security and Privacy Protection

Protection against unauthorized manipulation is, needless to say, extremely important in the context of V2X-based assistance systems and functionalities. Furthermore, the continuous transmission of data by vehicles, like their location and speed, also raises questions concerning data protection and privacy protection for the vehicle occupants.

This section will first of all examine the basic problems surrounding data security and privacy protection in vehicular communication. Afterwards, existing potential solutions and the current state of technology will be briefly outlined.

4.1 Security Issues

Without suitable protective measures, there is a risk in vehicular communication that fake messages can be sent and legitimate messages can be altered. Using such manipulation, attackers could, for instance, fake nonexistent traffic congestion or a construction site with the aim of hindering or diverting other vehicles; and with messages that fake extreme situations, like an emergency stop, they could put other road users in danger. These and similar scenarios make it necessary for manipulated messages and messages sent by illegitimate sources to be recognized as being invalid. That is to say, the authenticity (i.e., the message really is from the alleged sender, in this case, a legitimate vehicular communication partner) and the integrity (i.e., the message was not manipulated) of the messages must be assured.

Since vehicular communication primarily involves broadcast messages that are of relevance to safety and that should be readable for everyone, the confidentiality of the message content (i.e., the secrecy of the sent data) is not a top priority. In other areas of application, such as payment services, this may be quite different, though. For this reason, the capability for confidential communication must also be provided.

4.2 Privacy Aspects

If messages sent within the context of vehicular communication could be associated with a vehicle or its driver with little effort, this could have serious consequences for the privacy of the driver – and other vehicle occupants. If vehicles are continuously sending messages, they can be received by any individual or organization within range and also stored. This could lead to critical situations arising if, for instance, it were easy to determine or even prove that a particular vehicle was present at a potentially sensitive event, like a political party convention, general assembly, etc. The possibility that logged messages could be of assistance in the automated punishment of traffic offenses is an issue that is also worthy of discussion.

To avoid the vehicle occupants' privacy being severely affected as a result of vehicular communication, an effort should be made to prevent any permanent identifiers that are easily associated with the vehicle from being used in these messages. Furthermore, only the data that is necessary for each specific usage scenario should be transmitted (principle of "data minimization"). In particular, when devising a security solution, attention must be paid to ensuring that the cryptographic keys cannot be used to identify and trace vehicles easily or even verifiably.

4.3 Protection Objectives and Challenges

The following fundamental protection objectives emerge from the above considerations:

- **Integrity**: The messages must be protected against manipulation.
- **Authenticity**: It must be ensured that only messages from legitimate participants are accepted. In special cases, such as the investigation of a serious crime, it may even be desirable that the origin of a message can be proven (non-repudiation) also to a third party, such as a law court, after the necessary order to do so has been issued by a sovereign authority.
- **Anonymity/pseudonymity**: To keep curtailment of users' privacy to a minimum, support for anonymous or pseudonymous communication should be provided. When using pseudonyms, attention must be paid to ensuring that de-pseudonymization (i.e., the association of various pseudonyms with a particular user) can at the most only be easily performed by authorized parties.
- **Confidentiality**: It must be possible to optionally ensure the confidentiality of messages, if the application requires this.

In addition to this, the following aspects also play an important role in the security solution:

- **Performance**: The security mechanisms must be sufficiently efficient not to impair communication. In particular, steps must be taken to ensure that the realistically expected quantity of received messages can indeed be processed.
- **Costs/effort**: The costs and effort entailed in equipping vehicles and roadside units with security modules should be kept as low as possible. The same applies to the infrastructure required for certificate and key management.
- **Maintainability/future proofing**: Since a system that has been rolled out in the automotive sector must function in the field for decades, it is essential that one consider the aspects of maintainability and future proofing before the system's introduction. This naturally also applies to the security solution, which should provide – at least as far as can be judged from today's perspective – the necessary security over corresponding periods of time.

4.4 Potential Solutions and Their Mechanisms

To secure communications adequately from a cryptographic perspective, the current intended solution is for all messages to be digitally signed. The signatures generated by vehicles can be verified using pseudonymous certificates that change on a per vehicle basis and that are issued by a dedicated public key infrastructure (PKI) (Kargl et al. 2008; Papadimitratos et al. 2008). Figure 4 provides an overview

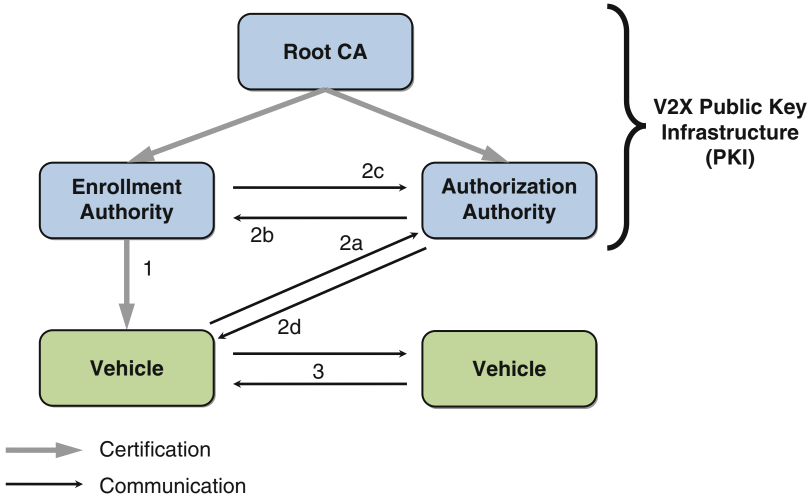


Fig. 4 Securing vehicular communication. Phase 1: Vehicle receives an enrollment credential (long-term certificate). Phase 2: Vehicle receives authorization tickets (pseudonymous certificates). Phase 3: Vehicles communicate with one another pseudonymously (Source: Bosch)

of this approach as defined by C2C-CC and ETSI (ETSI ITS 2012; Bißmeyer et al. 2011).

The PKI comprises three different types of certificate authority (CA): root CA, enrollment authority (EA, also called long-term CA), and authorization authority (AA, also called pseudonym CA). The trust anchor in the security architecture is the root CA, which is known to all system participants. The root CA issues certificates for the EA and AA and monitors compliance with the guidelines governing the issuing of certificates. It is envisaged that there will be several instances of enrollment and authorization authorities that are operated by different organizations.

A vehicle that wants to participate in the communication system first of all requires an enrollment credential (EC, also called a long-term certificate) from an enrollment authority (see Fig. 4, Step 1).

After receiving an enrollment credential, the vehicle can generate pseudonyms (cryptographic key pairs) and have these certified by an authorization authority (see Fig. 4, Step 2). To do this, the vehicle encrypts its enrollment credential for the enrollment authority and sends the encrypted credential together with the newly generated public keys (the pseudonyms) to the authorization authority (Step 2a). The authorization authority then sends the encrypted credential to the responsible enrollment authority, which then examines whether it is a proper certificate that qualifies for issuance of pseudonymous certificates (Step 2b). After a positive response from the enrollment authority (Step 2c), the authorization authority certifies the public keys with a digital signature and thus generates the so-called authorization tickets (ATs). These are now transmitted to the vehicle (Step 2d), which can then use them for pseudonymized communication.

In order to secure the messages for communication between vehicles and for communication between vehicles and infrastructure components (see Fig. 4, Step 3), the messages are now signed digitally by the sender. A private key is used for this – the sender possesses a valid authorization ticket for the public counterpart to this private key. The recipient of the message can now verify that the message was generated by a legitimate communication partner and was received unaltered. To do this, the recipient must first check the signature using the public key from the sender's authorization ticket; then, the recipient must verify that the authorization ticket is valid (in particular, that it has not expired) and that it was signed by a legitimate authorization authority. If the issuing authorization authority is not yet known to the recipient, the recipient must verify the AA's certificate (which was signed by the root CA) in order to make sure that the authorization authority in question is indeed legitimate. This therefore makes it possible for secure communications to also take place between vehicles that are encountering one another for the first time and that both simply trust the same root CA. If required, the system can be expanded to integrate several root CAs that certify one another by the so-called cross-certification. Optionally, confidential messages can also be transmitted in encrypted form. Since this is not necessary for the main currently intended (safety-related) application scenarios, it will not be dealt with in more detail here.

Effective pseudonymization is made possible through the procedure described here on the premise that the vehicles replace their authorization ticket often enough to prevent the association between the AT and vehicle from being easily identified. In order to ascertain the identity of the sender (which is equivalent to the sender's enrollment credential) on the basis of received messages, the authorization authority (that issued the authorization tickets) and the enrollment authority (as the only CA that has seen the vehicle's enrollment credential in unencrypted form) must cooperate with one another. If this identification of the sender is to be made possible (for instance, for the purpose of revealing it to a sovereign authority), the AA and EA must keep a log of the issuance of the ATs and of the examination of the ECs, respectively. By merging these logs, it is possible to revoke the pseudonymity of the sender of a message.

4.5 Current State of Technology and Implementation

The cryptographic mechanisms and data formats necessary for validation are already specified. The certificate and message formats are defined in the ETSI TS 103 097 technical specification (for Europe) (ETSI ITS 2013b) and in the IEEE 1609.2 standard (for North America) (IEEE Standards Association 2013). For the digital signatures, the Elliptic Curve Digital Signature Algorithm (ECDSA; see, for instance, NIST 2013) can be used. For optional encryption, the intention is to use the Elliptic Curve Integrated Encryption Scheme (ECIES; see, for instance, IEEE 1363a (IEEE Standards Association 2004)). Furthermore, the entire PKI is already implemented as a prototype (currently operated by C2C-CC) and will soon be going live. It is intended that certain differences between the US and European standards will be harmonized. Also, prototypes and projects (e.g., the EU "PRESERVE" project (PRESERVE 2013)) already exist in the area of design and deployment of cryptographic accelerators, which ensure the necessary in-vehicle performance.

5 Vehicle-2-X Applications

5.1 Requirements and Basic Functionality

Like conventional advanced driver assistance systems (ADAS) that are based on autonomous onboard sensor technology, the V2X-based ADAS are intended to assist the driver with the task of driving. This includes such aspects as informing the driver about events on the road up ahead and providing hazard warnings. In order to achieve this, the system must detect such events, and a decision must be made about their relevance to the driving task. If a detected event proves to be of relevance, the system will relay this to the driver: in the simplest case, this will take place as a warning via the HMI (e.g., a warning tone, a notification in the display, haptic feedback, etc.), though, in principle, it is also conceivable that the system could intervene in the driving function.

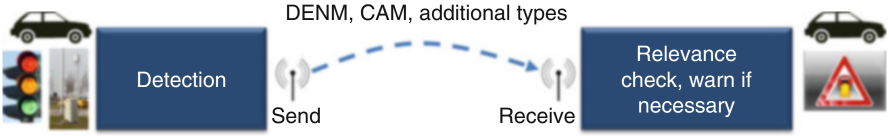


Fig. 5 In the case of V2X functions, the detection of a hazard and the interpretation of the associated message often take place at different nodes (Source: Bosch)

In contrast to ADAS based on autonomous onboard sensor technology, the processes of detection and interpretation often take place at two different nodes (see Fig. 5). These two nodes, however, need not both be vehicles as they are in the case of V2V. They could, for instance, be traffic lights that provide information about their current and future phases (infrastructure-to-vehicle – I2V) or a roadside unit (RSU) that collects vehicle movement data for generating a congestion forecast (vehicle-to-infrastructure – V2I).

Data is relayed between the two nodes through transmission of standardized messages. For V2V communication, the following two message types are used most frequently (ETSI ITS 2013c, d):

- Decentralized Environmental Notification Message (DENM) for event-driven warnings about local hazards
- Cooperative Awareness Message (CAM) for continuous observation of vehicles in the vicinity

In addition to this, there are also a large number of highly specific message types used for providing details about, e.g., traffic light phases, road signs, and intersection topologies.

In the case of a DENM, the detection of a local hazard (a process that is often very complex) takes place in the transmitting vehicle. Generally, this involves analysis and assessment of the vehicle sensor data obtained via the CAN bus; but it is also conceivable that drivers themselves could issue reports about, for instance, obstacles they have spotted on the road. All the data necessary for identifying the detected hazard is added to the message and sent. The receiving vehicle must then only evaluate the relevance of the hazard. This essentially involves determining whether the hazard is located on the vehicle's route and, if applicable, when the hazard will be reached, so that in that case a warning is issued in good time.

CAM-based functions, on the other hand, analyze the data sent periodically by all vehicles about their respective state of motion (position, driving direction, speed, etc.) only at the receiving end. This involves a sophisticated analysis of the available CAM data as well as an assessment of its relevance. Thus, for instance, the risk of collision near intersections that are not visible or difficult to see can be computed using the CAM data from all vehicles in the area, so that a warning can be issued, if applicable.

Table 1 DENM and CAM compared

	DENM	CAM
Sender	Complex detection, e.g., sophisticated analysis of CAN data	Read out CAN data about the vehicle's state of motion
Message content	Processed specific data, e.g., position, time, obstacle type, period of validity, etc.	CAN data about the vehicle's state of motion (position, time, driving direction, speed)
Receiver	Straightforward examination of relevance: is the hazard located on the route, and, if applicable, when will it be reached?	Sophisticated analysis of the CAM data and examination of relevance, e. g., calculation of the collision risk using the CAM data from all vehicles
Transmission mode	Event driven	Periodic

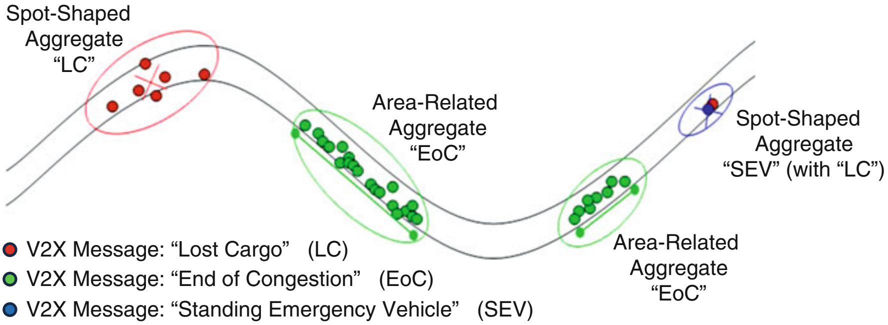


Fig. 6 Aggregation of several messages about the same event (Source: Bosch)

The fundamental characteristics of DENM and CAM are compared with one another in Table 1.

Since several V2X functions are generally active at the same time in the vehicles, a large number of different messages will arrive at the receiving parties. The individual function then only selects the ones that are relevant to it, e.g., an obstacle warning system will select the messages concerning obstacles.

A further problem has to do with the fact that generally several messages will arrive concerning the same event, since a hazard has been detected by several transmitting parties. To ensure multiple warnings do not occur, the received messages are aggregated. Depending on the type of event, a distinction must be made between spot-shaped aggregates (e.g., in the case of an obstacle) and area-related aggregates (e.g., the tail end of congestion whose position changes with time or an area of adverse weather). Figure 6 shows some examples of spot-shaped and area-related aggregates. The type of aggregate depends on the context of the notification: thus, for example, in the case of a "lost cargo," various vehicles will report numerous positions spread over an area due, for instance, to imprecise sensor systems or reports triggered manually when driving past. Since it is a spot-shaped

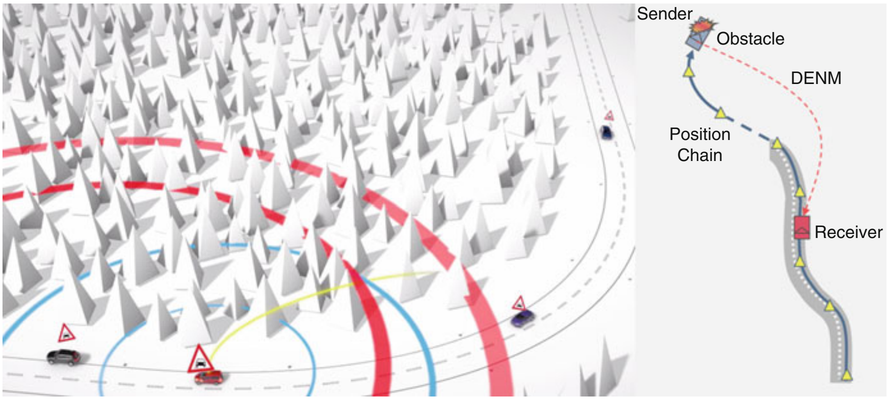


Fig. 7 Example of application: the obstacle warning system (Source: Bosch)

event (even though it can exhibit a certain amount but not a large amount of spread), the reports are aggregated accordingly, as is indicated by the cross in the diagram.

5.2 Practical Examples

Below are two specific examples of application:

5.2.1 Obstacle Warning System

The obstacle warning system issues warnings about the presence of such obstacles as broken-down vehicles or persons, animals, and objects on the road. In the case of the latter, their detection may simply be performed by the driver who, upon seeing the hazard, triggers a warning report through manual interaction with the HMI. If the vehicle is additionally fitted with, say, a camera that has image processing capabilities, this process could also be automated. Obstacles on the road can also be detected through the analysis of evasive maneuvers.

If a vehicle becomes an obstacle itself due to a breakdown or an accident, the detection takes place automatically through the analysis of various CAN data. In the simplest case, a broken-down vehicle detects itself as an obstacle due to it being stationary and having the hazard lights switched on (see Fig. 7 in which the vehicle at the very left of the diagram is stationary and its hazard lights are on).

A detected hazard leads to a DENM being sent that is received by the vehicles traveling behind. The message contains, among other things, details about the obstacle's type, time, and position and about the message's area of coverage. All vehicles that are located within the sender's range of communication will receive the message directly. This method of transmission cannot be guaranteed for receivers that are further away, and so, in this case, the message is relayed between several vehicles using the multi-hop method, as indicated by the yellow line.

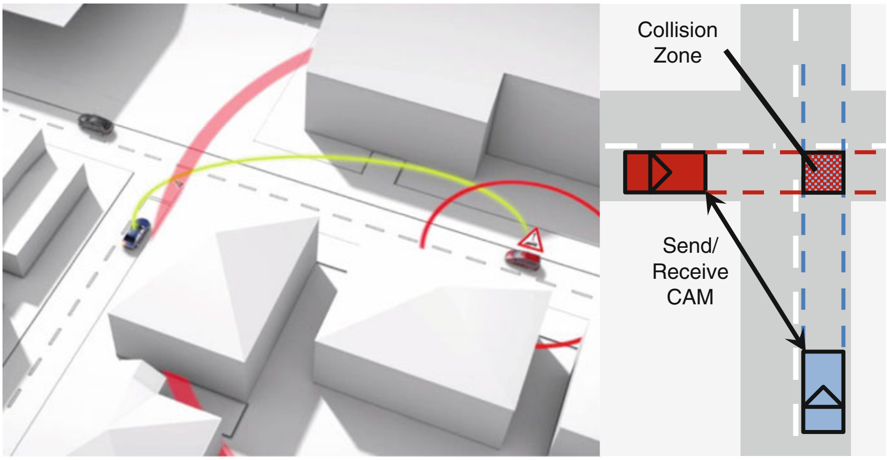


Fig. 8 Example of application: the intersection/cross-traffic assistant (Source: Bosch)

Receiving vehicles check whether they are heading toward the obstacle (spatial relevance) and whether there is any urgency (temporal relevance), that is to say, whether it is appropriate to inform the driver or display a warning on the HMI. The spatial relevance can be checked by comparing the obstacle's position with the vehicle's route using a digital map or by comparing the current position of the receiver with a position chain (i.e., the sequence of historical positions of the sending vehicle) that is contained in the DENM (see Fig. 7, right). The received details about the distance to the obstacle can be used in conjunction with the vehicle speed to determine the "time to obstacle," which can then be used to ascertain the temporal relevance.

5.2.2 Intersection/Cross-Traffic Assistant

The intersection/cross-traffic assistant informs or warns the driver in case of potential collision with turning traffic or cross-traffic at intersections and at road entrances (see Fig. 8; cf. also ▶ Chap. 50, "Intersection Assistance"). The vehicles approaching the intersection and the ones at the intersection send CAMs at a sufficiently high rate containing position and motion data (speed, driving direction, etc.). At the same time, they receive such data from other vehicles. The motion of one's own vehicle and that of the other vehicles is predicted and used to ascertain a collision risk.

First of all, a collision zone is determined for each sending vehicle (as shown on the right in Fig. 8). This is essentially the location where the driving trajectories of the vehicles intersect. If the red vehicle turns right, no such zone exists. If, however, such a zone does exist, the system checks whether the vehicles involved will be within the zone at approximately the same time or whether one of the two vehicles will pass through the zone significantly earlier than the other. In this case, the system in the vehicle that does not have the right of way computes the time it will

Fig. 9 Functions implemented and examined in the sim$^{\mathrm{TD}}$ project (Source: sim$^{\mathrm{TD}}$)

take for it to reach this zone; if this value falls below a critical time threshold, the system warns the driver in good time. This computation takes into account an analysis of the driver's intentions (cf. ▶ Chap. 37, "Driver Condition Detection"); so, for instance, if the driver has already begun to apply the brakes, no warning will be issued.

5.3 Implementation and Testing in the sim$^{\mathrm{TD}}$ Project

In Germany, the suitability of the V2X technology for practical use as well as its effectiveness and benefit was explored extensively within the scope of the sim$^{\mathrm{TD}}$ project by means of a large-scale field trial performed under real traffic conditions in the Frankfurt/Main metropolitan area. The traffic infrastructure was equipped with more than 100 roadside stations, and a road test fleet of 120 vehicles was established. During the six-month test period, a total of around 1.65 million kilometers was driven by more than 500 normal drivers.

Besides development and provisioning of the technical subsystems, a total of 21 functions were selected, specified, implemented, and made ready for road use in a self-contained automotive proving ground in order to then put them to the test in the field trial (sim$^{\mathrm{TD}}$ 2009); see Fig. 9.

By performing appropriate experiments based on the measurement data collected during the field trial, it was possible to accept or reject previously proposed hypotheses on the benefit and effect of the individual functions. These experiments were performed according to detailed scripts that contained the precise experimental procedure and corresponding instructions for the drivers. The following example serves to illustrate this process. In an experiment to test the obstacle warning function, the script assigned one vehicle the role of a broken-down vehicle. It was to drive along a predefined route to a position that could not be seen by the following traffic, and from there it was to identify itself as a broken-down vehicle, causing a corresponding DENM to be sent out. All the other vehicles were to drive along the same route at time intervals to check whether they would be warned about the hazard in good time. Among the data that was measured in this experiment were the points in time a DENM was sent and received as well as the point in time the warning was indicated. Using the respective vehicle positions and the speed of the receiving vehicle, it was possible to ascertain the range of communication as well as the timeliness of the warning and thus verify the corresponding hypothesis pertaining to the benefit of the obstacle warning system.

Based on the data from the field trial – and supported by driving and traffic simulations – it was possible to identify further positive effects on driving/road safety and efficiency. For instance, as a result of the obstacle warning, the brakes were applied up to 50 m sooner, and the vehicles drove past the hazard spot up to 15 km/h more slowly; furthermore, in the case of the electronic brake light, an improvement in the reaction time of around 60 % was observed for vehicles traveling behind. Of particular note in terms of an increase in traffic efficiency is the "traffic light phase assistant" ("green wave assistant"): in a traffic simulation calibrated with field trial data, it could be shown that lost time and the number of stops decrease. These effects already occur at penetration rates of 5 % and increase further at higher penetration rates; furthermore, the speed limit is exceeded less frequently on the approach to an intersection.

In addition to this, a very high acceptance among users was identified. Depending on the function, between 50 % and 80 % of drivers responding to a written survey stated: "I would like the application in my vehicle." A majority of drivers expressed the desire to use the V2XC functions when they are introduced to the market.

6 Economic Assessment and Introduction Scenarios

6.1 Effect and Benefit

The effect and benefit of the V2X functions were investigated and assessed in sim^{TD} using the procedure shown in Fig. 10 (sim^{TD} 2013b).

The upper path describes the procedure for determining the effect on road safety. The GIDAS accident database for accidents involving personal injury served as the basis for this research (GIDAS 2015). In the first step, an analysis of the field of effect was performed. The field of effect describes the accidents that are able to be addressed by a function as a proportion of the total incidence of accidents (i.e., it indicates the theoretical maximum possible potential of a function). So, for instance, the assessment for an intersection collision warning system is performed

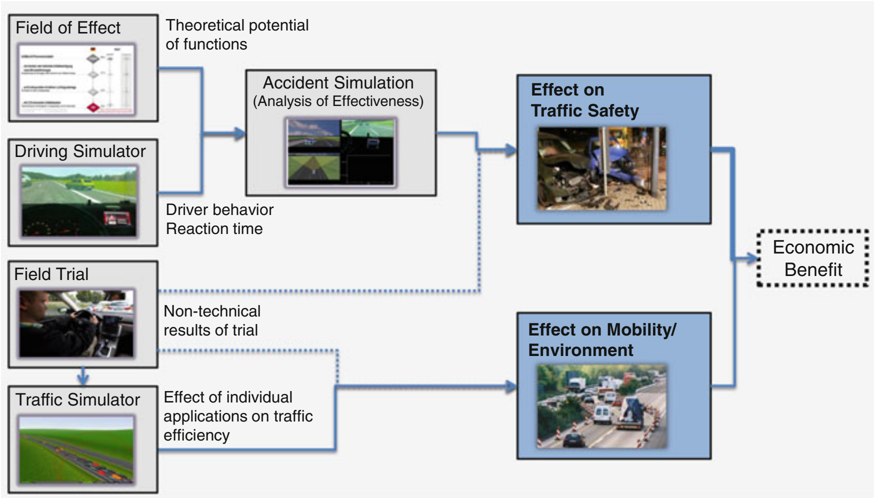


Fig. 10 Procedure employed in sim^{TD} to ascertain the effect and economic benefit (Source: Bosch)

Table 2 Results of the accident simulation for the degree of effectiveness analysis

Function	Number of simulated accidents in the respective field of effect	Number of avoided accidents	Degree of effectiveness (%)
Electronic brake light	173	33	19
Road sign assistant (stop signs)	92	13	14
Intersection/cross-traffic assistant	450	243	54

only on the basis of accidents involving intersections. For the functions that have an effect on safety, the field of effect analysis revealed a field of effect of more than 30 % – though the exact design of the function was not taken into account and the actually attainable value may be lower. The actual effectiveness of a function is identified by the degree of effectiveness: the degree of effectiveness specifies the portion of accidents in the field of effect that can be avoided with the help of the function; the product of the degree of effectiveness and the field of effect thus identifies the maximum possible effect of an individual function on the total incidence of accidents.

The degree of effectiveness was then determined through simulation of real accidents taken from the GIDAS database. Due to the considerable effort associated with this evaluation, it could only be performed for the three functions with the greatest field of effect. The results are shown in Table 2.

The degree of effectiveness analysis only takes into account those accidents that could be avoided, thanks to the function. Besides this, further parameters were also ascertained during the evaluation of the accident simulations that are likewise relevant to an economic assessment: for instance, the reduction in the severity of injuries in accidents involving personal injury and the avoidance of accidents involving damage to property.

The lower path in Fig. 10 shows the procedure for using traffic simulations to identify the effect of functions with regard to mobility and the environment. This investigation showed that, among other things, journey times could be reduced by up to 7 % through dynamic detour information and by up to 9 % through traffic light phase assistance.

6.2 Economic Assessment

When examining the described findings concerning the effect of V2X functions, it is possible to deduce a significant benefit to the economy (sim^{TD} 2013c).

- With complete penetration of V2X functions, up to €6.5 billion of the economic costs of road traffic accidents could be avoided every year.
- In addition to this, due to efficiency effects and through avoidance of environmental pollution, it is possible to realize an economic benefit of €4.9 billion.

On the assumption of ideal overall conditions, a maximum benefit-cost ratio greater than 8 was calculated for the period 2015–2035. If one then takes as the basis the typical development trend for penetration rates of new standard features, one can expect a cumulative benefit-cost ratio of 3 over the first 20 years. This figure is generally accepted as justification for relevant investments by the public sector in the infrastructure.

6.3 Introduction Scenarios and Outlook

The main problem faced by the introduction of V2X systems is the strong dependency on the penetration rate. Investment in the infrastructure only begins to make sense when there are a sufficient number of suitably equipped vehicles; and installing the respective equipment in vehicles only begins to pay off when marketable V2X functions exist that provide customers with a benefit – preferably a benefit that they are able to experience for themselves directly. Many different introduction scenarios have been considered: for instance, introduction through a self-imposed commitment by carmakers, introduction by the state through statutory provisions, and introduction by means of support from sponsors, such as insurance firms. The latter would pay all or at least part of the vehicle equipment costs and would in return receive access to mobility data that they can use to modify the structure of their policies and premiums. Since there does not appear to be an ideal solution in this regard, the conclusion has been reached that an introduction will

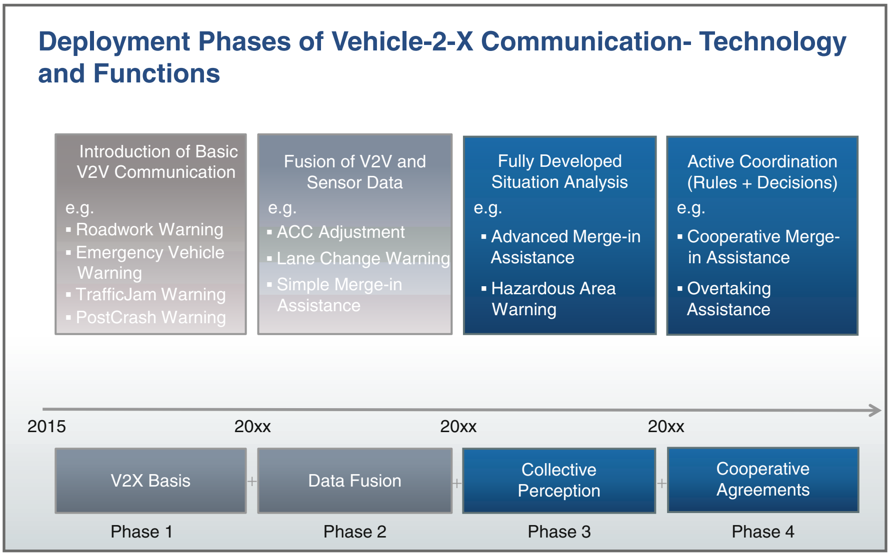


Fig. 11 Phase model for the introduction of vehicle-2-X systems and associated applications (Source: Volkswagen AG)

only be possible by way of a joint effort involving vehicle manufacturers and the public sector. Within the scope of the C2C-CC, the majority of vehicle manufacturers have therefore signed a memorandum of understanding (MoU) for the introduction of V2X systems; and, through the C-ITS Eurokorridor project (BMVI 2013), the public sector is attempting to provide a starting impetus for equipping the road transport infrastructure so that it will be possible for users to make initial early experiences with V2X functions.

The dependency on the penetration rate varies considerably from function to function. The C2C-CC has therefore developed a phase model for a staggered introduction of vehicle-2-X applications. Figure 11 shows a diagram by Volkswagen AG that is based on this.

The first phase includes the so-called “Day 1” functions that have a low dependency on the penetration rate; the functionality and therefore also the associated complexity increase as the penetration rate increases from phase to phase. Ultimately, it is expected that the V2X technology will also contribute decisively to automated driving.

To lower the obstacles to introduction, the architecture for a C-ITS systems network is, for instance, being developed in the current CONVERGE project. It involves looking more closely at the following points in particular:

- Flexible and future-proof concepts with distributed rights of ownership and distributed control in order to decouple technical solutions from the operator-specific requirements

- Openness to new actors, new services, and transnational operation through the new approach of institutional role models
- Hybrid communication, that is to say, access via a variety of communication technologies and operator platforms
- Guarantee of end-to-end security and privacy protection

References

Bißmeyer N, Stübing H, Schoch E, Götz S, Stotz J, Lonc B (2011) A generic public key infrastructure for securing Car-to-X communication. The 18th world congress on intelligent transport systems, Orlando

BMVI – German Federal Ministry of Transport and Digital Infrastructure (2013) Cooperative ITS corridor joint deployment. http://www.bmvi.de/SharedDocs/EN/Anlagen/VerkehrUndMobilitaet/Strasse/cooperative-its-corridor.pdf?__blob=publicationFile. Accessed 02 Mar 2015

CAR 2 CAR Communication Consortium (2015) Consortium home page. http://www.car-to-car.org. Accessed 02 Mar 2015

CEN/TC 278 (2015) Intelligent transport systems. http://www.itsstandards.eu. Accessed 02 Mar 2015

CONVERGE Project (2015) Project home page. http://www.converge-online.de/?id=000000&spid=en. Accessed 02 Mar 2015

COOPERS Project (2010) Project home page. http://www.coopers-ip.eu http://cordis.europa.eu/projects/rcn/79301_en.html. Accessed 02 Mar 2015

CVIS Project (2010) Project reference page in CORDIS. http://cordis.europa.eu/projects/rcn/79316_en.html. Accessed 02 Mar 2015

DRIVE C2X Project (2014) Project home page. http://www.drive-c2x.eu/project. Accessed 02 Mar 2015

ETSI ITS (2010) EN 302 665: ITS communications architecture. http://www.etsi.org/deliver/etsi_en/302600_302699/302665/01.01.01_60/en_302665v010101p.pdf. Accessed 02 Mar 2015

ETSI ITS (2012) TS 102 940: ITS communication security architecture and security management. http://www.etsi.org/deliver/etsi_ts/102900_102999/102940/01.01.01_60/ts_102940v010101p.pdf. Accessed 02 Mar 2015

ETSI ITS (2013a) TR 101 607: cooperative ITS (C-ITS). http://www.etsi.org/deliver/etsi_tr/101600_101699/101607/01.01.01_60/tr_101607v010101p.pdf. Accessed 02 Mar 2015

ETSI ITS (2013b) TS 103 097: security header and certificate formats. http://www.etsi.org/deliver/etsi_ts/103000_103099/103097/01.01.01_60/ts_103097v010101p.pdf. Accessed 02 Mar 2015

ETSI ITS (2013c) EN 302 637-2: specification of cooperative awareness basic service. http://www.etsi.org/deliver/etsi_en/302600_302699/30263702/01.03.00_20/en_30263702v010300a.pdf. Accessed 02 Mar 2015

ETSI ITS (2013d) EN 302 637-3: specifications of decentralized environmental notification basic service. http://www.etsi.org/deliver/etsi_en/302600_302699/30263703/01.02.00_20/en_30263703v010200a.pdf. Accessed 02 Mar 2015

ETSI C-ITS (2015) ETSI C-ITS home page. http://www.etsi.org/technologies-clusters/technologies/intelligent-transport/cooperative-its?highlight=YToxOntpOjA7czozOiJpdHMiO30=. Accessed 02 Mar 2015

ETSI ITS (2015) ETSI ITS home page. http://www.etsi.org/technologies-clusters/technologies/intelligent-transport. Accessed 02 Mar 2015

European Commission (2009) M/453 EN: standardisation mandate addressed to CEN, CENELEC and ETSI in the field of information and communication technologies to support the interoperability of co-operative systems for intelligent transport in the European community. http://ec.europa.eu/enterprise/sectors/ict/files/standardisation_mandate_en.pdf. Accessed 02 Mar 2015

Festag A, Noecker G, Strassberger M, Lübke A, Bochow B, Torrent Moreno M, Schnaufer S, Eigner R, Catrinescu C, Kunisch J (2008) NoW – network on wheels: project objectives, technology and achievements. In: Proceedings of the 5th international workshop on intelligent transportation (WIT), Hamburg

Franz W, Hartenstein H, Mauve M (eds) (2005) Inter-vehicle-communications based on ad hoc networking principles: the FleetNet project. Universitätsverlag Karlsruhe, Karlsruhe

GIDAS (2015) German in-depth accident study. http://www.gidas.org. Accessed 02 Mar 2015

IEEE Standards Association (2004) 1363a-2004: IEEE standard specifications for public-key cryptography. http://grouper.ieee.org/groups/1363/. Accessed 02 Mar 2015

IEEE Standards Association (2013) 1609.2-2013: IEEE standard for wireless access in vehicular environments – security services for applications and management messages. http://standards.ieee.org/develop/wg/1609_WG.html. Accessed 02 Mar 2015

Kargl F, Papadimitratos P, Buttyan L, Müller M, Schoch E, Wiedersheim B, Thong T, Calandriello G, Held A, Kung A, Hubaux J (2008) Secure vehicular communication systems: implementation, performance, and research challenges. IEEE Commun Mag 46(11):110–118

National Institute of Standards and Technology (2013) The digital signature standard (DSS). FIPS Publication, U.S. Department of Commerce, Gaithersburg, MD, 186–4

Papadimitratos P, Buttyan L, Holczer T, Schoch E, Freudiger J, Raya M, Ma Z, Kargl F, Kung A, Hubaux J (2008) Secure vehicular communication systems: design and architecture. IEEE Commun Mag 46(11):100–109

PRESERVE project (2013) Project home page. http://www.preserve-project.eu. Accessed 02 Mar 2015

PReVENT project (2008) Project reference page in CORDIS. http://cordis.europa.eu/result/report/rcn/45077_en.html. Accessed 02 Mar 2015

SAFESPOT project (2010) Project home page. http://www.safespot-eu.org. Accessed 02 Mar 2015

Safety Pilot (2014) Safety pilot fact sheet. http://www.its.dot.gov/safety_pilot/pdf/safetypilot_nhtsa_factsheet.pdf. Accessed 02 Mar 2015

sim^{TD} project (2009) Deliverable D11.2: Ausgewählte Funktionen (Selected Functions). http://www.simtd.de/index.dhtml/object.media/enEN/6495/CS/-/backup_publications/Projektergebnisse/simTD-Deliverable-D11.2_Ausgewaehlte_Funktionen.pdf. Accessed 02 Mar 2015

sim^{TD} project (2013a) Project home page. http://www.simtd.de/index.dhtml/enEN/index.html. Accessed 02 Mar 2015

sim^{TD} project (2013b) Deliverable D5.5 Part B-1A: Simulation realer Verkehrsunfälle zur Bestimmung des Nutzens für ausgewählte simTD-Anwendungsfälle auf Basis der GIDAS Wirkfeldanalyse (Simulation of real road traffic accidents for determining the benefit for selected simTD application scenarios based on the GIDAS field of effect analysis). http://www.simtd.de/index.dhtml/object.media/enEN/8145/CS/-/backup_publications/Projektergebnisse/simTD-TP5-Abschlussbericht_Teil_B1-A_GIDAS-Wirkgradanalyse_V10.pdf. Accessed 02 Mar 2015

sim^{TD} project (2013c) Deliverable D5.5 Part B-4: Ökonomische Analyse (Economic analysis). http://www.simtd.de/index.dhtml/object.media/enEN/8109/CS/-/backup_publications/Projektergebnisse/simTD-TP5-Abschlussbericht_Teil_B-4_Oekonomische_Analyse_V10.pdf. Accessed 02 Mar 2015

Skupin C (2014) Zur Steigerung der Kommunikationszuverlässigkeit von IEEE 802.11p im fahrzeugspezifischen Umfeld (Enhancing the communication reliability of IEEE 802.11p in the vehicular environment). Shaker Verlag, Aachen

Backend Systems for ADAS

28

Felix Klanner and Christian Ruhhammer

Contents

Abstract

Nowadays, a wide variety of in-vehicle services connect to a backend system via Internet. The key is to deliver information to the vehicle that is not locally available but accessible via Internet. For example, systems such as Google Traffic use fleet data to analyze the current traffic situation. This chapter gives an overview of available technologies for transmitting, storing, and analyzing data in a backend system. Based on simulation and measurement methods, we investigated the time required for transmitting data via cellular networks. The estimated transmission time is about 400 ms, whereby it can increase to 1 s, depending on the traffic situation and the condition of the cellular network. The transmitted data are then available in the backend system for further analysis.

F. Klanner (✉)
BMW Forschung und Technik, München, Germany
e-mail: felix.klanner@bmw.de

C. Ruhhammer
BMW Group, München, Germany
e-mail: christian.ruhhammer@bmw.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_29

The technological background of the methods used for data storage and analysis is introduced by an example of a minimalistic programming for a local danger warning database system. The example of extracting parameters in intersections to support driver assistance systems illustrates how relevant information can be generated from fleet data. Hence, these data allow an enhancement of as yet prototypically developed driver assistance systems and enable the development of new systems.

1 State-of-the-Art Backend-Based Driver Assistance Systems

Many latest driver assistance systems transfer data with backend systems via cellular communication. These include the visualization of the actual traffic flow in vehicles (e.g., BMW Real-Time Traffic Information, Audi Traffic Information Online) in Internet browsers (e.g., Google Traffic) or in smartphone apps (e.g., INRIX Traffic).

Backend-based systems can also report local dangers such as accidents or slippery road surfaces to a central data processing system. Local dangers are either detected automatically by in-vehicle systems or manually reported by the driver who selects a specific event from different categories (e.g., accident, animals on the road, or wrong-way driver).

When a car sends a request of “danger” information to the central processing system, the current position of the car is part of this message. Thus, the central backend system is able to filter and transfer only the relevant information among all available. Such a service is known as “location-based service.” That is why the current applications of web-based services in vehicles mostly focus on navigation and local danger warnings.

2 What Are Backend-Based Systems?

In the following sections, the backend-based system is defined as a client–server system, where cars are the clients whereas the backend corresponds to a server system, to which the clients are connected via Internet. The clients in backend-based driver assistance systems combine two different functionalities. On the one hand, they deliver data to the functionalities provided by the backend, and on the other hand, they receive information from the backend and use it to assist the drivers. In the backend, the data delivered from the clients is aggregated, and relevant information is extracted. An example is the real-time estimation of actual traffic flow. Apart from data from clients, a backend system can also use information from other sources. Task the traffic flow estimation as an example; in addition to the client data, the central backend system can get more relevant information by connecting to traffic control centers and other mobile clients (e.g., mobile phones,

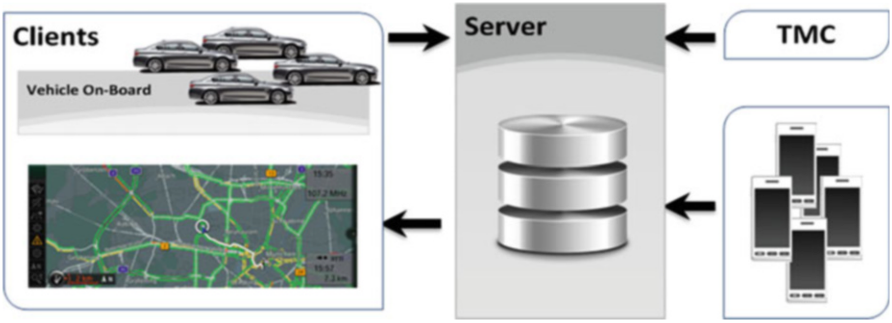


Fig. 1 Client–server systems for realtime traffic information (Source: BMW Group)

mobile navigation systems), which can provide extra valuable information to the system. Figure 1 illustrates the basic structure of a backend-based driver assistance system. The central part of the system includes a transmitter and receiver unit, a central server system, as well as a digital map (see ▶ Chap. 26, "Digital Maps for ADAS") in the vehicle, where the information from the backend is stored.

2.1 Digital Maps

Digital maps are the foundation of navigation systems (see ▶ Chap. 54, "Navigation and Transport Telematics"). These maps provide information to driver assistance systems by calculating the electronic horizon based on current positions (Blervaque et al. 2006). Using the horizon information, assistance functions can forecast the oncoming road situation. Hence, to gain an extended foresight, the digital map plays a crucial role in backend-based driver assistance systems.

In a client–server system, extra information from the backend can be attached to the map and be distributed to the assistance systems, in addition to the distribution of the electronic horizon. For example, in autonomous driving research projects (see ▶ Chap. 62, "Autonomous Driving"), high-precision maps are used in most systems. Thus, existing digital maps have to be enriched with more details, e.g., stop-line positions and traffic regulations. Through the connection to a backend, it is possible to detect and transmit deviations between a digital map and the real road situation. This knowledge transfer establishes the basis for a better actuality of digital maps. Therefore, the backend connection is beneficial for the quality assurance of many driver assistance systems up to autonomous driving.

2.2 Server Technologies

In a central server system, the so-called backend information of the vehicles is collected and stored in a database to enable efficient processing. Considering the example of reported local dangers, each piece of collected information includes the absolute position and the global time of the reported event. Such data with a position reference are called “geographical referenced data.” A computer system which captures, manages, analyzes, and depicts such location-related data is defined as a “geographic information system (GIS)” (Bill 2010).

2.2.1 Spatial Databases

A particular part of a GIS is the database which stores georeferenced data – also known as the spatial database. There are special kinds of relational databases which are suitable for this task, namely, the spatial relational databases. For instance, PostGIS (Obe and Hsu 2011) is a popular software, which expands the open source database PostgreSQL by spatial objects and requests.

A PostGIS database can store different geometries like a point, a polyline, or an area, as well as other types derived from these basic types. In the example of local danger warning, as the observation always occurs at a specific location, the information is represented as a point. Therefore, a position coordinate is added to each event.

Apart from the storage of objects, a spatial database can also define additional geometric objects based on specific requirements. For example, to get all relevant reports surrounding a specific vehicle, a rectangular box is defined as a geometry instance, where the current position of the vehicle appears as the center of the box. Through a query to the database, it is possible to get all events within a certain period of time. To accelerate such queries, an index on the geometric objects in the database can be generated.

In the following sections, we introduce a minimalistic example of a local danger warning system, to illustrate the application of standard backend technologies. The architecture of the application is shown in Fig. 1. A computer installed with the software PostgreSQL (www.postgresql.org) and PostGIS (www.postgis.net) is the precondition to run the minimalistic example. The computer represents the backend server in this case. To manage the database, either a tool with a graphical user interface (e.g., pgadmin) is used or PostgreSQL is started via the command line (pgsql.exe). The following shows the commands in the command line–based interface.

First of all, we create a new PostgreSQL database on the server computer and connect the chosen database management tool to this database. Secondly, this database is extended to a geospatial database using PostGIS.

```
– Creation of a new database
CREATE DATABASE local_hazards_db;
– Establish a connection to the database
\c local_hazards_db;
– Extension to a geospatial database
CREATE EXTENSION postgis;
CREATE EXTENSION postgis_topology;
```

To illustrate a “local danger warning,” a table within a geospatial database is created. New events, which are identified by vehicles and sent to the backend, are stored in this table. In addition to the original data received, the position of the vehicle where the report is generated is converted to a PostGIS point geometry and added to the corresponding entry in the database. If the position is sent in WGS84 format, the correct SRID number is 4326. This number is the ID of the WGS84 reference system in the standard table SPATIAL_REF_SYS. This table is generated by PostGIS automatically during the extension process from a database to a geospatial database.

```
– Create a table
CREATE TABLE local_hazards_tab
(
id SERIAL PRIMARY KEY, –– Sequential ID of all entries
geom GEOMETRY(Point, 4326), –– PostGIS geometry
latitude double precision, –– WGS84 latitude
longitude double precision, –– WGS84 longitude
heading double precision, –– Heading of the vehicle to the
north
speed double precision, –– Velocity of the vehicle
hazard VARCHAR(128), –– Type of the reported danger
hazard_time bigint – Timestamp of the report
);
```

In the geospatial database, the column “geom” is the WGS84 coordinates (latitude, longitude), encoded as PostGIS geometry. The coordinates are transformed into a binary value. To accelerate read accesses to the database, a geospatial index is generated based on the “geom” column.

```
– Add a spatial index
CREATE INDEX local_hazards_idx ON local_hazards_tab USING
GIST(geom);
```

Subsequent to these steps, the backend is ready to receive events from vehicles and store them in the database. An artificial data set of events is generated for this example. The events include general dangerous locations, slippery road, and deer crossing. The data set is shown in Fig. 2.

If a new event is identified by a vehicle, this information is sent to the backend via Internet. A correspondent program, which receives the data, writes the event to the database. As an example, a slippery road event is stored with the following command:

```
INSERT INTO local_hazards_tab
(geom, latitude, longitude, heading, speed, hazard,
hazard_time)
VALUES (
ST_SetSRID(ST_MakePoint(11.695927, 48.333459), 4326),
–– geom
48.333459, –– latitude
11.695927, –– longitude
```

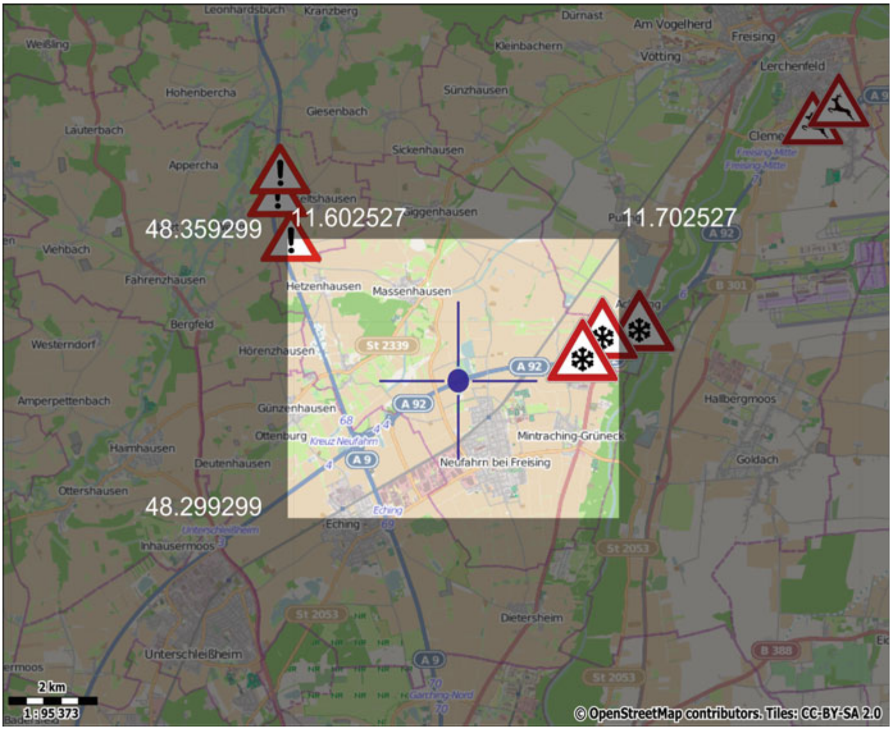


Fig. 2 Artificial example of reported events from different vehicles for a local danger warning (Source: BMW Group)

```
    0.2143, -- heading
    41.34, -- speed
    'Slippery Road', -- hazard
    1392675237); -- hazard_time
```

Corresponding to the reported events in Fig. 2, a database is built, see Table 1.

Based on this database, a vehicle equipped with local hazard warning systems can request critical events that happened in the local environment. In this case, the position of the vehicle is marked with a blue cross, as shown in Fig. 2. This position (11.652527° longitude, 48.329299° latitude) is sent to the backend. In the backend database, all events within a defined window (±0.05° longitude, ±0.03° latitude) around the position of the vehicle are queried. Usually, the size of this window is defined in meters, and the coordinates of the edges are transformed to WGS84. This is because WGS84 coordinates have different resolutions, depending on the absolute position on earth. In this example, we simplified this procedure by using a window in WGS84 coordinates.

```
  - Vehicle position: 11.652527° longitude, 48.329299°
latitude
```

Table 1 Content of the database – local danger events

Id	Geom	Latitude in °	Longitude in °	Heading in rad	Speed in m/s	Hazard	Hazard_time in s
1	0101000020E610000089. . .	48.333459	11.695927	0.2143	41.34	Slippery Road	1392675237
2	0101000020E6100000E8. . .	48.334643	11.701622	0.7624	39.56	Slippery Road	1392679267
3	0101000020E6100000661. . .	48.340933	11.712878	0.9538	26.89	Slippery Road	1392686384
4	0101000020E6100000EF. . .	48.385613	11.772187	0.1455	38.24	Deer Crossing	1392729461
5	0101000020E61000004A. . .	48.384673	11.769054	3.4907	32.72	Deer Crossing	1392762418
6	0101000020E610000023. . .	48.37347	11.595891	1.5345	23.46	Danger	1392660319
7	0101000020E6100000CD. . .	48.374782	11.596191	1.5323	34.58	Danger	1392664163
8	0101000020E610000008. . .	48.357789	11.597178	1.7256	13.56	Danger	1392669230

```
- Query window: ±0.05° longitude, ±0.03° latitude
SELECT * FROM local_hazards_tab
WHERE ST_Contains(ST_SetSRID(ST_MakeBox2D(
ST_Point(11.602527, 48.299299),
ST_Point(11.702527, 48.359299)),4326),geom);
```

This query yields two reported events in the local surrounding of the vehicle, as shown in Table 2. These events are then sent to the vehicle as a result. An assistance system in the vehicle receives the two events' data and evaluates the reports. The detected dangers are filtered according to the difference in the headings of the event and the vehicle to guarantee the relevance. Furthermore, the decision whether a warning is sent to the driver or not is based on the number of reported events and the timestamp of the reports. Thus, if there is potential for danger, the driver receives a warning.

2.2.2 Scalable Architectures for Geospatial Data Analysis

In the future, the amount of information to be transmitted and stored will increase further as a result of more vehicles will be connected to the Internet.

The challenge of quickly growing amounts of data has already raised concerns during the development of the Internet. As a consequence, file systems with redundancy and scalable performance, fail-safe database systems, and efficient concepts to process the data are necessary. Commonly used technologies in this field have been developed by the company Google Inc. To persistently save data, Google developed the "Google File System (GFS)" (Ghemawat et al. 2003) as well as the constitutive database concept "Big Table" (Chang et al. 2006).

Let us take a scalable programming model introduced by Google Inc., "MapReduce" (Dean and Ghemewat 2004), as an example. The basic concept of this model is the partitioning of the data analysis in the so-called Map and Reduce phases. In the Map phase, all input data are divided into different parts. For every part, a Map process is started. Every Map process processes the corresponding part of the data independent of other Map processes. Thus, all the processes can be executed in parallel. Every single process produces intermediate results, and the results are then integrated to a final result in the Reduce phase. In this phase, it is possible to start an independent Reduce process for every final result.

The described concepts have been implemented and extended in free open-source software by the "Apache Software Foundation" to a complete framework, "Hadoop" (White 2009), which is freely available. The main component is constituted of the file system "Hadoop Distributed File System (HDFS)"; the database concepts "Pig," "HBase," and "Hive"; as well as a MapReduce implementation.

These basic technologies are used in many different applications, for example, in Facebook. The MapReduce concept and "Apache Hadoop" are also applied in geospatial projects. Different extensions have been developed ever since, like "MRGIS" (Chen et al. 2008), "Hadoop-GIS" (Aji et al. 2013), and "SpatialHadoop"

Table 2 Result of the event request from a vehicle

id [−]	Geom [−]	Latitude [°]	Longitude [°]	Heading [rad]	Speed [m/s]	Hazard [−]	Hazard_time [s]
1	0101000020E610000089…	48.333459	11.695927	0.2143	41.34	Slippery Road	1392675237
2	0101000020E6100000E8…	48.334643	11.701622	0.7624	39.56	Slippery Road	1392679267

(Eldawy and Mokbel 2013). These developments combine the scalable architectures with specialized solutions for geospatial data, the geographic information systems. This progress constructs the base for highly scalable geospatial applications, which is highly important for the automotive industry in the future.

2.3 Transmitter Unit in the Vehicle

The telematic component (see ▶ Chap. 54, "Navigation and Transport Telematics") enables the connection of a vehicle to a backend system. This component can be built as an individual control unit or integrated into the head unit. The implementation is called "telematic control unit (TCU)" (BMW, Daimler). TCU requires connection to the Internet and access to the sensor data of the vehicle as well as to the current geographic position.

There are two ways to enable the Internet connection:

- A fixed SIM card with a surcharge for the data rate (e.g., BMW, Daimler)
- A slot for a SIM card owned by the vehicle user (e.g., Audi, Daimler)

The TCU is connected to the vehicle buses to get the access to the sensor data. The current position is delivered via GPS by the navigation system. With such a component, information like detected local dangers can be transmitted and delivered via the Internet.

3 Data Transmission Properties

Data communication between clients and servers is characterized by a multilevel architecture. The first step in the process is to capture data from different sensors. The second step is to preprocess and merge these data in order to provide a wider knowledge concerning both the vehicle and the objects in the local environment. This object-oriented modeling comprises static objects (e.g., lane markings), dynamic objects (e.g., other traffic participants), and metadata (e.g., speed limit). The combination of this information creates a new knowledge. For instance, if the current vehicle speed is significantly below the speed limit, this suggests relevant information for traffic disruption detection. This data analysis takes place especially at the frontend, e.g., at the vehicle. The results are transmitted via cellular network and IP connection to the backend server Fig. 3). Next, the backend server processes

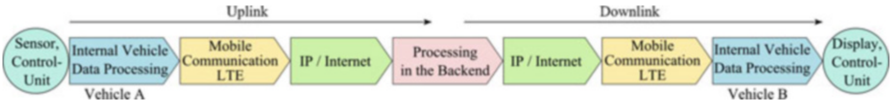


Fig. 3 Multi-level architecture to exchange data between vehicles (frontends) and backend servers (Source: BMW Group)

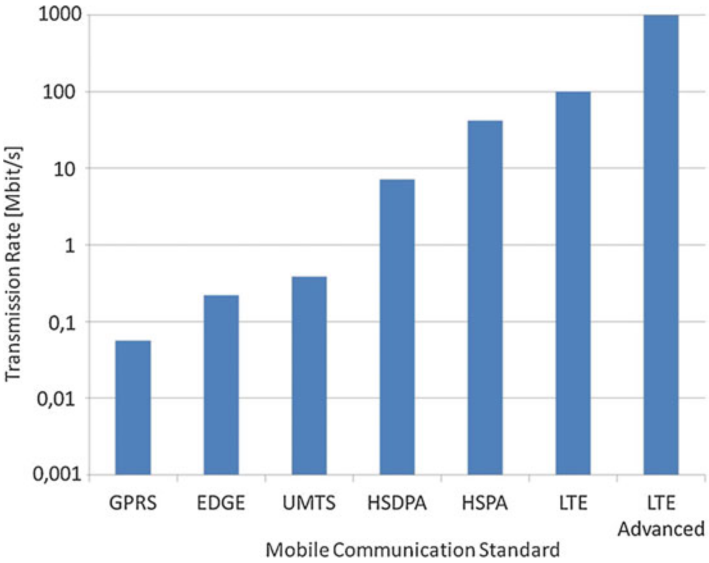


Fig. 4 Available communication standards and their theoretical transmission rate (Source: BMW Group)

the information from both vehicles and additional metadata, for example, combining the reported accidents and areas of roadworks. Results of this backend-based information fusion are then returned to the vehicles to support in-vehicle applications. It needs to be noted that the body of acquired knowledge is customized to the demand of the driver; e.g., through a traffic disruption ahead, the time traveled would increase by 23 min.

In order to reduce the cost of data transmission, it is necessary to restrict the amount and frequency of data transmitted. Therefore, it is strongly recommended to transmit abstracted knowledge instead of raw data. Furthermore, a decentralized knowledge management system between vehicles and backend servers can provide the possibility of exchanging data only if new information emerges. A comparison of the anticipated vehicle status (e.g., vehicle in a traffic jam) and the actual status (e.g., current vehicle speed close to speed limit) often identifies an information gap. Where deviation occurs, data will be exchanged, and the backend server will get updated information from the vehicles. In the demonstrated example, vehicles act as mobile sensors, which provide a better understanding of the current on-road situation.

To gain backend-based knowledge which is up to date, inter alia a fast data transmission is necessary. Theoretical transmission rates of different mobile communication standards are shown in Fig. 4.

Highest data transmission rates are provided by LTE (Long-Term Evolution, 3.9G) with up to 100 Mbit/s and LTE Advanced (4G) with up to 1000 Mbit/s. The properties of the multilevel architecture's latency have been analyzed in Bartsch

			Optimist	Realist (Expectation)	Pessimist
↓ UPLINK	Vehicle	CAN-Bus	< 5 ms	< 10 ms	< 30 ms
		CAN-Converter	< 0,04 ms	< 0,07 ms	< 0,4 ms
		Buffering and Calculation of the Environment Model	< 30 ms	< 100 ms	< 400 ms
		Ethernet with Switch	< 0,3 ms	< 0,3 ms	< 0,3 ms
		PC and USB-Bus	< 0,5 ms	< 1 ms	< 3 ms
	Mobile Communication LTE	Transition Idle-Mode to Active-Mode	-	< 0,35 ms	< 102 ms
		LTE Terminal (LTE USB-Stick)	< 1,5 ms	< 5 ms	< 7 ms
		LTE Radio Access Network (RAN, Wireless Interface)			
		Urban Scenario, mid load	< 11 ms	< 12 ms	< 40 ms
		Urban Scenario, high load	< 11 ms	< 12 ms	< 48 ms
		Rural Scenario, mid load	< 12 ms	< 16 ms	< 83 ms
		Rural Scenario, high load	< 18 ms	< 29 ms	< 240 ms
		Motorway Scenario, mid load	< 28 ms	< 54 ms	< 340 ms
		Motorway Scenario, high load	< 39 ms	< 80 ms	< 600 ms
		Handover between radio cells	-	< 0,001 ms	< 25,5 ms
		Handover failure handling	-	< 0,001 ms	< 110 ms
		LTE Evolved Packet Core (EPC, Core-Net)	< 15 ms	< 20 ms	< 25 ms
	IP	Transmission over the internet	< 8 ms	< 30 ms	< 500 ms
	Backend	Autentification and authorization	< 1 ms	< 2 ms	< 5 ms
		Communication middleware	< 0,1 ms	< 0,1 ms	< 0,2 ms
		Function / Application	< 1 ms	< 2 ms	< 5 ms
DOWNLINK		Autentification and authorization	< 0,1 ms	< 0,1 ms	< 0,2 ms
		Communication middleware	< 1 ms	< 2 ms	< 5 ms
	IP	Transmission over the internet	< 8 ms	< 30 ms	< 500 ms
	Mobile Communication LTE	LTE Evolved Packet Core (EPC, Core-Net)	< 15 ms	< 20 ms	< 25 ms
		LTE Radio Access Network (RAN, Wireless Interface)			
		Urban Scenario, mid load	< 3 ms	< 11 ms	< 14 ms
		Urban Scenario, high load	< 4 ms	< 12 ms	< 15 ms
		Rural Scenario, mid load	< 7 ms	< 13 ms	< 52 ms
		Rural Scenario, high load	< 13 ms	< 28 ms	< 12000 ms
		Motorway Scenario, mid load	< 3 ms	< 11 ms	< 20 ms
		Motorway Scenario, high load	< 3 ms	< 12 ms	< 7000 ms
		Transition Idle-Mode to Active-Mode	-	< 0,35 ms	< 102 ms
		Handover between radio cells	-	< 0,001 ms	< 25,5 ms
		Handover failure handling	-	< 0,001 ms	< 110 ms
		LTE Terminal (LTE USB-Stick)	< 1,5 ms	< 5 ms	< 7 ms
↓	Vehicle	PC and USB-Bus	< 0,5 ms	< 1 ms	< 3 ms
		Ethernet with Switch	< 0,3 ms	< 0,3 ms	< 0,3 ms
		Buffering and Calculation of the Environment Model	< 30 ms	< 100 ms	< 400 ms
		CAN-Converter	< 0,04 ms	< 0,07 ms	< 0,4 ms
		CAN-Bus	< 5 ms	< 10 ms	< 30 ms
		Expectation value End-to-End latency: (for LTE rural scenario, mid load)		**< 369 ms**	

Fig. 5 Multi-level architecture's latency properties estimation for the communication channel: Vehicle to backend-server to vehicle (Source: BMW Group)

et al. (2012). In order to investigate the performance of different architecture layers, a prototypical data link between a vehicle (frontend) and a backend server as well as an LTE simulation environment is constructed. Six different scenarios are analyzed based on simulation studies (Lottermann et al. 2012), in Urban, Interurban, and Motorway, with average and high network load in each case. Measurements and simulation results are summarized in Fig. 5. These numeric values show the time range for data transmission. The optimistic results indicate the highest theoretical possibility, and the realistic results show the anticipated transmission rate on average. Higher transmission rates can occur depending on the current traffic density and the condition of the communication network. In Fig. 5, these results are labeled as pessimistic. Latency variability of hundreds of milliseconds has to be taken into account if functions use data provided via mobile communication.

The analysis shows that the data exchange based on LTE in the urban scenario with an average network load is characterized by an end-to-end latency up to 369 ms.

4 Next-Generation ADAS

Data connection between vehicles and backend servers provides large potentials to enable new functions and services in advanced driver assistance systems. The data connections between vehicles and backend servers provide dynamic traffic information, position of hidden queues, and identification of incorrect map information which can also enhance existing ADAS. For example, the backend can provide detailed traffic information, e.g., the position of a hidden queue. As described in Klanner et al. (2013) and depicted in Fig. 6, although a hidden queue is already handled safely by a highly automated driving vehicle even without backend information, the limited detection range leads to strong braking maneuvers. The hidden queue information provided by backend systems can largely enhance the range of detection and can guide vehicles operated in highly automated driving mode to approach the queue smoothly.

Another example is the traffic sign recognition. In traffic sign recognition, data from video cameras and maps are merged to enhance the accuracy. However, errors can occur due to incorrect or outdated map information, or environmental influences like fog or rain which reduce the camera's detection performance. To enhance the robustness of detection, backend information can be used in addition. The backend server can observe the current traffic flow, and identify and update incorrect map information.

Additional map attributes are important inputs for urban adaptive cruise control and urban highly automated driving. For example, the stop line position is an important input for advanced automatic engine start/stop control, which tries to minimize the number of unnecessary stops of engines. If a vehicle stops at a four-way stop, it is more likely to be a short stop, and with this information, advanced automatic engine start/stop control systems can prevent the engine to switch off. However, it is challenging to detect stop lines with high reliability and low false detection rates with a camera, because only a few features are available in a camera picture. Backend-based knowledge can help to validate and merge camera-based stop line detections, as this type of information can be learned based on artificial intelligence methods. Thereby, false detections can be prevented.

5 Extracting Driver Assistance–Related Information from Fleet Data

Many vehicles connected to backend servers provide a huge number of georeferenced data. In order to extract information which is relevant for driver assistance, an adequate architecture is necessary to gather and manage data, see Sect. 3.2. as well

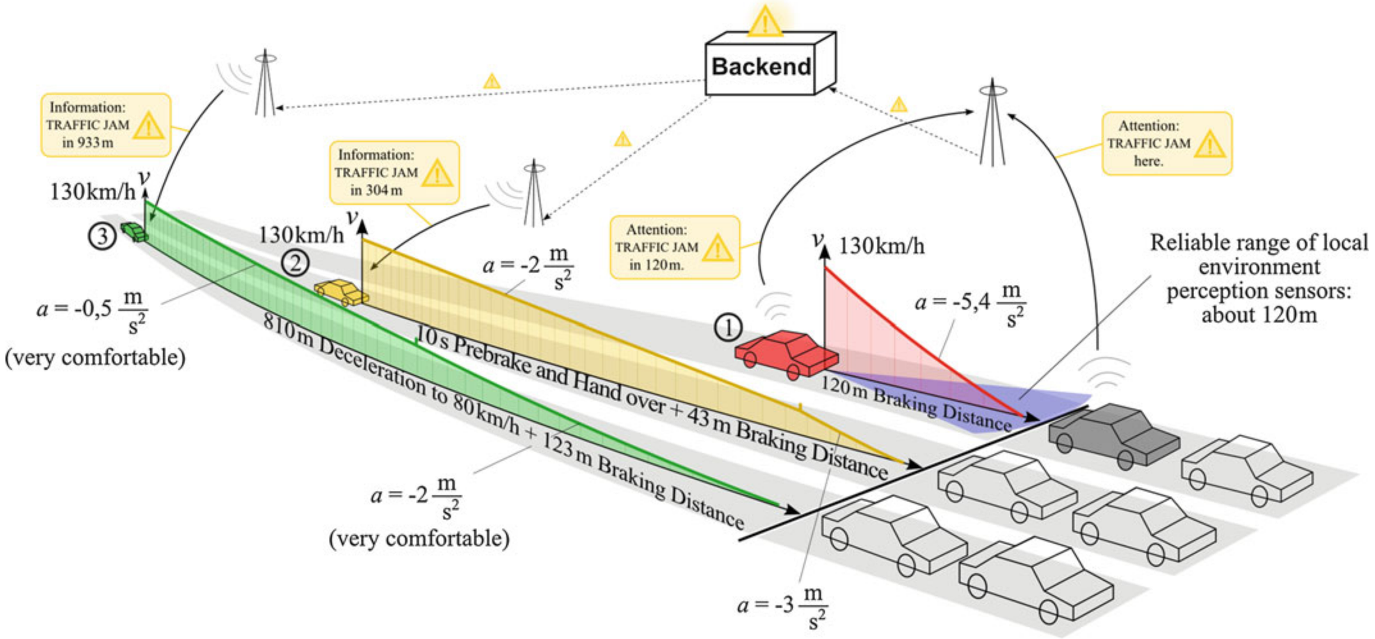


Fig. 6 Highly automated driving vehicle approaches to a hidden queue (Source: BMW Group)

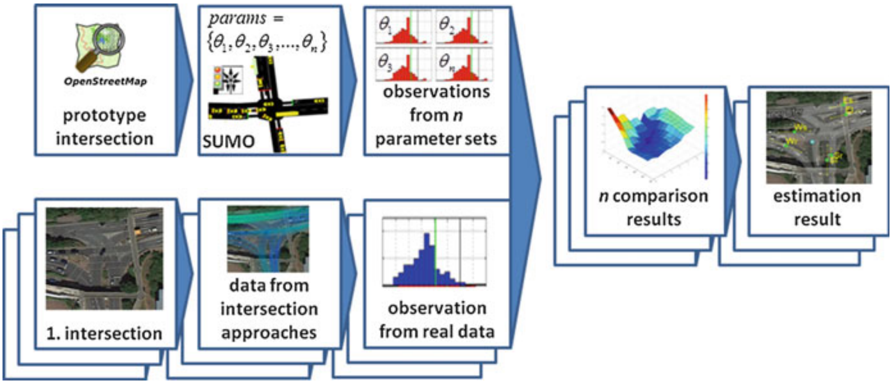


Fig. 7 Generic acquisition of intersection parameters based on car fleet data (Source: BMW Group)

as algorithms to analyze and georeference these data. The algorithms for data analysis use distributed systems, whose structures depend on the use case.

In the case of "Local Danger Warning," the main part of data processing typically is frontend based. Hidden queues are detected by vehicle-based algorithms, and their positions are transmitted to backend servers. After verifying the hidden queue reports, backend servers distribute the warnings to other vehicles. After this, the relevance of a local danger warning is tested by each vehicle individually.

In contrast, driver assistance systems which require additional map information collect and analyze data in backend servers. For example, fleet data can be used to extract intersection parameters (Ruhhammer et al. 2014a). Figure 7 shows the steps of this method. The algorithm is based on the specific fleet data values, which are related to the intersection parameters. A microscopic traffic simulation at each intersection is executed with the required intersection parameters. At the same time, the unknown intersection parameters are varied in a discrete value range and the traffic scene of the intersection simulated based on the combination of each parameter. Analogous to the fleet data, the same values are observed based on the simulation of each parameter combination. The parameter combination which is characterized as the best one is identified by the comparison of simulation-based observations and fleet measurement data. Apart from exact parameter estimations, driver assistance systems require the estimation of a confidence region. A method to estimate the intersection parameters' confidence can be found in Ruhhammer et al. (2014b). Most calculation is conducted in backend servers due to the high computing capacity these algorithms require.

6 Conclusions

The connection of vehicles to backend systems enables a lot of possibilities to deliver more information about the driving environment to drivers as well as to the in-vehicle systems. In addition to the existing entertainment services via Internet, we face an increasing demand toward the connectivity for driver assistance systems in the future. The possibility to send latest information about dangers or to create high-precision maps form the data sent by on-road vehicles points out some interesting future applications. The investigation of the capability of the current 4G wireless technology shows that the transmission latency with around 400 ms is low enough to enable real-time applications. However, the technology has to be improved further on the stability of this latency. The progress in technologies for data analysis and data transmission will enable the development of new systems that assist the drivers up to highly automated driving.

References

Aji A, Sun X, Vo H, Liu Q, Lee R, Zhang X, Wang F (2013) Demonstration of Hadoop-GIS. ACM Press, New York

Bartsch A, Klanner F, Lottermann C, Kleinsteuber M (2012) Latenz-Eigenschaften prototypischer Datenverbindungen zwischen Fahrzeugen über Backend (Latency properties of prototypic data connections between vehicles over a backend). Forschungspraxis, TUM Lehrstuhl für Medientechnik, München

Bill R (2010) Grundlagen der Geo-Informationssysteme (Basics of geo-information systems). Wichmann, Heidelberg

Blervaque V, Mezger K, Beuk L, Loewenau J (2006) ADAS Horizon – how digital maps can contribute to road safety. In: Valldorf J, Gessner W (eds) Advanced microsystems for automotive applications 2006. Springer, Berlin, pp 427–436

Chang F, Dean J, Ghemawat S, Hsieh WC, Wallach DA, Burrows M, Chandra T, Fikes A, Gruber RE (2006) Bigtable: a distributed storage system for structured data. USENIX Association

Chen Q, Wang L, Shang Z (2008) MRGIS: a MapReduce-enabled high performance workflow system for GIS. In: 2008 I.E. fourth international conference on eScience. IEEE

Dean J, Ghemewat S (2004) MapReduce: simplified data processing on large clusters. USENIX Association

Eldawy A, Mokbel MF (2013) A demonstration of SpatialHadoop: an efficient mapreduce framework for spatial data. In: Proceedings of the VLDB Endowment

Ghemawat S, Gobioff H, Leung S (2003) The Google file system. ACM Press, New York

Klanner F, Ruhhammer C, Bartsch A, Rasshofer R, Huber W, Rauch S (2013) Mehr Komfort und Sicherheit durch zunehmende Vernetzung (More comfort and safety through increasing connectivity). WEKA FACHMEDIEN GmbH, Elektronik automotive 6/7.2013

Lottermann C, Botsov M, Fertl P, Müllner R (2012) Performance evaluation of automotive off-board applications in LTE deployments. In: 2012 I.E. vehicular networking conference (VNC 2012), Seoul, pp 211–218

Obe R, Hsu L (2011) PostGIS in action. Manning Publications, Shelter Island

Ruhhammer C, Atanasov A, Klanner F, Stiller C (2014a) Crowdsourcing als Enabler für verbesserte Assistenzsysteme: Ein generischer Ansatz zum Erlernen von Kreuzungsparametern (Crowdsourcing as an enabler to improve driver assistance systems: a generic approach to learn parameters of intersections). In 9. Workshop Fahrerassistenzsysteme: FAS2014, Walting. Uni-DAS e.V

Ruhhammer C, Hirsenkorn N, Klanner F, Stiller C (2014b) Crowdsourced intersection parameters: a generic approach for extraction and confidence estimation. IEEE intelligent vehicles symposium

White T (2009) Hadoop: the definitive guide. O'Reilly Media, Sebastopol

Part VI

Actuation for DAS

Hydraulic Brake Systems for Passenger Vehicles

29

James Remfrey, Steffen Gruber, and Norbert Ocvirk

Contents

Abstract

Within the standard architecture, hydraulic brake systems of passenger vehicles must decelerate the vehicle according to the driver's request and to legal requirements (i.e., ECE R13H). Wheel forces generated during braking are transferred via tires to the road surface in such a way that the vehicle remains stable and controllable and always follows the driver's intention. The basis for this is optimized pedal feel and optimized distribution of brake forces left/right and front/rear.

Architectures can be extended to influence fuel consumption and emissions. The combination of internal combustion engines and electric machines

---

J. Remfrey (✉)
Continental Corporation, Frankfurt am Main, Germany
e-mail: james.remfrey@continental-corporation.com

S. Gruber • N. Ocvirk
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: steffen.gruber@continental-corporation.com;
norbert.ocvirk@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_30

("hybrid drive") as well as electric drive is becoming more widespread in passenger vehicles. Coupling of electric machines and drivetrain generates electric power by brake energy recuperation. The impact on brake system design is to offer the same pedal feel, independent of whether the vehicle is braked by an electric machine and/or by friction brakes (brake blending).

Electronically controlled hydraulic brake systems (e.g., ABS, TCS, ESC) optimize vehicle dynamics. Together with beam and image sensors, this opens various opportunities to utilize additional brake system functions, i.e., for advanced driver assistance systems (ADAS), to fulfill future vehicle safety requirements. The performance of these advanced assistance systems mainly depends on vehicle system and component layout, hardware, software, sensors, and HMI.

1 Standard Architecture

The task of hydraulic brake systems is to decelerate a car safely in accordance with the driver's wishes while conforming to minimum legal requirements (e.g., ECE R13H) (Breuer and Bill 2013). The brake forces exerted at the wheels should be transmitted to the road via the tires so that the car always follows the driver's intention. This presupposes distribution of brake forces between the front and rear axles and the right and left sides of the car. The distribution of brake forces is governed by regulations which automakers and their suppliers are required to follow.

Standard brake systems boost the force that a driver exerts on the brake pedal so that it will be sufficient to act on the wheel brakes. They also control the distribution of brake forces to the axles, depending on the design. Additionally mechanical "brake force distributors" can help to change brake force distribution as a function of load on a passenger car.

The introduction of electronically controlled hydraulic brake modulators aided by wheel speed sensors (such as ABS and ESC) means that it is now possible to control brake torque to each individual wheel brake and, if necessary, independently of the driver.

This opens up a number of possibilities for utilizing the brake system, for instance, driver assistance functions that extend way beyond pure brake functions.

The brake pedal serves as the human-machine interface in passenger cars (also in view of legal requirements).

Today's hydraulic brake systems (Fig. 1) therefore contain components that cover the following functions:

- Initiate (foot) brake force
- Boost brake force
- Convert brake force into brake pressure/volume flow
- Transfer pressure/volume
- Convert pressure/volume into brake force at wheels

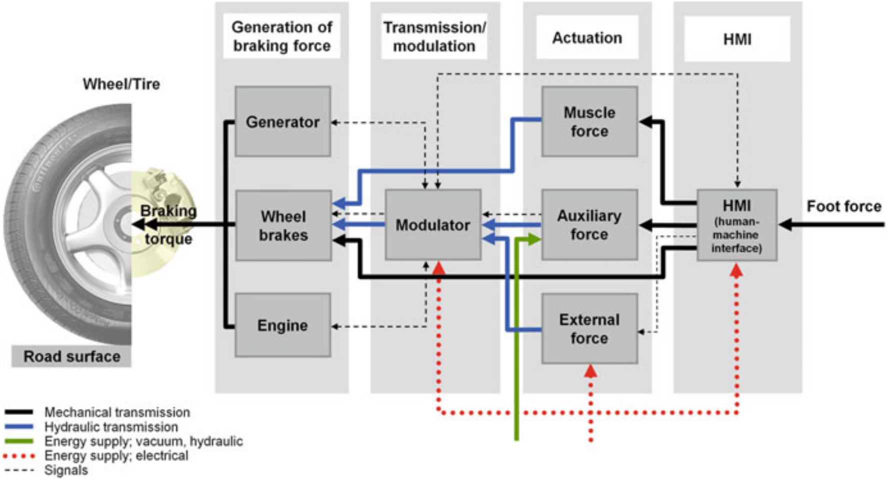


Fig. 1 Chain of effects, architecture for hydraulic brake systems in cars

Modulation of brake force in hydraulic systems is achieved by placing a modulator between brake actuator (booster) and generation of brake force at the wheels. Sensors, moreover, that monitor vehicle behavior are used in modulating brake forces.

Components Among the typical components of passenger car brake systems are:

- Pedal box with brake pedal
- Brake actuator (booster, (tandem) master cylinder, reservoir)
- Electronic Brake System (EBS) with sensors:
  - Wheel speed sensors at all four wheels
  - Acceleration sensors
  - Yaw rate sensor
  - Steering angle sensor
- Brake calipers and disks
- Drum brakes
- Brake lines and hoses
- Tires

1.1 Actuation

Four main components comprise the actuator: the brake booster, the (tandem) master cylinder, the brake-fluid reservoir, and the brake lines/hoses.

Brake Booster Brake boosters boost the force that a driver's foot exerts on the brake pedal. In a vacuum booster this is done by generating boost force from the

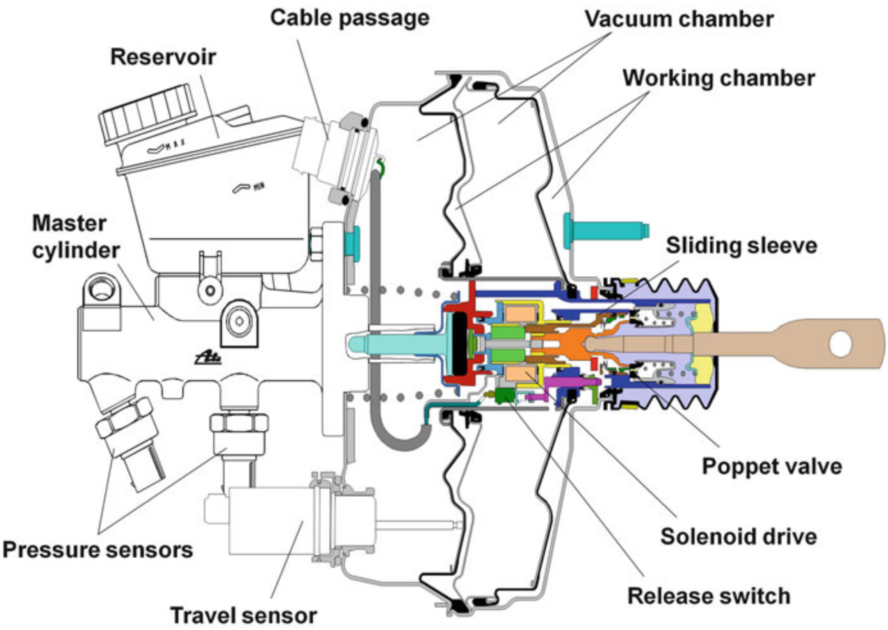


Fig. 2 Active brake booster in tandem design

differential between atmospheric pressure and vacuum (see below). Brake boosters thus provide increased comfort and safety. Three main types of booster find application:

- Vacuum brake boosters
- Hydraulic brake boosters
- Electromechanical brake boosters

Vacuum Brake Boosters Vacuum brake boosters have prevailed over hydraulic brake boosters in spite of their much larger size. The main reasons for this are that they are inexpensive to manufacture and normally aspirated engines produce vacuum energy that is readily available.

The suction chamber of the brake booster is connected to the engine's intake manifold by a vacuum line or a separate vacuum pump, as is the case with diesel engines, direct injection, or charged gasoline engines.

The size of the booster is normally given in inches. Common sizes run between 7″ and 11″. The working capacity of these single devices, however, is insufficient for larger cars. Tandem brake boosters do duty here, meaning two single devices are arranged, one behind the other, and integrated into one device. Common sizes run from 7″/8″ to 10″/10″. Figure 2 shows such an electronically controllable tandem brake booster design.

Increasing fuel economy (especially through reduction of intake throttle losses) has led to a continual reduction of the vacuum available to operate the brake booster. One obvious countermeasure would be to increase the size of the booster. This however incurs packaging problems in the engine compartment. The second countermeasure would be to install a vacuum pump.

The following describes two special variants of vacuum brake boosters in use in addition to the conventional design described above.

(a) Active Brake Booster
 Active brake boosters provide assistance and additional functions. These functions are independent of driver input, are electrically controllable, and thus generate brake pressure (see Fig. 2). Active brake boosters support such functions as pre-charging ESC, brake assistant (BA) function, and adaptive cruise control (ACC). Active brake boosters feature a solenoid trigger built into their control housing. A sliding sleeve (collar) activates the poppet valve with the electrically activated solenoid trigger. This initially closes the connection between the suction chamber and the working chamber; another electrical signal opens the working chamber to the outside air and activates the brake booster (independent of driver input). A release switch is built into the control housing to ensure correct detection of a driver's input.

(b) Mechanical Brake Assist
 This concept makes use of a modified mechanical design to utilize the inertia of the brake booster. When activated suddenly (in an emergency situation), it opens a poppet valve beyond a predefined aperture. This causes the poppet valve to remain open even if the foot force on the pedal diminishes (Fig. 3). Maximum brake boost is brought to bear and not just the amount provided by the driver's foot on the pedal.

Hydraulic Brake Booster Hydraulic brake boosters have advantages over vacuum brake boosters with regard to energy density and size. Their scope of application is chiefly for heavy vehicles (such as armored limousines) and electric cars.

Electromechanical Brake Boosters An electromechanical brake booster (Bosch 2013) with (partially) uncoupled brake pedal is practical, especially in view of the demands that hybrid and electric cars make on the brake system. These types of cars supply no vacuum at all to a vacuum brake booster or do so only periodically. Drivers, moreover, must not notice any differences in pedal feel while braking, such as might arise from the blending of generator and friction brake.

In an electromechanical brake booster (Fig. 4, photo Bosch), travel or angle sensors monitor whenever the brake pedal is activated. The ECU assesses this information and converts it into appropriate control signals for an electric motor. This motor enhances the force that a driver exerts on the brake pedal via a gearbox, which ultimately acts on the master cylinder that generates the hydraulic brake pressure. Software can influence the characteristic feel in the pedal that a driver

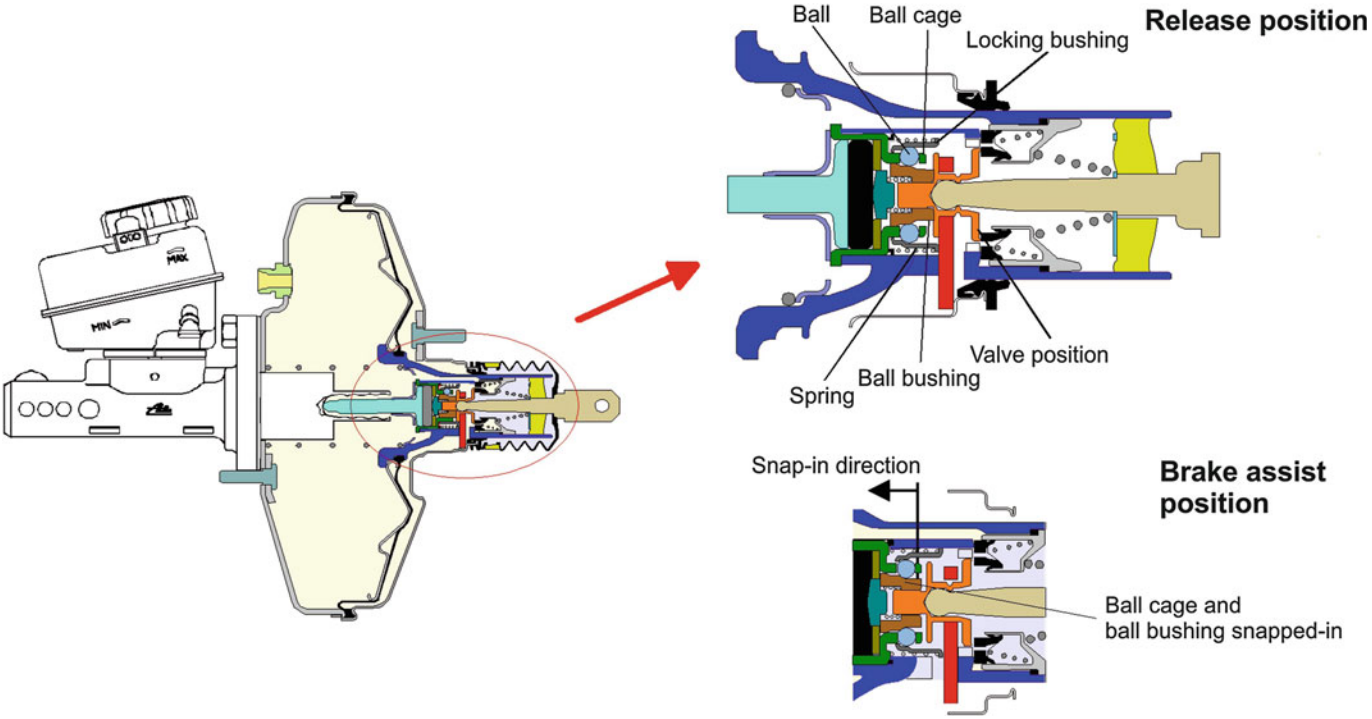


Fig. 3 Mechanical brake assist

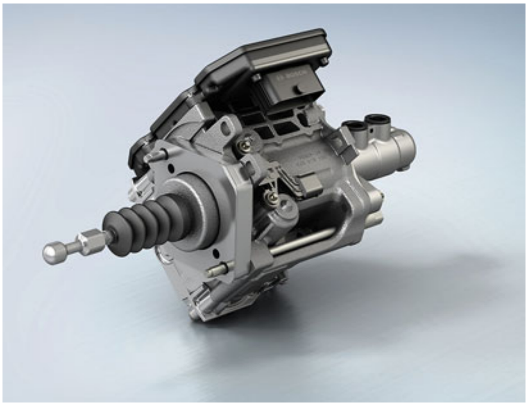


Fig. 4 Electromechanical brake booster – iBooster (Photo Bosch)

experiences with this approach. A secondary HECU for hybrid systems takes care of the blending mentioned above.

Brake Master Cylinder The task of the (tandem) master cylinder is to convert the mechanical force of the brake boost into hydraulic pressure. In addition to generating pressure against (limited) elasticity of the brake calipers and pads and overcoming the clearance, it has to supply the fluid volume.

Since legislation requires dual-circuit brake systems, single master cylinders typically are only installed in special applications such as racing cars.

The tandem master cylinder in general use today is actually a combination of two master cylinders in series incorporated in a single housing. It allows an increase and decrease of pressure in both brake circuits. If the volume changes in the brake system due to a change in temperature or due to worn pads, the brake-fluid reservoir compensates for this change in volume.

Brake lines apply pressure and hydraulic volume to the calipers in accordance with brake force distribution and then convert it to mechanical clamping force. The brake pads transfer the brake clamping force to the disks. The wheels and tires transmit the resultant (brake) torque to the road, with the car decelerating as a result.

Brake-Fluid Reservoir The brake-fluid reservoir compensates for additional necessary volume caused by pad wear, compensates for changes in volume within the brake system under different conditions, prevents air from getting into the system in different situations, reduces brake-fluid foaming, and separates the reserve volumes of the master-cylinder circuits if fluid level is low.

A brake-fluid reservoir can moreover serve as a reservoir for a hydraulically operated clutch or even an ESC pump. Moreover, it can contain the brake fluid necessary for pressurizing an accumulator.

Brake Fluid Brake fluid is the usual medium in the hydraulic section of the brake system for transferring energy between the (tandem) master cylinder or the hydraulic modulator and the wheel brakes. It also performs an additional function in that it lubricates moving parts such as gaskets, pistons, and valves and protects against corrosion. Brake fluid is hygroscopic: moisture from the atmosphere can cause the boiling point of brake fluid to deteriorate, which can lead to foaming and, hence, to problems if the brake system is subjected to high temperatures. This is why it is necessary to check the brake-fluid condition regularly to monitor whether the boiling point has dropped below a critical value (cf. prescribed intervals for changing brake fluid).

Brake fluid must also flow freely at low temperatures (down to −40 °C), i.e., it must exhibit a sufficiently low viscosity so as to permit good brake response during activation and release and also proper function of the electronic control systems. Moreover, brake fluid must exhibit a high boiling point so that it does not foam even under extremely high temperatures. The fact that air bubbles are compressible means that insufficient brake force would result due to the limited output volume of the (tandem) master cylinder. Brake fluids are either glycol or silicon based.

Brake Lines and Hoses High-pressure lines, hoses, and flexible reinforced tubing connect the hydraulic components of a brake system. The main requirements are that they can withstand pressure, can withstand mechanical wear, take up little space, and can withstand chemicals such as oil, fuel, and saltwater and that they are indifferent to temperature fluctuations.

1.2 Modulation

The HECU (hydraulic/electronic control unit) is connected to the brake circuits of the (tandem) master cylinder by two hydraulic lines. Brake lines lead from the hydraulic modulator to the wheel brakes.

Hydraulic/Electronic Control Unit (HECU) Today's ABS/TCS/ESC systems (such as the Continental MK60, Fig. 5) consist of a central hydraulic block with solenoid valves, an integrated piston pump powered by an electric motor (HCU = hydraulic control unit), and a bracket containing coils and electronics (ECU = electronic control unit). The coils contained in the bracket are aligned with the valve actuation in a layout known as "magnetic connector" concept.

The energy supply of an ESC unit consists of a dual-circuit hydraulic piston pump (pump cartridges incorporated in the valve block) driven by an electric motor with eccentric drive shaft. It releases excess brake fluid during ABS/TCS/ESC modulations from the low-pressure accumulators back into the brake circuits of the (tandem) master cylinder.

The intake and outlet valves integrated into the valve block are designed as 2/2 (two ports, two switching positions) solenoid valves. They permit modulation of wheel-brake pressure (see Fig. 6). The intake valve (normally open) fulfills two functions. It opens or closes the connection from the master cylinder to each wheel-brake circuit when activated in order to hold necessary brake pressure. The

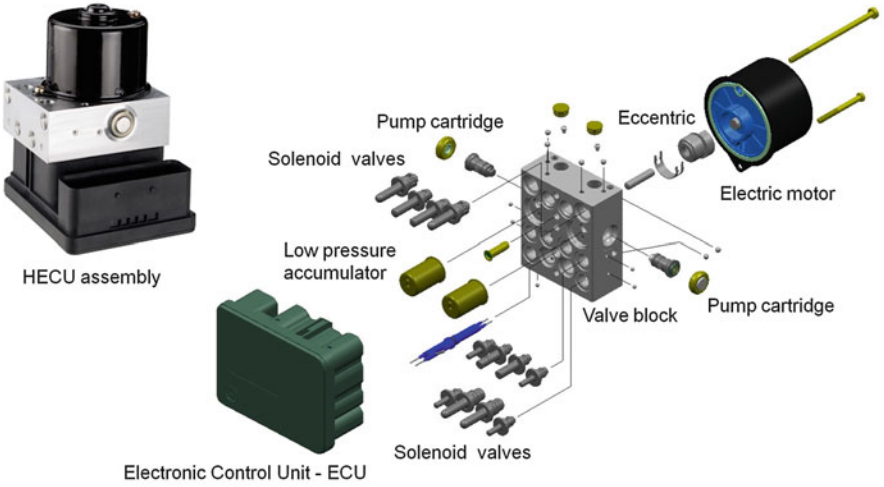


Fig. 5 Exploded view of an ESC system

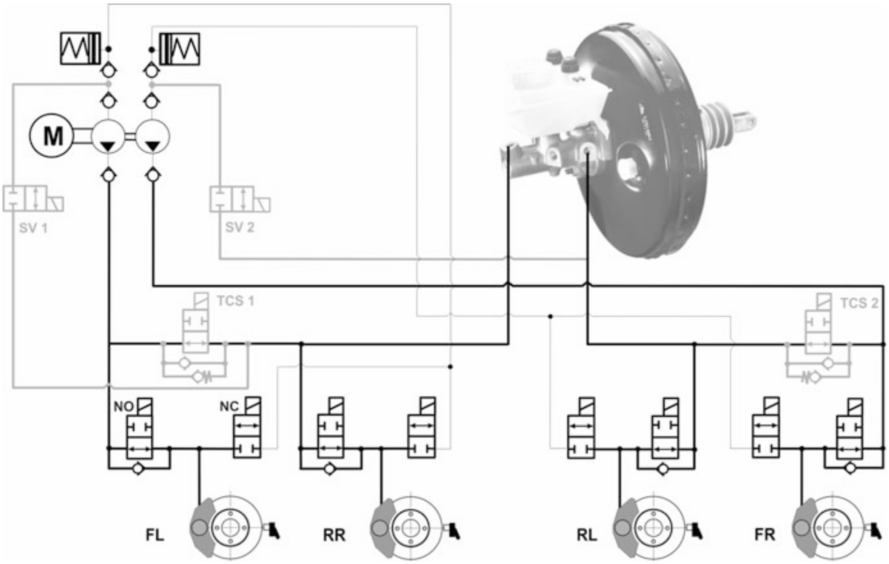


Fig. 6 ABS circuit diagram (*black*) with additional components for TCS (*gray*) for cars with front-wheel drive and diagonal brake circuit distribution

in-parallel check valve permits a reduction of brake pressure and, thus, of brake force, when the driver reduces foot force at the brake pedal, regardless of the status of the solenoid valve.

The 2/2 outlet valve (normally closed) opens the connection from the wheel-brake circuit to the low-pressure reservoir. Therefore, it permits a reduction in brake force to each individual wheel by reducing pressure.

The low-pressure accumulators serve to store brake fluid during brake pressure modulation. Each brake circuit requires one low-pressure accumulator.

Pump pulsation dampers are integrated to minimize any hydraulic noises and vibration (pedal feedback).

The HECU is connected with different sensors, depending on how it is configured and on the extent of its functions. It communicates, moreover, via bus systems such as CAN or FlexRay with other control devices found in a car. An array of safety and assistance functions constitute today's car brake systems with the aid of these mechatronic components. A discussion of these follows.

Electronic Brake Force Distribution (EBD) The task of electronic brake force distribution is to prevent the rear wheels from locking up while brake pressure is increasing. This keeps the car stable. If the wheel speed sensors detect increased wheel slip at the rear axle, the EBD function will limit any further increase in pressure to the rear axle by closing the inlet valves. As such, one can see EBD as a preliminary step to any potential ABS intervention.

Antilock Brake System (ABS) The brake force that corresponds to the friction coefficient is the only force that can be used to decelerate a car. If a driver increases brake force to one or more wheels that exceeds the maximum amount, the wheels will begin to lock up. A car will particularly become unstable if this occurs at the rear wheels; a driver will no longer be able to control the car.

The ABS wheel speed sensors permanently monitor the speed of each wheel and compare it to a (calculated) reference speed of the car. If the wheel slip calculation indicates a tendency to lock, the ABS HECU will first reduce brake pressure and therefore brake torque to the affected wheel(s) to maintain the lateral guiding forces at the wheel and, thus, the car's overall stability (cf. Kamm circle of frictional forces). Brake torque will be increased again until it matches the friction coefficient. This enables nearly ideal braking while maintaining stability and steerability (see ▶ Chap. 39, "Brake-Based Assistance Functions").

Traction Control System (TCS) If a car's rate of acceleration exceeds the coefficient of friction of the drive wheels, it will trigger traction control function. If the comparison of the drive wheel speed reveals that they are spinning while the non-drive wheels are not, the modulator will firstly reduce engine torque. If one of the drive wheels continues to spin as a result of different coefficients of friction between the left and right side of the road and transmission of drive torque via the differential, brake torque is applied to the affected drive wheel. This occurs completely automatically without any intervention on the part of the driver. The car will remain stable and will accelerate according to the available friction level (see ▶ Chap. 39, “Brake-Based Assistance Functions”).

Electronic Stability Control (ESC) ESC increases brake torque to each wheel individually and can thus generate yaw moment of the car around the car’s vertical axis. The steering angle sensor continually determines in which direction the driver wishes to go. If the car deviates from this course, monitored by yaw rate and longitudinal and lateral acceleration sensors, targeted brake torque to one or more wheels will generate a yaw torque around the car’s vertical axis and thus stabilize the passenger car or make it follow the intended course (see ▶ Chap. 39, “Brake-Based Assistance Functions”). The engine interface is utilized to adapt drive torque output.

1.3 Wheel Brakes

Wheel brakes generate friction via the brake pads which translate brake forces to the wheel. Most common types are disk brakes and drum brakes. Nearly all front brakes today are disk brakes, while many cars also feature rear disk brakes.

1.3.1 Disk Brakes

Disk brakes are axial brakes. The clamping force of the calipers is transmitted via hydraulic cylinders in an axial direction toward the brake pads which clamp the brake disk (also known as a rotor) from both sides. The pistons and pads are housed in a caliper that encompasses the outer circumference of the disk. The brake pads are supported in the direction in which the disk rotates in a caliper bracket mounted to the axle stub.

The brake pad surfaces cover only part of the ring-shaped portion of the disk (the so-called partial disk brake). Generally speaking, when people talk of disk brakes, they mean caliper disk brakes. Full disk brakes, in which the pad covers the entire disk, are not commonly used in cars. There are three subcategories: fixed caliper, fist-type caliper, and combined type caliper. Fixed caliper brakes have pistons on both sides of the brake disk (see Fig. 7); fist-type caliper and floating frame-type caliper have pistons on only one side. In the case of the latter type of brake, the housing slides to and from within the caliper frame (see Fig. 8).

Fixed Caliper Fixed-type calipers are usually the equipment of high-powered and sports cars. The reasons for this are that they exhibit good stiffness, are very responsive, and normally have low residual torque. The look and the image of fixed calipers also play a part. Aluminum alloys are usually the material of choice.

FN Floating Caliper A special fist caliper design (FN) permits a relatively large disk diameter. The advantage is a larger radius and thus greater brake torque at same brake pressure. The housing bridge can be quite long and therefore thin at the

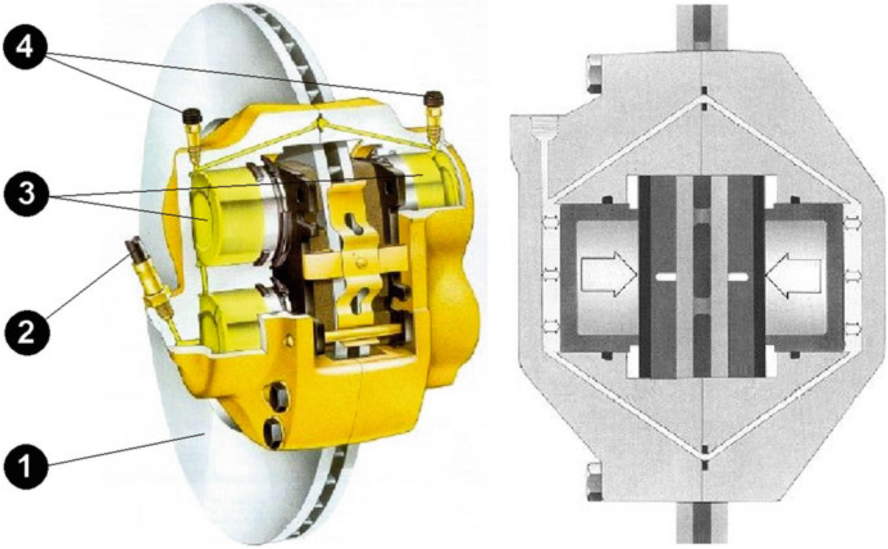


Fig. 7 Fixed caliper: *1* brake disk, *2* hydraulic connection, *3* brake piston, *4* bleeding screws

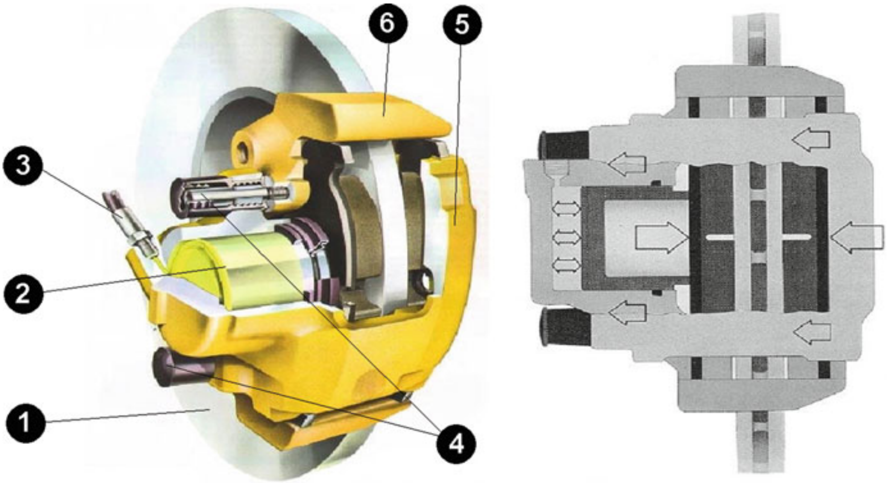


Fig. 8 Continental FN fist caliper: *1* brake disk, *2* brake pistons, *3* hydraulic coupling, *4* bushings, *5* caliper frame, *6* frame

narrowest part of the wheel without any deterioration of caliper stiffness (due to increase in hydraulic volume).

Combined Caliper The FNc combined caliper (Fig. 9) unites the service and parking brake functions into one disk brake caliper utilizing the same friction

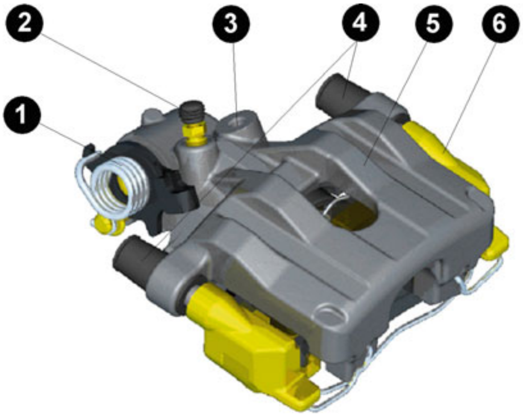


Fig. 9 FNc combined caliper: *1* leg spring, *2* bleeding screw, *3* hydraulic connection, *4* bushings, *5* housing, *6* anchor

pairing for both tasks. The service brake works in analogous fashion to the floating frame caliper. The parking brake function is actuated by hand brake via a cable. A lever within the caliper mechanism turns an actuating shaft, thereby generating parking brake force mechanically by pressing the brake pads onto the disks.

1.3.2 Drum Brakes

Drum brakes are radial brakes, a combination between a brake lining attached to a steering knuckle and a brake drum attached to a wheel. They have two brake shoes with linings that hydraulic cylinders press outwards against the drums. Once braking is completed, springs retract the shoes to their original position.

The geometry of drum brakes makes them robust. They are very efficient and are very common as a low-cost combined service and parking brake at the rear wheels. Even on cars with disk brakes on the rear wheels and heavier weight (>2.5 t), a dual-servo drum brake is integrated into the disk brake pot.

2 Enhanced Architectures

In order to continue reducing fuel consumption and emissions, combinations of internal combustion engines and electric motors known as hybrids have become increasingly common, as have purely electric cars. Depending on the situation, hybrid cars drive either in a purely conventional manner, purely electric, or a simultaneous combination of both types of propulsion. By coupling (temporarily) the drive wheels to an electric motor, it is possible to operate this engine drive both as a motor and as a generator (Fig. 10 explains the principle).

An electric motor can help decelerate a car when it is operating in generator mode. The amount of brake power available from the generator depends heavily on the installed generator output, the speed of the passenger car (rpm torque curve of the generator), and the amount of storable electrical energy. Installed generator capacity today is sufficient for light braking at best: for high-deceleration braking,

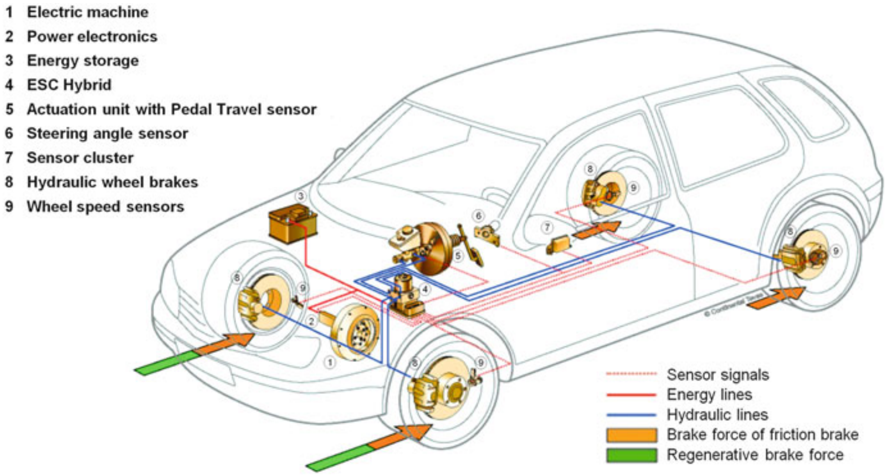


Fig. 10 Hybrid car with a recuperating brake system

however, it is insufficient. It is for this reason “generator brakes” can currently only support the conventional friction brake system.

For hybrid cars, therefore, it is necessary to switch dynamically between friction brakes and generator brakes or overlap, depending on the driving situation and the condition of the system. This must be possible with as little difference in normal brake pedal feel to the driver as possible. The process, known as brake blending, is controlled electronically. In order to achieve pure recuperative (i.e., without friction) braking, it is necessary to decouple the brake pedal from the hydraulic system. In the following is a description of the design of such a system (Fig. 11).

2.1 Recuperative Brake System (RBS-SBA)

The Recuperative Brake System with SBA (Simulator Brake Actuation) (Fig. 11) described here contains essentially the following components in addition to the standard architecture described above together with the electric drivetrain:

- Pedal feel simulator (PSU) with pedal angle sensor
- Active brake booster
- Vacuum pump to supply energy to the brake booster should the internal combustion engine be deactivated
- ECU and software for recuperation, blending, vacuum pump control, monitoring, etc.
- Additional sensors, especially for monitoring status

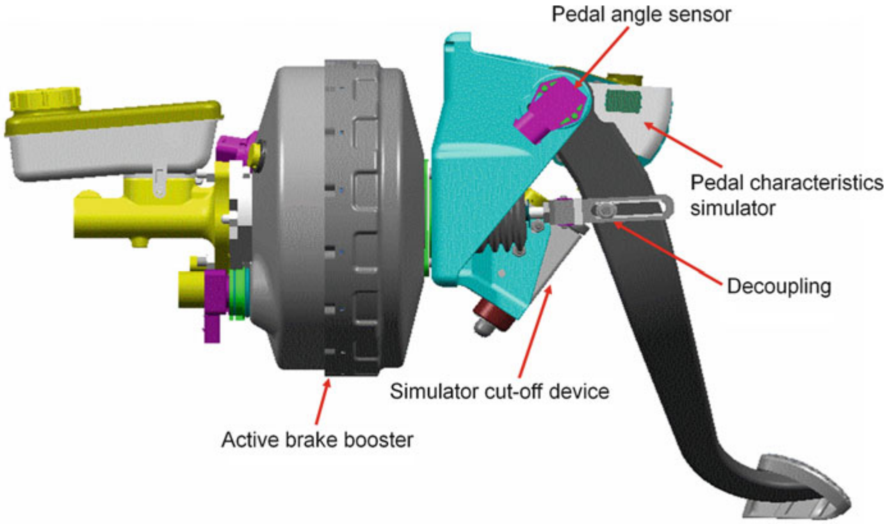


Fig. 11 Actuation unit (human-machine interface, HMI) for a recuperative brake system that permits decoupling of the brake pedal from the hydraulic system as required

When a driver depresses the brake pedal, the angle sensor captures the amount of brake force desired.

The mechanical decoupling of the brake pedal from the brake booster/(tandem) master cylinder does not initially result in a buildup of hydraulic brake pressure. The driver's foot meets with resistance caused by the simulator (primarily influenced by pedal travel) so that the driver does not get the false impression that the brakes are not responding. Brake pedal travel is linked via governing software to a deceleration curve that allows for the simulation of different pedal characteristics – within limits.

The HECU electronic components receive information, via a corresponding generator control interface, concerning what brake torque the generator is able to supply at the moment. If this torque is sufficient to supply the amount of braking the driver desires, generator brake force alone will decelerate the car. If more torque is desired than the generator is currently capable of delivering, the system will generate additional torque through the friction brake by triggering the active brake booster electrically. This increase in brake pressure is independent of the brake pedal actuation and the resultant friction brake torque and supplements the available generator brake torque.

Since some internal combustion engines do not generate sufficient vacuum and since no vacuum at all is generated when a hybrid car drives entirely on an electric motor while the internal combustion engine is shut off, an electrically driven pump supplies the necessary vacuum.

Should the system fail in whole or in part, it will still be possible to brake using conventional mechanical/hydraulic means.

2.2 Electrohydraulic Brake (EHB)

Figure 12 shows the chain of effects in a passenger car EHB system. The layout of the EHB with its three main subassemblies is as follows:

- Actuation unit with tandem master cylinder, integrated pedal travel simulator, backup travel sensor, and brake-fluid reservoir.
- Hydraulic control unit (HCU) with motor, three-piston pump, and metal diaphragm reservoir and furthermore a travel sensor to supply pressure, a valve manifold with analogue control valves, two isolation valves, and two balancing valves forming a unit. Additionally, there are six pressure sensors (See Fig. 13).
- An integrated electronic control unit (ECU).

When the brake pedal is depressed, the isolation valves close immediately. This separates the accumulated pressure in the master cylinder from the brake calipers. The brake simulator comes into play. Pedal travel plus the pressure accumulated in the simulator dictates the amount of deceleration desired.

A conventional power-assisted brake system amplifies foot force via a vacuum brake booster in the actuator unit and converts it into hydraulic pressure. With an EHB, sensors in the activating unit evaluate how much braking is desired and transmits the signal by wire to the ECU (electronic control unit) (Jonner et al. 1996). The conversion to hydraulic pressure takes place via valve action in the HCU (hydraulic control unit). As in conventional brake systems, brake lines and hoses then transmit the hydraulic pressure to the wheel brakes.

The EHB presents a series of advantages to drivers and passenger car manufacturers. Since the brake pedal is decoupled from the wheel brakes, drivers always

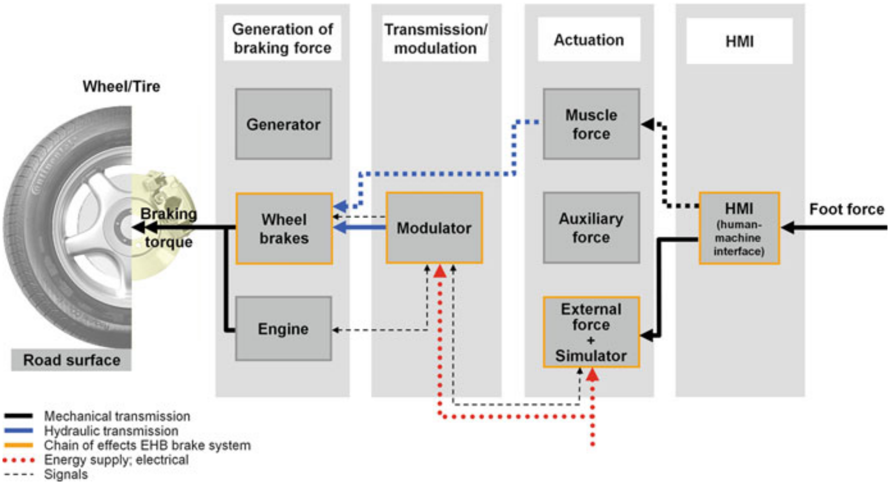


Fig. 12 Chain of effects, EHB brake system in a passenger car

Fig. 13 EHB components by Continental (*left*, electrohydraulic control unit; *right,* brake actuator with pedal characteristic simulator)

sense optimum pedal characteristics as determined by the passenger car manufacturer, with short pedal travel and little effort required. This makes it easier to translate a driver's intention more quickly and to generate brake force more efficiently. Control functions such as ABS occur without feedback, i.e., without the usual pedal vibration. Since instead of a vacuum brake booster there is a high-pressure accumulator, brief lag time occurs while pressure is generated. This permits optimum control functionality for EBD, ABS, TCS, and BA in conjunction with more responsive wheel pressure control.

The extensive network of sensors also allows for precise system diagnostics and a much more comprehensive failure reaction than a conventional brake system. Because the driver is disconnected from the wheel brakes, he cannot feel the system fault. This means that the EHB must also assume responsibility for diagnosing defects. Figure 13 shows the EHB system with a tandem master cylinder, integrated pedal travel simulator, brake-fluid reservoir, and pedal travel sensor.

In order that a driver does not experience unusually "increased" pedal travel, the actuation shuts off the simulator, thus preventing additional brake fluid from flooding into the simulator (see also Fig. 15).

Eliminating the vacuum brake booster has the advantage for passenger car manufacturers that the actuator unit is smaller, which makes it easier to integrate and makes easier to adapt from left-hand drive cars to right-hand drive cars. The shorter actuation unit, moreover, carries advantages in designing the engine compartment around the driver's footwell. The chance of injury from the pedal unit intruding into the passenger compartment during a frontal collision is thus lower.

The pedal travel sensor also serves to monitor the fluid volume of the brake system. Whereas a driver with a conventional brake system can notice increased

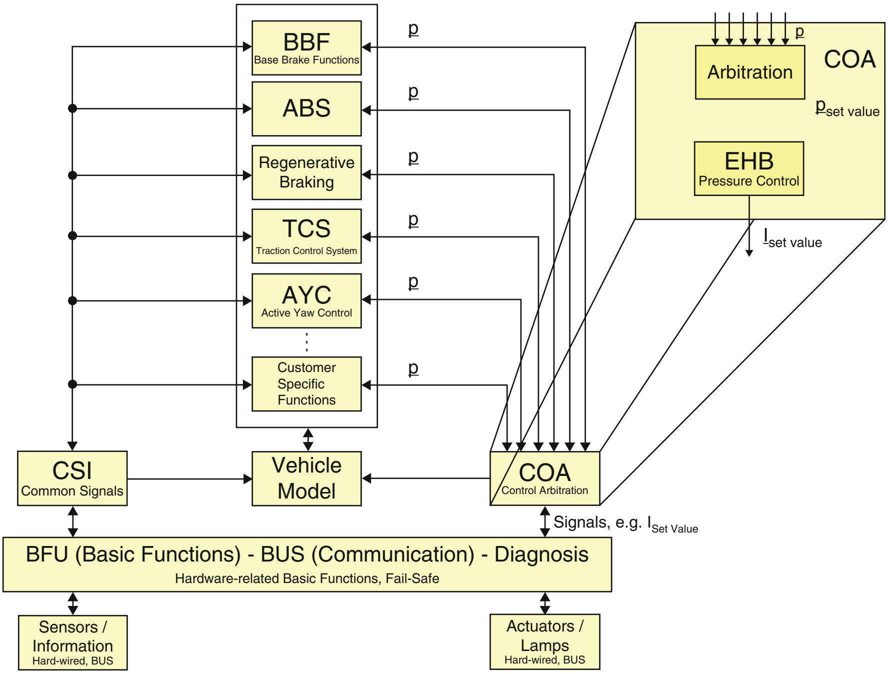


Fig. 14 Modular EHB software structure (Stölzl et al. 2000)

volume (e.g., if air enters the system) because pedal travel has increased, this occurs with EHB by comparing the volume taken from the accumulator with the brake pressure accumulated in the wheel brakes using the system's volume uptake curve (Albrichsfeld et al. 1998). This serves primarily to monitor the total functionality of the hydraulic fallback level because, in by-wire mode, the brake-fluid volume contained in the accumulator can compensate nearly imperceptibly for increased volume uptake by the system.

Control and Monitoring Methods Since the EHB can be considered essentially a universal actuator for braking which makes use of extensive sensors, its software, the control, and the monitoring methods are of cardinal importance. The only way to achieve broader functionality is to develop new software modules.

The modular software structure (Fig. 14) represents a logical extension of this philosophy that Continental has been following (Fennel et al. 1998; Rieth 1999).

Each of the functions displayed prescribes a projected pressure. The arbitration module (COA) calculates a projected value for the brake circuits using intelligent weighting and by setting priorities. The use of modular structure made it possible to adopt higher control functions from previous ABS and ESC projects with relative ease.

The process of discerning a driver's intention serves to calculate wheel-brake pressure at each wheel from the values that sensors have obtained from the

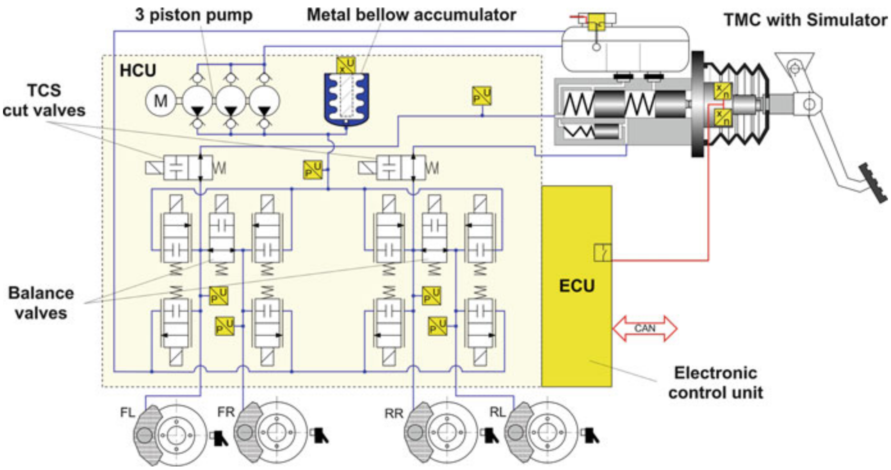


Fig. 15 EHB circuit diagram with schematics of the system components

redundant pedal travel sensor and the pressure in the tandem master cylinder. After these signals have undergone a plausibility check, the system determines the amount of deceleration desired with the aid of derivative factors. This allows for a continual brake boosting function distributed to each individual wheel that is apportioned according to the situation and is then converted into different brake pressure at each wheel. A downstream wheel pressure target control sets the projected brake pressure at each wheel.

Function Description of Electrohydraulic Brakes (EHB) The electrohydraulic brake is a full-power brake system. The main characteristics are compact size, optimum reaction times, and brake pedal characteristics which are scalable. The EHB is decoupled from the brake pedal and is thus feedback-free, both under normal braking conditions and during wheel slip control. An example of how the subassemblies are arranged is given in Fig. 15.

A high-pressure accumulator generates brake energy with the aid of the hydraulic control unit in accordance with the driver's input. An integrated motor pump unit pre-pressurizes the accumulator.

During braking, the hydraulic connection between the tandem master cylinder and the hydraulic control unit is disconnected. The pre-pressurized accumulator unit adjusts the brake pressure at the wheel via control valves.

Summary of the advantages in comparison to a conventional brake system:

- Shorter braking and stopping distances (shorter threshold time, accumulator system)
- Optimum braking and stability because of fast reaction time
- Optimum pedal feel, easy to adapt to customer specifications

- Noise-free operation without distracting brake pedal feedback during standard braking
- Better crash properties due to decreased intrusion
- Easier mounting due to elimination of vacuum brake booster on the firewall
- Use of uniform subassemblies
- Easy to realize full-power braking energy for various auxiliary functions such as ACC, disk drying function during rainstorms, anti-fade, etc.
- Independent of vacuum and, thus, perfect for new vacuum-optimized internal combustion engines
- Can be networked easily with future traffic guidance systems

The design of the control unit permits integration of all of today's brake and slip control functions such as EBD, ABS, TCS, ESC, BA, ACC, etc., without any additional hardware (Albrichsfeld and Eckert 2003).

The system determines the intended brake maneuver with triple redundancy for safety reasons (redundant travel sensor and pressure sensor) so as to be able to run a plausibility check on the signals and so as to always have a redundant estimation of the driver's intention.

The ECU processes the signals along with additional signals that describe the car's status as well as full-power brake interventions. It then converts this information into optimized, stable brake pressure at each individual wheel. A pressure control circuit with an analogue intake and outlet valve for each individual wheel, located in the HCU, adjusts the brake pressure at each wheel. A unit, consisting of motor, pump, and accumulator, supplies the necessary high pressure. Depending on temperature, it realizes operating pressures of approx. 150–180 bar.

In many cases, braking does not have to be apportioned to each individual wheel. The balancing valves (the connection between the wheel brakes of each axle) stay open in order to attain equal pressure at the wheel brakes of an axle. This allows diagnostic functions (such as comparison of the pressure sensor values). In similar fashion, pressure can build up solely by triggering a pair of solenoid valves at an axle (by utilizing the open balance valve during gentle braking as pressure builds up slowly). This can lengthen the service life of the solenoid valves since the system drivers and thus the ECU are subject to less thermal stress.

The use of multiple sensors allows for very detailed self-analysis of the EHB system. It also makes it possible to implement various fallback levels should one of the components fail. The hydraulic fallback level assumes particular importance here, should the electrical supply fail, for example. All solenoid valves will be in the position indicated in Fig. 15. The increase in pressure to the wheel brakes occurs in similar fashion to a conventional brake system when there is no vacuum, namely, through the force that a driver's foot exerts via the open cutoff and balance valves. In order to spare drivers unnecessarily long pedal travel, the simulator is shut off hydraulically, thus preventing additional increases in volume.

Two fallback levels are available in case of malfunction:

First level: If the high-pressure accumulator fails, the brake-by-wire function will remain available. However, only the pump will supply the necessary brake fluid to the brakes.

Second level: Should the brake-by-wire function fail (e.g., if the electrical system fails), the hydraulic links of the master cylinder to both brake circuits will remain intact. All four wheel brakes will be activated without any boost, proportional to the foot force applied. The simulator remains shut off in this case (Stölzl et al. 2000).

The dual-circuit feature required by legislation will remain, in spite of a partial failure.

Market Overview and Differences in Design of EHB Systems The first mass market EHB system was the one developed by Japanese automotive supplier ADVICS in early 2001, which was installed in the Toyota Estima (only in Japan). Late 2001 saw the introduction of the EHB system that Robert Bosch GmbH developed under the name Sensotronic Brake Control (SBC) in the Mercedes SL Roadster and, in early 2002, in the high-volume production Mercedes E Class. 2004 saw the introduction of the Continental EHB system in the Ford Escape Hybrid.

ADVICS presented a reworked version (ECB II) in the Toyota Prius II Hybrid in late 2003. The actuation system, with integrated brake pedal travel simulator and motor-pump-accumulator unit, had been combined into one large module. The second generation saw it split into individual modules. The introduction of the Continental EHB system in the Ford Escape Hybrid took place in the United States in late 2005.

ADVICS presented its enhanced ECB III system in early 2009 in the third-generation Toyota Prius, reducing the number of main assemblies from three to two. The simulator was integrated into the actuation unit.

All of the EHB systems on the market (see Fig. 16) function according to the by-wire principle described above. The main differences are in the way in which components are grouped together and also that they exhibit different fallback level concepts.

Potential Functions Up until now, the functions of electrical brake systems have dealt mainly with driving dynamics that assist a driver in critical situations such as ABS, TCS, ESC, and BA. As such, they intervene relatively seldom in situations in which a driver has to concentrate entirely on driving. By contrast, today's electronic brake systems increasingly contain assistance and comfort functions that a driver can experience on a daily basis. They can assist drivers in situations in which a car's occupants can well notice the effects. For example, the traffic jam assistant, developed from the ACC function, requires finely applied brake pressure in the

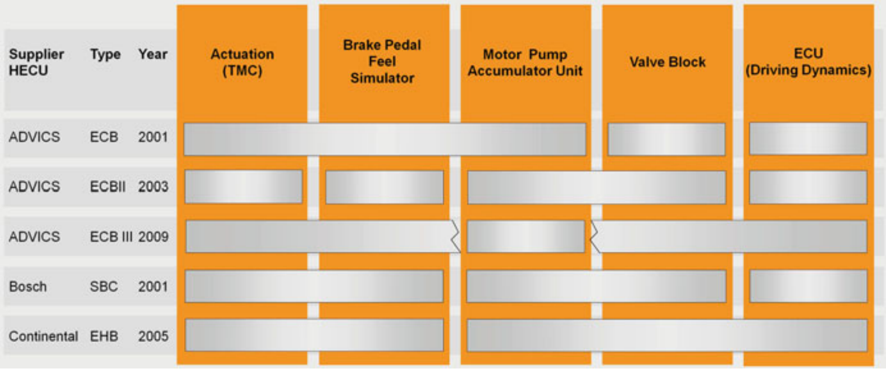


Fig. 16 Overview component integration comparison for various EHB systems

acoustically ultrasensitive low-speed range so that the car accelerates and brakes as a function of the distance to the car in front of it.

The development of economical, optimized internal combustion engines has led to a situation in which many cars have to rely on mechanical or electrical vacuum pumps because the combustion engine no longer generates enough vacuum for the brake booster. An EHB eliminates the need for vacuum so that the vacuum energy is no longer necessary.

Decoupling the brake pedal and triggering braking by electronic signals also permits optimum communication and exchange with other chassis control systems, making the EHB the ideal actuator for future Global Chassis Control (GCC) systems (Rieth et al. 2001). It likewise makes it an ideal brake system for recuperative braking, especially in hybrid cars. In this case, the generator produces as much brake torque as possible, while the friction brake needs only to make up the difference in brake force desired by the driver.

The increasing introduction of assistance functions and the comparatively high costs of EHB systems lead to heightened demands on function and comfort for future full-power brake systems as well.

2.3 Integral Brake Systems

MK C1 Electrohydraulic Brake Actuator (Continental) The electrohydraulic brake actuator currently under development (Fig. 17) is capable of building up brake pressure much faster than conventional hydraulic systems, plus it does not depend on engine vacuum (Feigel 2012). It also meets the heightened requirements of new driver assistance systems to avoid accidents and to protect other vulnerable road users such as pedestrians. What is more, the system can meet the requirements of a recuperative brake system with a high degree of comfort. A compact, lightweight brake module houses the actuation and boosting functions together with the control system (ABS, ESC, ACC, etc.).

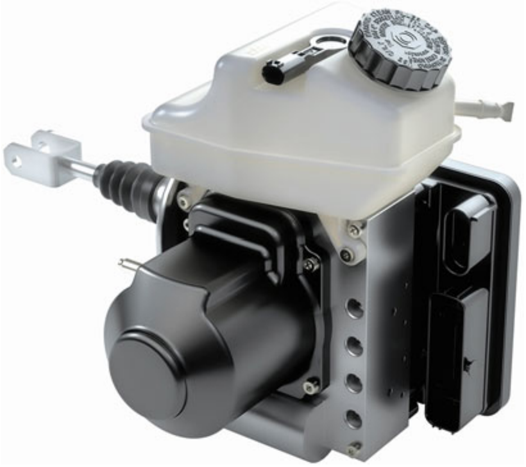


Fig. 17 MK C1 compact brake system

Fig. 18 Integrated brake control IBC

Multiplex Brake System: IBC (TRW) An alternative architecture for brake systems, still in the development stage, is based on the multiplex principle that supplies the wheel brakes with pressure or brake fluid (Vollmer 2013). In contrast to conventional systems, a central motor-master-cylinder unit supplies the multiplex system (see Fig. 18) with brake pressure. The driver actually depresses a pedal simulator when braking. Dedicated sensors capture the driver's intention and convert it into pressure by means of the motor-master-cylinder unit. Only one shift valve is available for each wheel brake; it is used for both building up pressure and releasing it. Pressure control in the sense of multiplexing occurs nearly sequentially: the master cylinder generates the pressure needed at each wheel and then transmits it to the respective wheel brake(s) via shift valves. In reality, it is not an exclusively sequential process because pressure control processes during wheel slip adjustments overlap sequentially and in parallel.

Since the above mentioned brake pedal travel simulator disconnects the driver input from the wheel-brake hydraulics, it is possible to combine the brake torque from the wheel brakes and generator dynamically without a driver noticing these adjustments through the brake pedal (brake blending).

3 Dynamics of Hydraulic Brake Systems

The issue dealt with here is the response time between when a braking maneuver commences and the rate at which a car actually decelerates. The main determining factors here are the components of the installed brake system, the dynamic axle loads, the chassis with its suspension and shock absorbers, and the tires. Moreover, peripheral influences such as the ambient temperature and the temperature of the components may play a major role.

This dynamic aspect of brake systems is of major importance for various driver assistance systems such as emergency brake assist. The performance of such an assistance system is assessed in the specifications and test procedures (see Euro NCAP procedures as an example, ▶ Chap. 39, "Brake-Based Assistance Functions"). One test is the stopping distance after a car has been confronted with a defined obstacle in its trajectory. Brake pressure buildup should begin automatically, mostly via the hydraulic pump of the HECU (hydraulic/electronic control unit) without any driver input, according to usual testing procedures. Time (and thus brake distance) claims a relatively large share of what happens between the point that an object is detected and the point that maximum deceleration is achieved, in comparison to the steady-state condition. Figure 19 demonstrates behavior with a "standard brake system" for mid-class passenger cars. In the case of emergency braking at high speeds on the other hand, most of the time is spent on

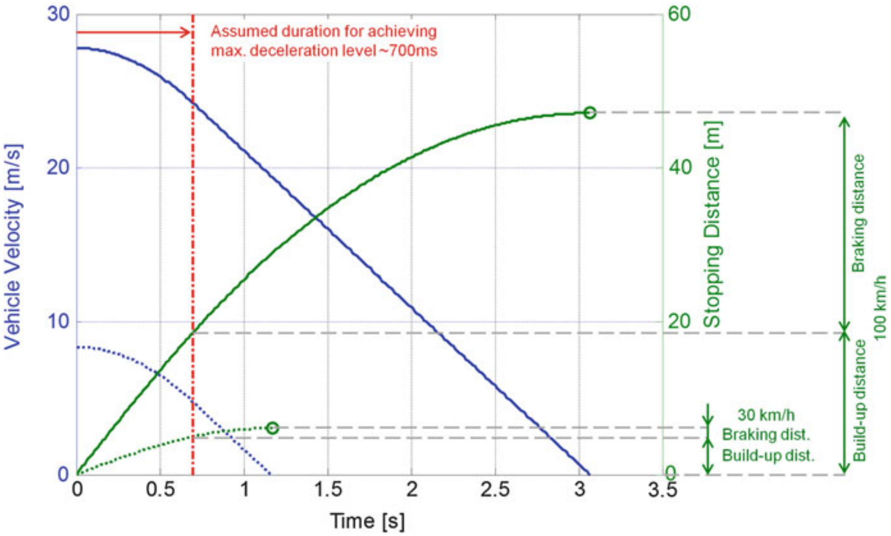


Fig. 19 Simulated autonomous emergency braking from 30 to 100 kph with identical threshold time

the actual emergency braking phase. The time at the very beginning of braking is particularly valuable in terms of reducing stopping distance, however, because the car is still traveling quite fast. Therefore one goal of system design is to reduce this threshold time – even though development today bears on avoiding collisions in urban traffic and thus at comparatively low speeds.

In summary, it is important to consider the entire system when developing such driver assistance systems, i.e., beginning with sensor-based capture of the surrounding, through signal assessment (analysis of the object and decisions on action to be taken) through triggering and pressure buildup of the HECU, to the response and behavior of the wheel brakes, tires, and chassis. It is important to consider all these elements, particularly when predicting how a system will react during simulation. Each of these subsystems inevitably causes a certain delay in the entire chain, be it through software runtime, transmission of signals/commands, mechanical or hydraulic inertia, elasticity, or damping effects. On the whole, however, these individual factors (which appear to be insignificant at the first glance) can largely be ignored. What makes things more difficult when contemplating the overall system is the unavoidable scattering of parameters which results from production effects or due to different peripheral factors influencing operation. The following passages will focus on the hydraulic brake system. Braking within this subsystem usually commences with a signal to the ESC-HECU such as a deceleration request. The HECU receives such a request either directly from the sensor's control device or indirectly through a higher-order control device, whereby communication flows via networks such as CAN or FlexRay. The HECU's control device checks to what extent the deceleration request is feasible in terms of dynamics and stability. Additional information such as wheel slip, transverse acceleration, yaw, etc., flows into what is known as the arbitration process. The HECU has permanent access to this information anyway for other control functions. The process described above can take up to many 10 ms. If the result of arbitration is positive, the power electronics built into the HECU control device activate the motor-driven hydraulic pump and create pressure for the wheel brakes. The amount of pressure needed for wheel lock can be reached in approx. 150 ms or, in the worst case, in nearly 1 s, depending on the size of the system and the ambient conditions. Figure 20 illustrates the important factors that influence this behavior.

The main reason for the widespread scattering that occurs when brake pressure increases is dependent on the different HECU classes available on the market today. Some customized solutions are available for a wide variety of automotive segments and price classes. From a historical point of view, HECUs used to have to either supply individual wheels with a hydraulic flow (such as during ESC control), or they were pre-pressurized from the tandem master cylinder (as in the case of hydraulic brake assist systems) when a driver steps fast on the brake pedal. In either case, lower pump output was sufficient to meet general market demands. Automated emergency braking on the other hand places greater demands on HECU performance at all wheels. The only way to meet this demand is to provide more powerful HECU units, which can be limited by the car's electrical system load capacity.

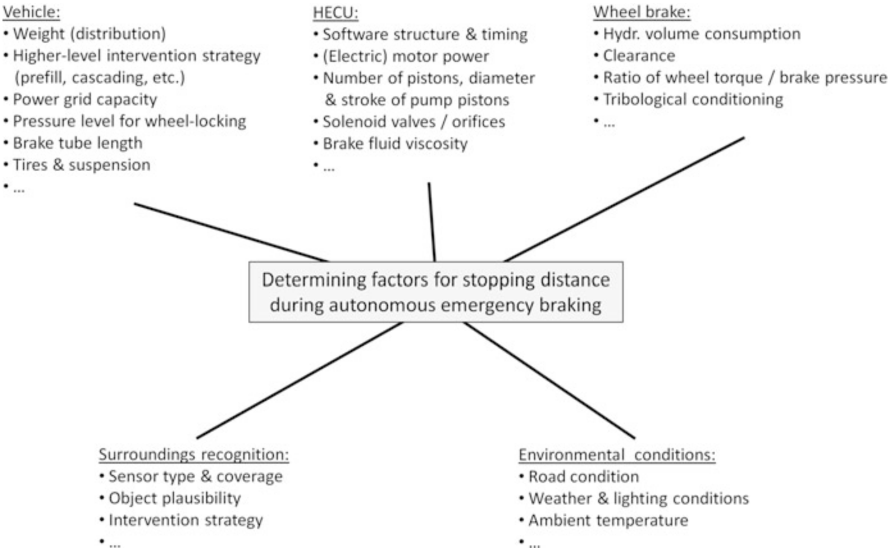


Fig. 20 Limiting influences on triggered emergency braking

The point of reference when it comes to brake pressure buildup time is currently (still) the human being. Trained drivers can reach wheel lock pressures typical for passenger cars within 150–200 ms by actuating the brake pedal (approx. 80–100 bar under normal conditions). The time for recognizing and reacting is not included. Only the fastest HECUs with fully integrated actuation unit (see Sect. 2.3, MK C1) are capable of comparable performance. Standard HECUs may be slower to build up pressure, as may be seen in Fig. 21. It is, however, possible to attain an acceptable level of performance by carefully optimizing the entire system with a targeted triggering strategy (such as stiff brake calipers, low brake-locking pressure, pre-pressurizing the brake system while the system runs an object plausibility check, etc.). A perhaps oversized hydraulic system does not present a solution if its high ability for pressure buildup dynamics is lost in the elasticity of tires and chassis.

Finally, Fig. 22 contains an example (mid-class sedan with sports chassis) for a cascading brake-trigger strategy. Following an initial pre-fill phase, brake pressure increases in accordance with an initial deceleration rate of approx. 4 m/s^2. The delay between wheel-brake pressure and deceleration is apparent in the left half of the illustration. This is chiefly due to the reaction time needed for tires and chassis. The wheel-brake pressure increases once again in accordance with a deceleration of approx. 8 m/s^2. The delay between pressure and deceleration in the right half of the illustration is less because the entire chassis is now pre-conditioned and compressed. Additional brake pressure now translates nearly immediately into higher deceleration.

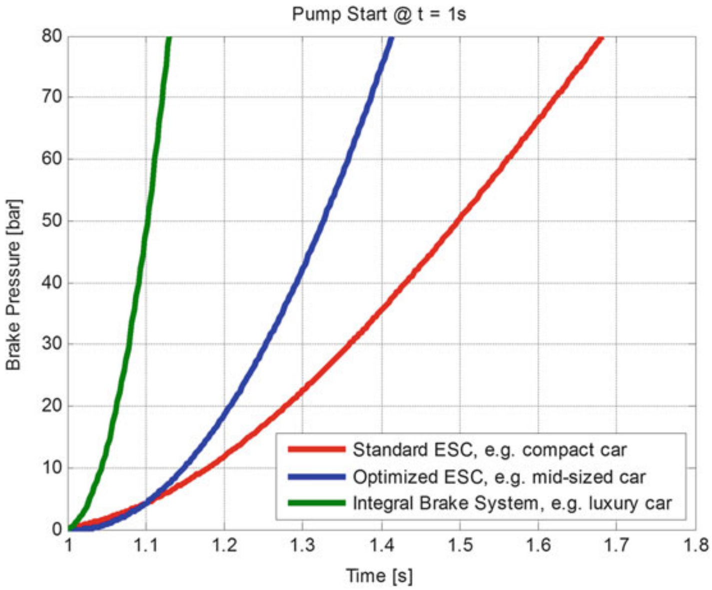


Fig. 21 Comparison of pressure buildup dynamics of various HECUs under conditions typical for different classes of cars

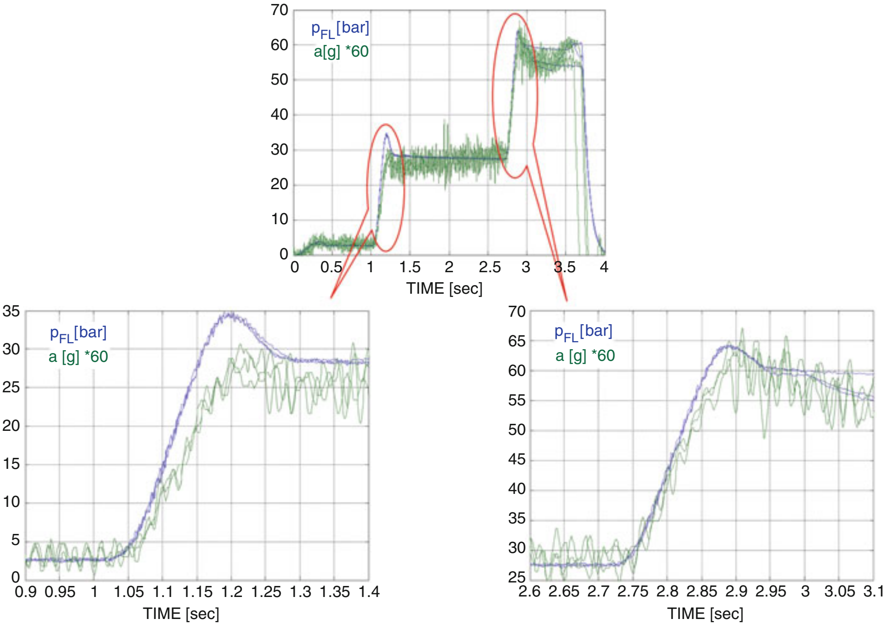


Fig. 22 Example of gradual automatic increase in brake pressure and deceleration buildup

References

Albrichsfeld C von, Eckert A (2003) EHB als technologischer Motor für die Weiterentwicklung der hydraulischen Bremse (EHB as technological driver for the further development of hydraulic brakes), HdT-Tagung Fahrzeug- und Verkehrstechnik

Bosch Mediaservice Kraftfahrzeugtechnik (2013) Presseinformation. http://www.bosch-presse.de/presseforum/details.htm?txtID=6278

Breuer B, Bill KH (2013) Bremsenhandbuch (brake technology handbook). Vieweg Verlag, Wiesbaden

Feigel HJ (2012) Integrated brake system without compromises in functionality. ATZ 114:46–50

Fennel H, Gutwein R, Kohl A, Latarnik M, Roll G (1998) Das modulare Regler- und Regelkonzept beim ESP von ITT Automotive (The modular controller and control concept of the ITT Automotive ESP System), vol 7. Aachener Kolloquium Fahrzeug- und Motorentechnik, Aachen

Jonner WD, Winner H, Dreilich L, Schunck E (1996) Electrohydraulic brake system – the first approach to brake-by-wire technology. SAE 1996-09-91, Detroit

Rieth P (1999) Technologie im Wandel X-by-Wire (Technology in change X-by-wire). IIR Konferenz Neue Elektronikkonzepte in der Automobilindustrie, Stuttgart

Rieth P, Eckert A, Drumm S (2001) Global Chassis Control – Das Chassis im Reglerverbund (Global chassis control – the chassis in the controller network). HdT-Tagung Fahrwerktechnik, Osnabrück

Stölzl S, Schmidt R, Kling W, Sticher T, Fachinger G, Klein A, Giers B, Fennel H (2000) Das Elektro-Hydraulische Bremssystem von Continental Teves – eine neue Herausforderung für die System- und Methodenentwicklung in der Serie (The electro-hydraulic brake system from Continental Teves – a new challenge for the system and methodology in production). VDI-Tagung Elektronik im Kraftfahrzeug, Baden-Baden

Vollmer A (2013) Powertrain, Bremsen und Sensoren, Eine kleine Vorschau auf Highlights der IAA (Powertrain, brakes and sensors, a brief preview of the highlights at the IAA). AUTOMOBIL ELEKTRONIK, pp 40–43

von Albrichsfeld C, Bayer R, Fritz S, Jungbecker J, Klein A, Mutschler R, Neumann U, Rüffer M, Schmittner B (1998) Elektronisch regelbares Bremsbetätigungssystem (Electronically controlled brake actuation system), Patentschrift DE 198 05 244.8. Deutsches Patent- und Markenamt

Electro-Mechanical Brake Systems

30

Bernward Bayer, Axel Büse, Paul Linhoff, Bernd Piller,
Peter E. Rieth, Stefan Schmitt, Bernhard Schmittner,
Jürgen Völkel, and Chen Zhang

Contents

Abstract

Electromechanical brake systems are already on the market as EPB (Electric Park Brake), in combination with conventional “wet” hydraulic service brake systems. In the future, so-called hybrid service brake systems will appear with

---

B. Bayer (✉) • A. Büse • P. Linhoff • B. Piller • P.E. Rieth • S. Schmitt • B. Schmittner • J. Völkel •
C. Zhang
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: bayercub@t-online.de; bernward.bayer@continental-corporation.com;
axel.buese@continental-corporation.com; paul.linhoff@continental-corporation.com;
bernd.piller@continental-corporation.com; peter.rieth@continental-corporation.com;
ls.schmitt@t-online.de; bernhard.schmittner@continental-corporation.com;
juergen.voelkel@continental-corporation.com; chen.2.zhang@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_31

the front axle still being hydraulically actuated and the rear axle having new "dry" electromechanical brake systems as a feasible "high-end" solution for advanced vehicles.

1 The EHCB System (Electric-Hydraulic Combined Brake, Hybrid Brake System)

The EHCB system combines hydraulic servo front-wheel brakes with electromechanical power rear-wheel brakes. The parking brake is fully integrated in the rear-wheel actuators (Electric Park Brake, EPB).

1.1 Objectives

An electromechanical brake system (EMB) consumes less electric power in rear-wheel applications than in front-wheel applications, since lower levels of clamping force and dynamic responsiveness are required. This level of power consumption can be met with a conventional 12/14V electrical system. Many of the advantages of a fully brake-by-wire system, such as an integrated parking brake, variable brake force distribution between front and rear wheels, and software-based design, are achievable even with a hybrid – electrohydraulic combined brake (EHCB) – system. Such a system also offers better performance and comfort than conventional systems when it comes to implementing driver-independent braking requests, e.g., from driver assistance systems. On electric vehicles featuring braking energy recuperation using an electric motor/generator on the rear axle, the system can be configured for brake blending at the rear wheels.

This system offers vehicle manufacturers a range of other advantages too. Since only the front brake is hydraulically actuated, it is possible to reduce the size of the actuator (vacuum brake booster) and thereby to significantly optimize the pedal feel characteristics. And since the electromechanical rear-wheel brakes are actuated independently of the hydraulic front-wheel brakes, it is possible to achieve better and more adaptable overall brake response characteristics. At the same time, the "dry" rear-wheel brake not only dispenses with hydraulic brake lines and hoses to the rear axle but also allows the axle to be preassembled using fully tested modules with simple interfaces.

1.2 System Architecture and Components

At the front wheels, the EHCB system (Strutz et al. 2013; Stemmer et al. 2012) comprises a hydraulic single- or dual-circuit actuation system acting on conventional hydraulic brake calipers. The hydraulic system is adapted for single-axle braking. The driver's wishes are identified by means of sensors at the pedal and in

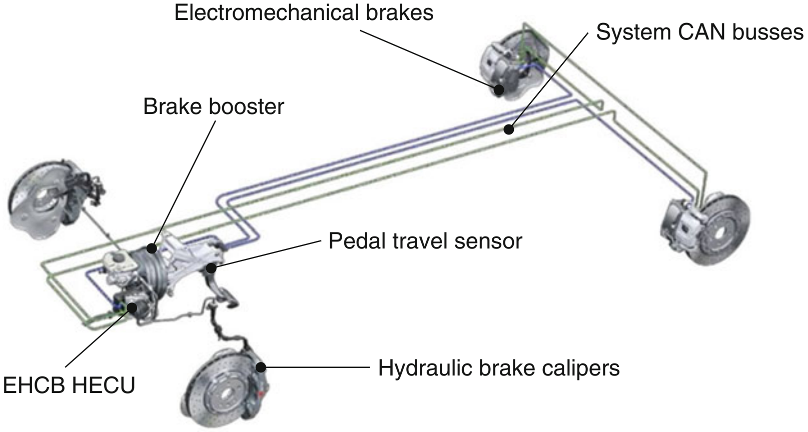


Fig. 1 EHCB system, layout (source: AUDI)

the central HECU (hydraulic electronic control unit). The two rear-wheel electromechanical actuators (brake calipers) complete the brake system (see Fig. 1). The driving dynamics sensors (for wheel speed, steering angle, yaw rate, and acceleration) remain unchanged. An electric park brake actuation button is also provided.

Slip control functions continue to be handled by the electronic control unit (ECU) within the hydraulic stability control HECU. Management of the regular service braking function is also handled by this HECU. Any braking requests from driver assistance systems, for example, adaptive cruise control commands, are likewise processed by this unit. The HECU also identifies driver wishes and, via a bus connection (e.g., CAN), requests the optimal amount of braking force from the rear-wheel brakes, taking into account driving and load conditions. To ensure the necessary high level of system functionality and availability for regular service braking, a variety of redundancies are built into the system, such as a ring structure for finely controllable signal transmission to the rear-wheel actuators.

At this stage, such a system does not require any significant changes in terms of vehicle equipment (see Fig. 2). The basic packaging remains unchanged.

1.3 Control Functions

The existing, familiar control functions are fully retained (Schmittner 2003). However, they are joined by various supplementary and new functions:

The regular service brake system is augmented with adaptive brake force distribution for optimal brake proportioning, taking into account different load and driving conditions. Vehicles with soft, comfort-biased suspension are prone to unpleasant fore-and-aft pitching when braking to a standstill. This effect can be virtually eliminated by reducing the braking force at one axle for short periods (“soft-stop”/“anti-jerk” function). With EHCB, this can be done in such a way that

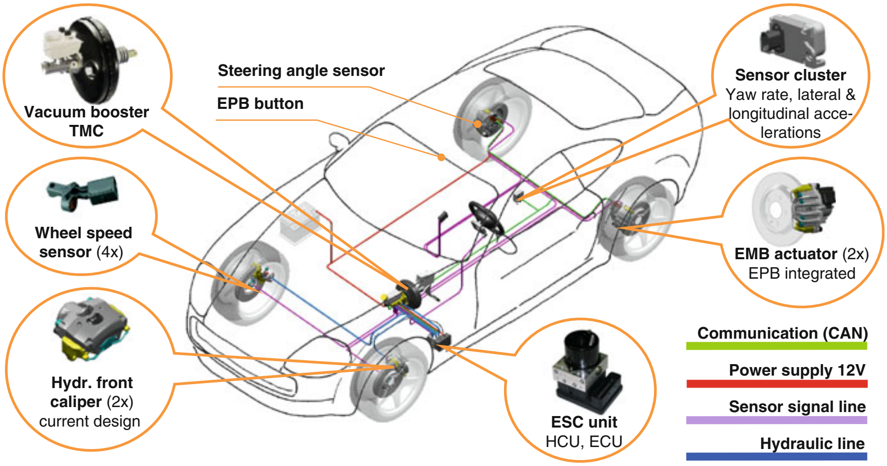


Fig. 2 EHCB system, components

the driver is not aware of the intervention at the brake pedal. Slip control functions meanwhile can potentially be optimized by increasing the brake force at the rear wheels beyond the level requested by the driver, while at the same time the reduction of the hydraulic brake system to the front axle solely makes for a more comfortable pedal feel. An integrated parking brake allows new control concepts that make full use of service brake and parking brake functionality, with highly dynamic, smooth transitions between the two. The resulting standstill management enables the vehicle manufacturer to implement nearly any kind of operating concepts (see Sect. 2.6). The system also caters to the increasing requirement for driver-assistance-system-prompted brake actuation, particularly in the low-duty deceleration spectrum (up to approx. 0.3 g). In this spectrum, optimally controllable braking can be achieved using only the rear-wheel brakes. The driver notices nothing unusual at the brake pedal.

Enormous comfort improvements, particularly in the case of acceleration skid regulation at the rear wheels, can be achieved due to the fact that control interventions can be more finely modulated and are no longer audible.

For rear-wheel-drive electric or partially electric vehicles (all-electric and hybrid vehicles), the EHCB system offers optimal opportunities for recovering braking energy and feeding it back to the battery using blended braking (whereby some friction braking is replaced by regenerative braking).

1.4 Rear-Wheel Brake Actuation System

The rear-wheel brakes in an EHCB system are actuated by all-electromechanical “dry” rear-wheel calipers without brake fluid. The system comprises floating calipers (fist calipers) with electric motor, reduction gearing, and spindle-nut

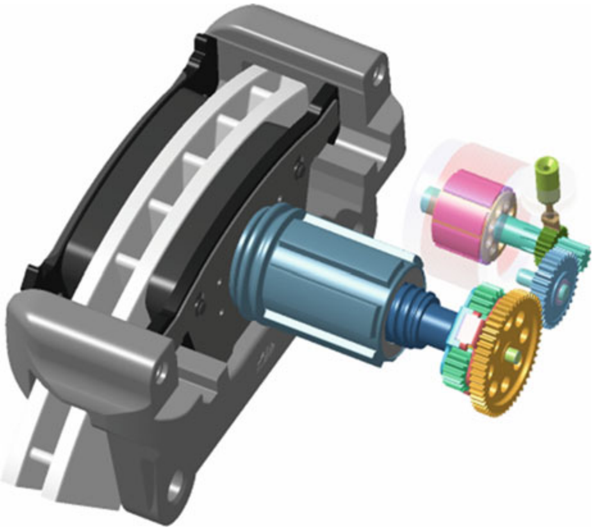


Fig. 3 EHCB system, actuator system

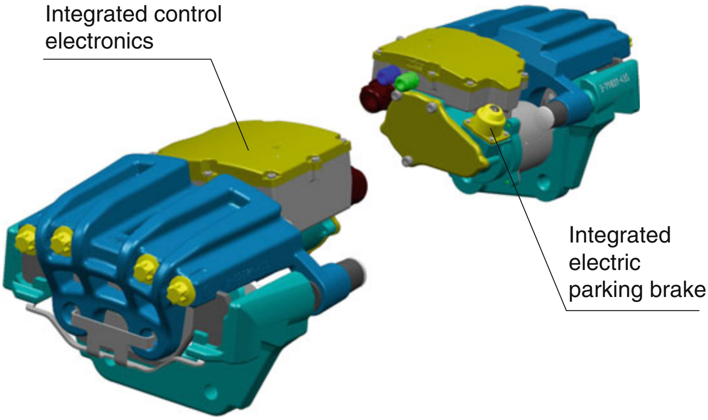


Fig. 4 EHCB system, electromechanical rear caliper

drive (Schmittner 2004) (see Figs. 3 and 4). The focus is both on integration of the parking brake and on a cost-optimized design that still maintains adequate levels of dynamic response. The system is currently designed to operate with a 12/14 V vehicle electrical system.

2 The Electric Park Brake (EPB)

In more and more vehicles, an electromechanical parking brake is being used in place of a driver-actuated mechanical parking brake, in order to meet growing demands in terms of reliability, driver comfort, and in-car connectivity. Numerous different versions of the EPB have been developed, to cater to the wide-ranging demands and diverse vehicle concepts.

2.1 Objectives

The main objectives are to ensure simple, convenient, low-noise, reliable parking brake functionality. Integrated with other vehicle systems and with sensor data, the electric park brake can provide enhanced driver convenience either in the form of assistance in complex driving situations (e.g., moving off on a slope) or in the form of automatic functions (e.g., autoapplication of the parking brake when the driver exits the vehicle). Replacing the hand-operated lever or foot-operated pedal of driver-operated mechanical systems by a push button frees up interior design options and improves passive safety. Additionally, the possibility of actuating the parking brake automatically and independently of the driver allows the stationary vehicle to be firmly secured without driver intervention. This is a feature that has been incorporated into the safety concept of numerous assistance systems.

Take-up for these benefits will be further enhanced by the steady reduction in system costs due to advances in integrating the control unit and actuation system.

2.2 System and Components

Electromechanical parking brake systems comprise control and display systems, a control unit with accompanying software, and an actuator system. On top of the legally required minimum specification, it is also common for the control and display systems to incorporate manufacturer-specific features in the form of additional comfort functions or additional information for the driver about system functions and status in the form of audible signals or text messages.

2.3 System Architecture

In the first EPB systems, the control unit and its software were designed as a standalone system (see Fig. 5).

The first step toward integration was to fit the hardware electronics circuit board in the central actuator housing (cable puller system, see Fig. 6).

Functional commonalities between the EPB and the ESC (electronic stability control) system and the shared use of vehicle interfaces provided an incentive for integrating the EPB and ESC ECUs in the HECU. Such a design first went into production in 2012, since when it has become increasingly common (see Fig. 7). One challenge associated with this system has to do with the length of the electric wiring connecting the control unit and the actuators. In particular, high wiring temperatures and/or high contact resistances tend to limit the current-carrying

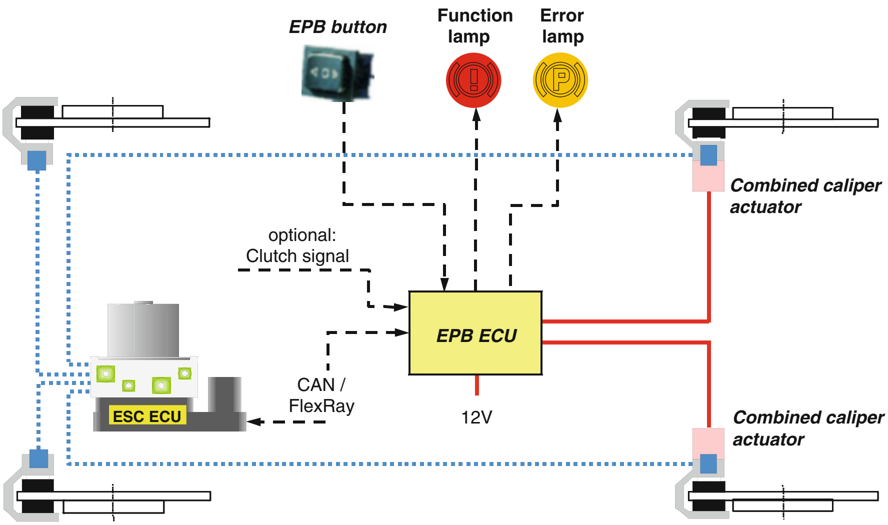


Fig. 5 System layout with stand-alone ECU

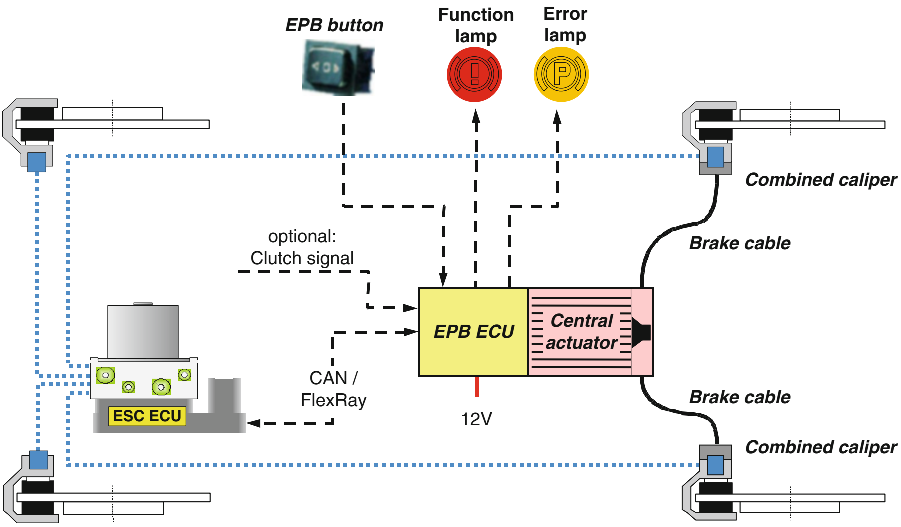


Fig. 6 System layout with central actuator

capacity of the wiring, so it is important to ensure that the actuators offer sufficiently high energy conversion efficiency (electrical energy into clamping force).

Initially, the integrated products comprised ESC and EPB electronic hardware and software from one and the same supplier. One of the requirements for second-generation integration, however, was to design a standardized interface (see Fig. 8) allowing EPB and ESC systems from different suppliers to be combined – a

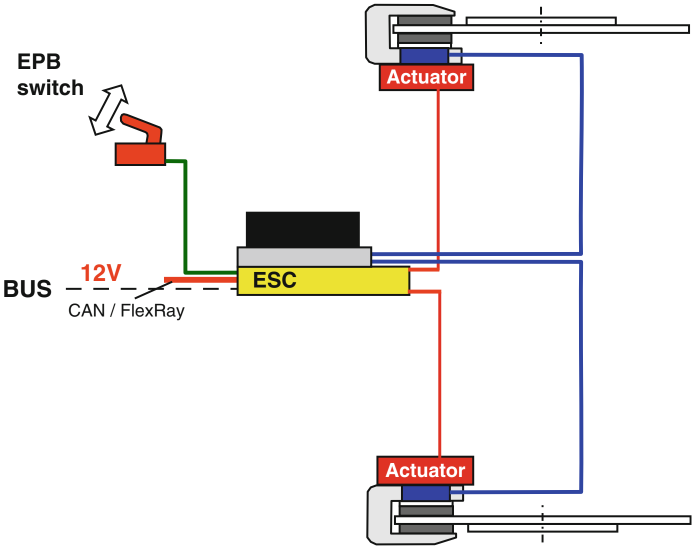


Fig. 7 System layout with ECU integrated in the ESC control unit

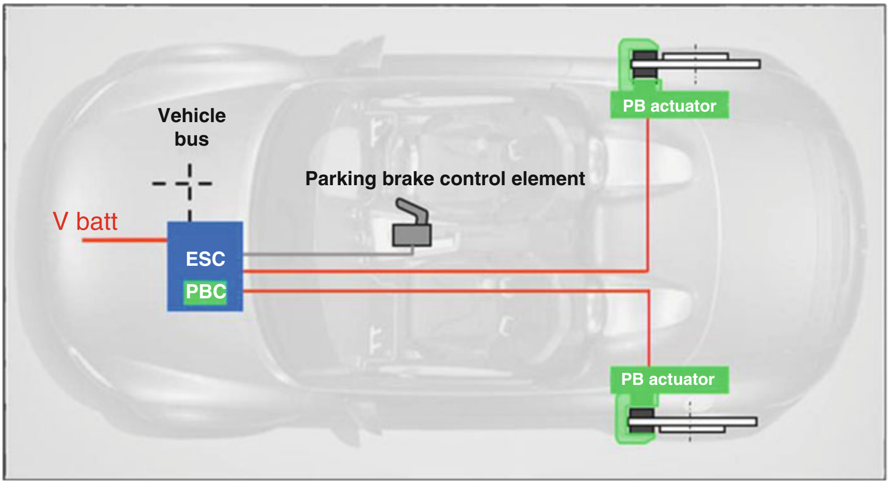


Fig. 8 System layout with open interface

criterion for the acceptance of integrated systems by the vehicle manufacturers. The standard was developed by members of VDA (Association of the German Automotive Industry) and published in VDA recommendation 305–100. This recommendation both describes the technical constraints for the interface and assigns responsibilities for development and implementation.

2.4 Actuator System

An electromechanical parking brake actuator primarily comprises an electric motor, gearing, and a transmission mechanism. Some versions may also include additional sensors.

2.4.1 Central Actuator

When first introduced on the market, electromechanical parking brakes normally featured a central actuator. A central actuator, acting on one or two cables, allows the previous driver-actuated system to be replaced without the need for changes to the foundation brake. This reduces the design effort, particularly when the system is to be fitted in existing vehicle architectures. The central actuator primarily comprises an electric motor, gearing, a spindle-nut transmission mechanism connected to a cable or cables, and a force or displacement sensor. The ECU is integrated in the central actuator housing (see Figs. 6 and 9). The advantage of using an integrated actuator and ECU is offset by the complex design requirements for the housing and electronics.

2.4.2 Foundation Brake Actuator

A foundation brake actuator system – where the EPB actuator is combined with the foundation brake – can either take the form of an integral system (i.e., the actuator is installed in the foundation brake) or can feature a separately located actuator.

A system with separate actuator is usually favored when using fixed rear-wheel calipers. In this case, two system versions are possible: a version with an additional small, electrically actuated floating caliper (see Fig. 10) or, alternatively, a version with a duo servo drum brake "hat" fitted in the brake disk hub. This drum brake is electrically actuated to provide the parking brake function (see Fig. 11).

Integration of the actuator in the hydraulic service brake, on the other hand, is favored in the case of floating and fist-caliper disk brakes and for drum brakes (cf. Fig. 11). The caliper-integrated actuator is the most common system (combined caliper, see Fig. 12). In this case, a drive module with plastic housing, incorporating the motor and reduction gearing, is usually fitted to the caliper housing. The principal

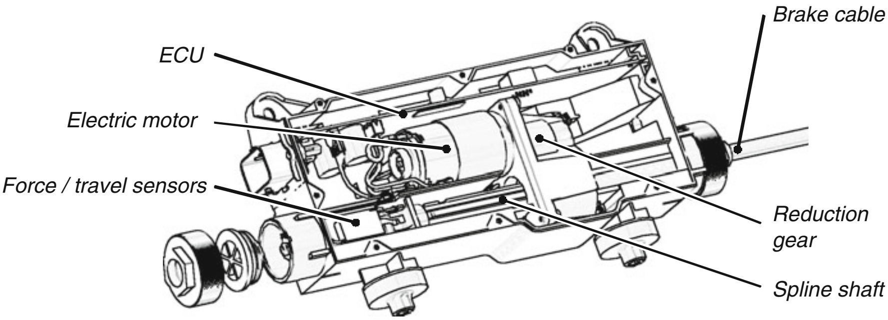


Fig. 9 Central actuator (single cable puller system)

Fig. 10 Fixed caliper (service brake) plus floating caliper (parking brake) (source: Brembo)

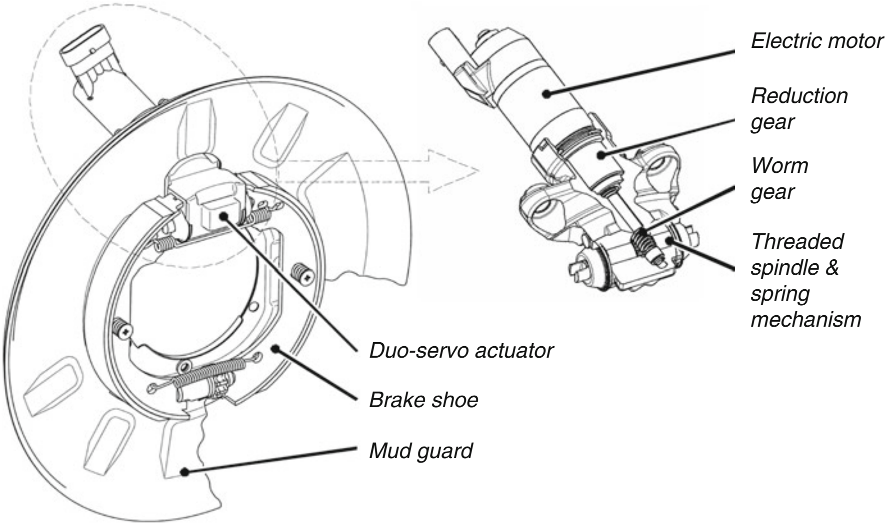


Fig. 11 Actuator for duo-servo drum-in-hat parking brake

components are spur gears, planetary gears, worm gears, and belt drives. Inside the caliper housing, a spindle-nut system converts rotational movement into translational movement. This system can be designed as a friction spindle or ball-screw system.

2.5 ECU Interfaces

The ECU requires at least the following information: wheel speed, vehicle speed information, and status information (e.g., ignition status). It has the following interfaces:

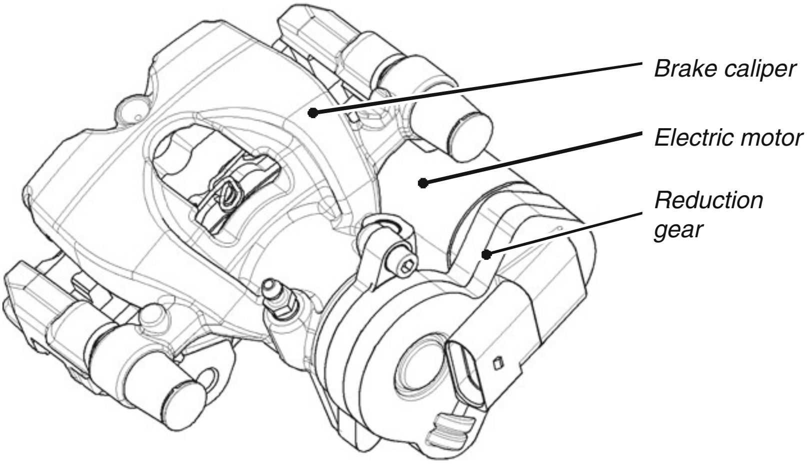


Fig. 12 Components of combined caliper-integrated actuator

- An interface to the user controls. These controls may be analog or digital, with single or multiple redundancies.
- Interface(s) to display system(s). In addition to the legally prescribed warning lights, audible signals and text messages, whose design will vary from one model to the next, are also customary solutions existing.
- Interface(s) to the vehicle bus system. This connection is provided via one or two high-speed CAN systems and/or via FlexRay. Communication protocol configuration varies from one model to the next.
- Interface to vehicle diagnostic systems (optional).
- Interface to vehicle power supply.
- Interface to the electric motor of the actuator.
- Depending on actuator technology and functional requirements: interfaces for sensor inputs.
- Optionally, inputs and outputs for switches, for function configuration.

If the EPB and ESC ECUs are integrated in the HECU, many of these interfaces are no longer required, either because the relevant information is already available in the HECU system or because the interfaces are implemented inside the integrated control unit. The resulting reduction in the number of components, and elimination of one housing, results in reduced costs, which has led to speedy take-up of this system version.

2.6 Functions

The basic function of the EPB, like that of its mechanical predecessor, is to provide parking brake functionality when requested by the driver. In addition, an electrically actuated parking brake also provides the basis for numerous driver assistance functions (Drive-Off Assist, Drive Away Release, Standstill Management, Hill Holder, Rollaway Prevention on a slope).

2.6.1 Secure Retention

Parking brake functionality primarily comprises the parking brake application/ release function.

When the brake is applied, a clamp force is produced which holds the vehicle in place under all operating conditions. Optionally, for convenience functions, a reduced clamp force/expansion force setting can be selected.

When the brake is released, the actuator is retracted far enough to completely eliminate the clamp force, while also maintaining the required air gap.

Although this core function of the parking brake is simple to describe, it is only possible to guarantee this functionality under all operating conditions using complex system functions and with the aid of signals from the vehicle. For example, automatic self-adjustment may be requested by a rollaway monitoring system or may be triggered on a time-controlled basis to compensate for cooling-related contraction of components.

2.6.2 Comfort Functions

The possibility of actuating the parking brake independently of the driver is used in a number of convenience functions. These convenience functions can be differentiated into two main groups: hold and drive-away assistance functions and automatic functions.

Comfortable drive-away assistance, whether after starting the engine of a vehicle with the parking brake engaged or after stopping on a hill, helps to make these driving situations easy to cope with for the driver. Depending on the vehicle manufacturer, functions in this group come in varying configurations, in some cases involving button-activated hold functions. With the engine running, implementation of the hold function is initially implemented by active pressure buildup in the hydraulic service brake, for example, via the ESC system, then automatically transferred, as and when required, to the EPB itself. If the engine stalls while moving off on a hill, the EPB is actuated directly.

Automatic functions are used to hold the vehicle in place and prevent it from moving – for example, after switching off the engine, after engaging an automatic transmission’s parking lock, after removing the ignition key, and/or after leaving the vehicle.

2.6.3 External Activation

Driver-independent activation of the parking brake is easy to implement by means of a software interface. This possibility is used by numerous assistance systems (e.g., adaptive cruise control with stopping capability). In some cases, parking brake involvement is an essential part of such systems’ safety concept.

Driver-independent EPB functionality capable of replacing the automatic gearbox lock and steering wheel lock and supporting the immobilizer has also been studied, but so far such a system has not gone into production.

2.6.4 Emergency Brake

Compliance with the legal requirement that it should be possible to use the parking brake/parking brake actuation system to rapidly decelerate a moving vehicle in the event of failure of the service brake actuation is, even on vehicles with electromechanical parking brake, normally ensured by an ESC function which actively builds up hydraulic pressure in the service brake at all four wheels – i.e., normally not by means of the EPB. If necessary, slip control is also integrated. Compared with a driver-operated mechanical parking brake, this system achieves a considerably higher braking effect while also maintaining vehicle stability and therefore offers a clear safety advantage. However, the proviso is that the vehicle must be equipped with an ESC system, in order to implement the necessary active pressure buildup.

A further form of emergency braking system can be implemented by EPB-actuated rear-wheel braking. Here too, slip control can be integrated. Such a function is used, for example, on vehicles with ABS (but not ESC) and as an additional fallback level for vehicles with ESC.

2.6.5 System Monitoring

Monitoring functions run continuously in the background. If they detect impaired functioning or function nonavailability, they prompt specific measures such as operation with function degradation, fault messages, and event logging in the fault memory. The diagnostic functions can be used to set the system during vehicle production, to read out the fault memory, and for flashing of control units with writable EPROMs.

2.6.6 Service Functions and Special Functions

When performing assembly or repair jobs, such as changing brake pads, the actuator can be retracted to its rearmost stop. This function is normally activated by means of a service or diagnostic tool.

During brake testing on a dynamometer, the front axle stands on a firm surface, while the rear axle is turned by a roller. Since this is not a normal driving situation, the EPB would normally register this as a system fault (front wheels stationary, rear wheels turning). To nevertheless permit dynamometer testing, the system incorporates a special function which automatically identifies this situation and ensures that the EPB can still be actuated.

3 Conclusion

After a long period dominated by hydraulic and mechanical systems, both for the regular service brake system and the parking brake, for some years now electromechanical solutions have also been in existence for both functions. Where parking brakes are concerned, electric parking brake (EPB) systems which use an electronic actuation system while retaining the friction foundation brake (disk, drum) are already in production.

As far as all-electromechanical service brake systems (EMB) are concerned, cost factors and the lack of full redundancy in the vehicle electrical system is currently still an obstacle to the adoption of such systems in production models. However, development of a production-standard hybrid brake system (EHCB) with integrated EPB functionality, which could serve as a stepping stone to all-“dry” mechatronic brake systems, is already well advanced.

References

Schmittner B (2003) Das Hybrid-Bremssystem – die Integration von ESP, EMB und EPB. In: Veranstaltungsunterlagen Autotec 2003. IIR Deutschland GmbH, Baden-Baden (Hybrid brake system – integration of ESC, EMB and EPB. In: Proceedings of Autotec 2003. IIR Germany GmbH, Baden-Baden)

Schmittner B (2004) Das Hybrid-Bremssystem – die Markteinführung der elektromechanischen Bremse EMB. In: Veranstaltungsunterlagen Bremstech Konferenz. TÜV Automotive GmbH/TÜV Akademie/TÜV Süd Gruppe, München (Hybrid brake system – market introduction of the electro-mechanical brake EMB. In: Proceedings of Braketech conference, TÜV Automotive GmbH/TÜV Academy/TÜV Group South, Munich)

Stemmer M, Münchhoff J, Schuster A (2012) Audi R8 e-tron: Pilotumsetzung und Potentialstudie des innovativen elektromechanischen Bremssystems EHCB. In: Veranstaltungsunterlagen 2. VDI-Fachkonferenz Innovative Bremstechnik. VDI Wissensforum GmbH, Stuttgart (Audi R8 e-tron: Pilot realization and feasibility study of the innovative electromechanical brake system EHCB. In: Proceedings of 2. VDI conference innovative brake technology, VDI Knowledge Centre GmbH, Stuttgart)

Strutz T, Münchhoff J, Rahn S, Schuster A (2013) Brake by Wire Anwendung am Beispiel des Audi R8 e-tron. In: Veranstaltungsunterlagen EuroBrake 2013, Dresden (Brake by wire by the example of Audi R8 e-tron. In: Proceedings of EuroBrake 2013, Dresden)

Steering Actuator Systems

31

Gerd Reimann, Peter Brenner, and Hendrik Büring

Contents

G. Reimann (✉) • P. Brenner • H. Büring
Robert Bosch Automotive Steering GmbH, Schwäbisch Gmünd, Germany
e-mail: gerdreimann@gmx.de; Peter.Brenner@bosch.com; Hendrik.Buering@bosch.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_32

Abstract

The steering converts the turning movement applied to the steering wheel by the driver into a change in the steering angle of the steered wheels. At the same time, its job is to inform the driver, by means of the haptic feedback, of the current driving situation and the road conditions.

Keywords

Angle superimposition • Electrohydraulic power steering • Electromechanical power steering • Hydraulic power steering • Rear-axle steering systems • Steer-by-wire steering system and individual wheel steering • Superimposing angles • Superimposing torques • Torque sensor • Torque superimposition

1 General Requirements Placed on Steering Systems

The steering converts the turning movement applied to the steering wheel by the driver into a change in the steering angle of the steered wheels. At the same time, its job is to inform the driver, by means of the haptic feedback, of the current driving situation and the road conditions. The steering system is therefore a crucial factor helping to ensure comfortable and safe control of the vehicle. Its main features are the following:

- The actuating force of the steering should be as low as possible and adjusted to the driving state. The requirement for low actuating forces in particular when the vehicle is stationary or slowly rolling has led to the situation where nearly all vehicles today are equipped with power steering. At the same time, however, it must be ensured when complying with this requirement that the low actuating forces do not lead to any loss in the haptic feedback from the roadway during fast driving and hence to uncertain and unstable straight-ahead driving.
- The number of steering wheel turns from one steering stop to the other should be as low as possible, while it is however also necessary to support the straight-line stability of the vehicle at higher driving speed by a not-too-direct steering ratio.
- The transmission of the steering wheel angle up to the wheel stop angle must be absolutely precise and backlash-free.
- As soon as the vehicle is moving, the wheels must revert back by themselves to the straight-line position once the steering wheel is released. This applies both for exiting from bends and for very minor steering movements on straight stretches, for example, during motorway driving.
- The feedback and jolts indicating the driving state and the roadway conditions must be strong enough to be noticed by the driver, but cushioned to prevent stress and fatigue for the driver.

The statutory requirements relating to steering systems in motor vehicles govern in particular the maximum permissible actuating force and duration in an intact and failure-prone steering system and are described in the European Directive 70/311/EEC.

2 Basic Solutions for Steering Assistance

Due to the requirements relating comfort and safety, power steering systems are now being used in all vehicle classes. Until a few years ago, these were mainly hydraulic systems. Continued development of electrical and electronic systems plus additional requirements, such as that of energy saving, has resulted in more and more electrically assisted steering systems being used, from small cars and compact and midrange vehicles up to those in the luxury class.

2.1 Hydraulic Power Steering (HPS)

A conventional hydraulic power steering system consists of integrated steering valve and hydraulic cylinder within the steering gear, a power steering pump, an oil tank, and tubes and pipes to connect these components (cf. Fig. 1).

The power steering pump powered directly by the combustion engine is designed to provide a sufficient oil pressure and oil quantity at the idling speed of the combustion engine. Since this design would lead to an excess delivery quantity at high speeds, for example, when driving on motorways, a valve is integrated to regulate the oil flow. To prevent overloading, for example, when steering against the end stop, a pressure relief valve is incorporated.

The power steering pump is connected to the steering gear by pipes/tubes. The expansion hoses used are able to absorb the pressure peaks caused by the power steering pump and by roadway impacts. They also ensure control stability of the hydraulic circuit.

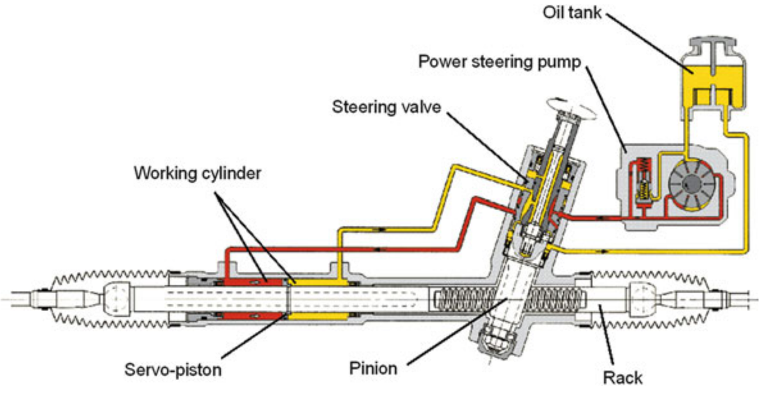


Fig. 1 System concept for hydraulic rack-and-pinion power steering

While hydraulic car steering systems once used above all the so-called recirculating ball-and-nut steering, the demands for compactness, low weight, and simple design have resulted in the use of rack-and-pinion steering gear in nearly all cars. The turning movement by the driver is converted by a pinion into a pushing movement of the rack. The connection to the wheels is made by means of tie rods and matching joints.

To control and convert the hydraulic auxiliary power, a control valve and a working cylinder are integrated into the steering gear. The control valve controls an oil pressure in the steering cylinders that corresponds to the turning effort of the driver. The turning of a torsion bar leads to a mechanical control travel in the steering valve which is proportional to the force. Due to this control travel, control edges designed as chamfers and facets are displaced and thus form the opening cross section for the oil flow. The steering valves are designed according to the "open-center" principle, meaning that the oil coming from the pump flows without pressure back to the oil tank when the control valve is not actuated.

The double-action steering cylinder on the rack converts the controlled oil pressure into an appropriate auxiliary force. Thanks to the control valve, the compartments of the steering cylinder are switched in the neutral position in such a way that an unhindered pushing movement of the rack is possible. By introducing a torque at the steering valve, the oil flow of the pump is diverted into the appropriate left-hand or right-hand cylinder compartment, generating the required force. Appropriate design and shaping of the chamfers on the control edges of the valve allows the relationship of the actuating torque at the steering valve and the force curve at the cylinder to be traced. This matching allows the intended individual steering characteristics of the respective vehicle to be achieved.

2.2 Parameterizable Hydraulic Power Steering

Rising demands for comfort and safety in the vehicle have resulted in the development of steering valves with electrically modulatable assistance characteristics. An electrohydraulic converter determines the hydraulic effect and hence the actuating force on the steering wheel. The electrical actuation of the converter is handled by an assigned electrical control unit. The main input signal for the control unit is the vehicle speed. The electric current in the converter is controlled in such a way that the steering assistance decreases as the vehicle speed increases. This results in high steering comfort, thanks to the low actuation forces at low vehicle speeds and high steering precision at high speeds (Fig. 2).

2.3 Electrohydraulic Power Steering (EHPS)

An alternative to conventional hydraulic power steering with a vehicle-engine-driven hydraulic pump is a system which uses a steering pump, driven by an electric motor. A major advantage of this system is the energy saving achievable with this

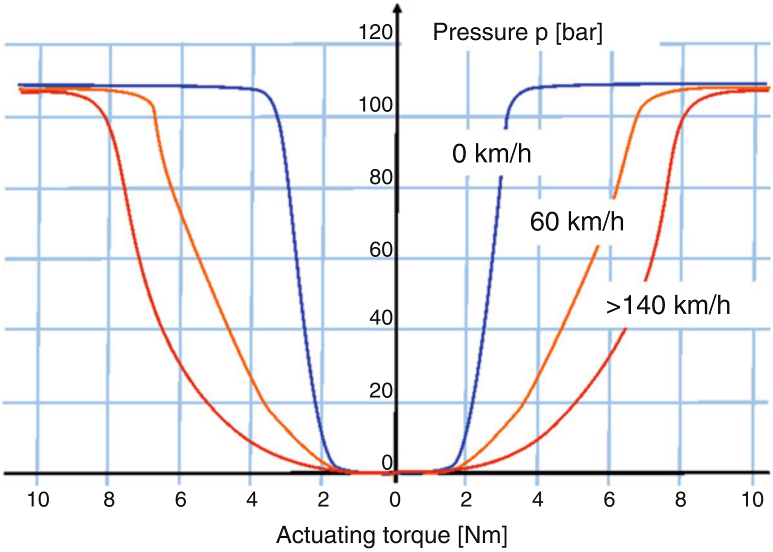


Fig. 2 Servotronic® valve characteristic

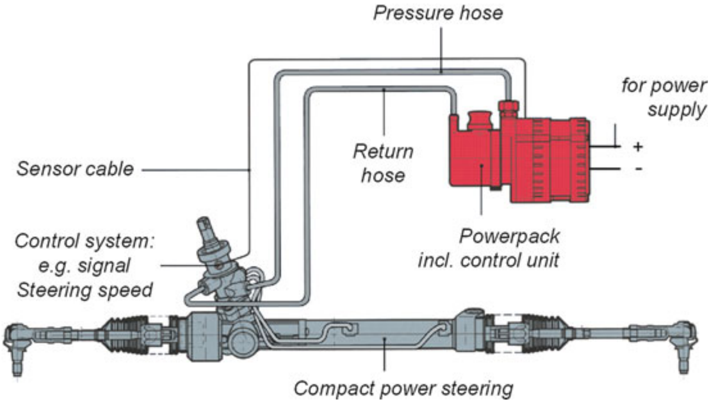


Fig. 3 System concept for electrohydraulic power steering

operation of the electric motor. The required electrical energy is drawn from the vehicle's electric power system, while the electric motor is operated by an electric control unit (Fig. 3).

The hydraulic pump is designed in current systems as a gear pump or roller vane pump. The electric motors used in the standard solutions are designed as brush-type or brushless DC motors. The required steering effort is ascertained by sensors in the steering system and from the vehicle. These are primarily the steering and vehicle speed which is evaluated by the electric control unit, from which the set speed for the electric motor is calculated and regulated using the integrated end power stage.

2.4 Electromechanical Power Steering (EPS)

Electromechanical power steering was developed to increase steering comfort, to further reduce energy consumption, and to simplify installation work in the vehicle. Originally used only in small vehicles, it is becoming increasingly widespread in all vehicles, including the luxury class. The basic operating principle is always the same: A torque sensor records the manual force of the driver, and an electric control unit evaluates these signals and calculates from them, taking into account further information from the vehicle such as vehicle speed, an appropriate set assistance torque for an electric motor. The latter is operated by an appropriate end power stage and passes its output torque to the steering via one or more gear stages. The type and design of the gear stages depend mainly on the requirements of the steering system in terms of installation space and the maximum steering assistance to be achieved.

2.4.1 Column-Type EPS

In case of vehicles with less strict demands placed on steering assistance and maximum steering speed, the servo force is usually introduced to the steering column. The servo unit, consisting of the torque sensor, the electric motor, and the reduction gear, is arranged inside the vehicle interior on the steering column. The electric control unit can here be designed separately as a remote solution or attached to the motor or sensor. The reducing gear is usually designed as a worm gear. In designing this gear stage, care must be taken that a sufficient back-turning efficiency is achieved in this unit in order to assure the necessary haptic feedback of the steering system to the driver or to assure back-turning of the steering system by itself when the steering assistance is switched off. The nonpositive connection to the steered wheels is achieved via the intermediate steering shaft and a mechanical rack-and-pinion steering gear (Fig. 4).

2.4.2 Pinion-Type EPS

A similar solution is pinion-driven electric steering. The servo unit, consisting of torque sensor, electric motor, gear stage, and a possibly integrated or attached electric control unit, is here arranged on the steering gear in the area of the steering pinion.

The assistance power provided by the electric motor via the worm gear is supplied directly to the pinion. The resultant advantages of this solution are a compact design and, in comparison to the steering column solution, a stiffer mechanical connection of the steering assistance to the rack. This leads not only to a possible higher assistance capacity but also to an improvement in steering precision. Disadvantages arise due to the more strict demands relating to environmental conditions, since the servo unit is installed in the engine compartment, exposing it to higher ambient temperatures and splash water (Fig. 5).

Fig. 4 Column-type EPS

Fig. 5 Pinion-type EPS

2.4.3 Dual Pinion-Type EPS

For a further increase in the assistance capacity and in steering precision, solutions are used which transmit the servo force directly to the rack. With the dual pinion solution, the force of the servomotor acts on a pinion, as in the pinion solution, via a worm gear. This is however arranged on a second and separate toothing on the rack.

Fig. 6 Dual pinion-type EPS

The spatial separation from the pinion permits higher flexibility during integration inside the vehicle. Thanks to the independence of the servo pinion to the steering pinion, the different objectives of these two pinion stages can be taken into account and so optimized in respect to comfort, capacity, and service life. The torque sensor for recording the steering torque introduced by the driver is, like in the following variants, arranged on the steering-spindle-side input of the steering gear (Fig. 6).

2.4.4 APA-Type EPS

A further possibility for converting the rotary movement of the servomotor into a pushing movement of the rack is to use a ball screw on the rack. This gear type combines very high mechanical efficiency, high loading capacity, and the absence of backlash necessary for precise steering. The transmission of forces with this gear type is from the ball nut via a continuous chain of hardened steel balls to the rack, which is provided with one or more ball screw threads. The ball nut is driven by an electric motor arranged parallel to the rack and connected to the ball nut via a toothed-belt gear stage. This gear stage too operates without backlash and with a very high mechanical efficiency. By appropriate design and selection of the transmission ratios, it is possible with a steering gear of this type, as is also the case with the dual pinion solution, to adjust the performance of the steering to the vehicle in question and to adjust the available motor capacity in the direction of high and maximum rack forces or toward high steering dynamics. Electrical steering gears with this design solution can be used in vehicles up to the luxury class and large SUVs (Fig. 7).

Fig. 7 APA-type EPS

2.4.5 Rack-Type EPS

The rack solution is a further option for transmitting the rotary movement of the electric motor to the rack. The ball nut of the ball screw is here driven directly by the electric motor without any additional gear stage. Therefore, the electric motor must be designed with a hollow shaft through which the rack is passed. A high degree of steering precision and dynamism can be achieved with this compact and direct connection of the motor, recirculating ball gear and rack. The lack of a gear stage when compared with the axis-parallel solution leads to the electric motor needing to have a comparatively high torque at lower speeds. The direct connection of the motor also requires a particularly high quality for the steering and motor control (Fig. 8).

2.5 Electrical Components

The general requirements placed on the electrical and electronic components of the EPS solutions presented are substantially identical. They differ only in the specific requirements relating to environmental conditions and the performance to be achieved.

2.5.1 Torque Sensor

The torque sensor is designed as a proximity-measuring angle sensor which records the angular rotation of a torsion bar and converts it into electric signals. The

Fig. 8 Rack-type EPS

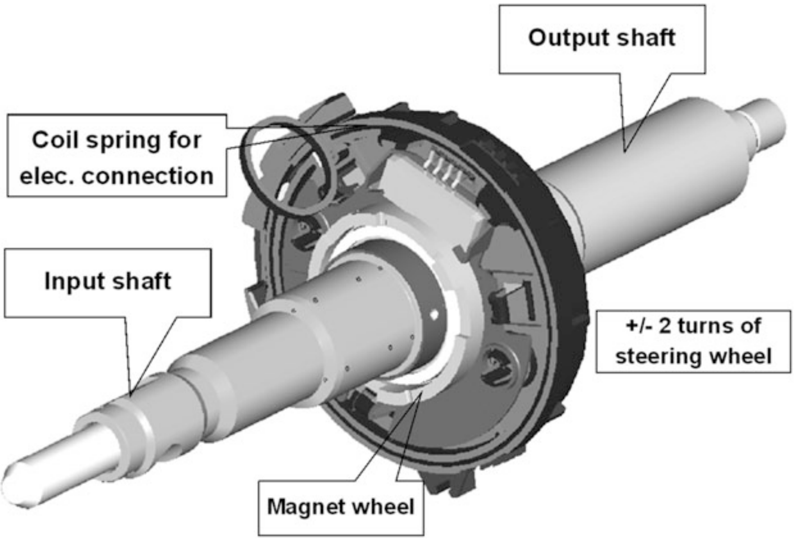
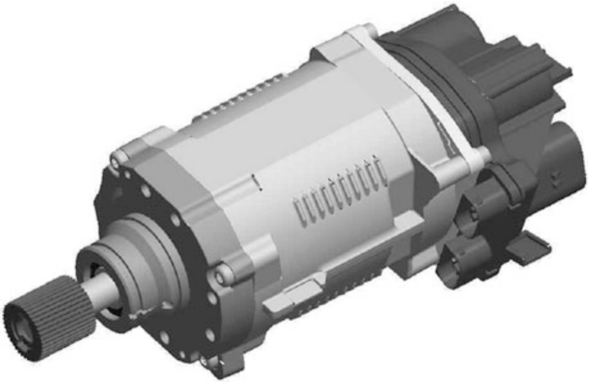


Fig. 9 Torque sensor

measurement range of a torque sensor for an electric power steering system is usually in the range from ±8 to ±10 Nm. In case of higher manual torques, a mechanical angle limiter on the torsion bar ensures that the latter is not overloaded. The electric control unit calculates the current torque value from the sensor signals. The high safety requirements expected from electric steering systems demand that all errors occurring in the sensor can be detected and result in a safe state in the steering system (Fig. 9).

Fig. 10 Electric motor



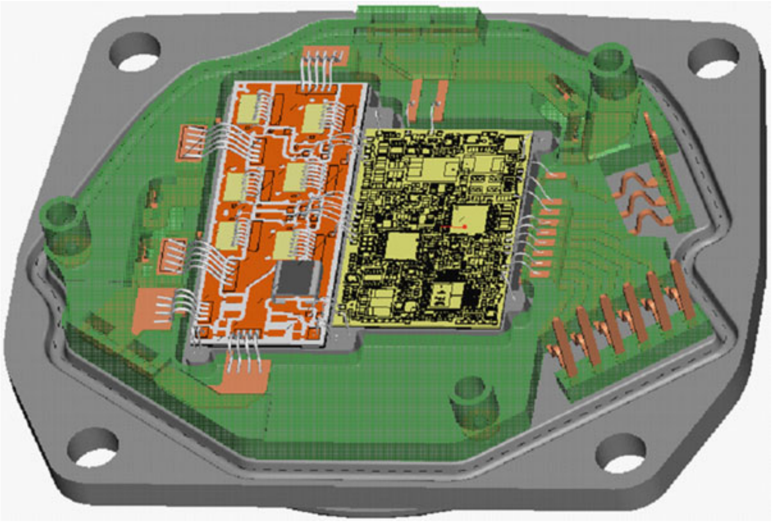


Fig. 11 Control unit

2.5.2 Electric Motor

The electric motors used in electric steering systems can be a brush-type DC or a brushless DC motor as well as an induction motor. Thanks to their sturdiness and the higher power output possible, it is the brushless motor variants that are increasingly being used. Steering power for vehicles in the upper midrange and luxury classes in particular calls for the use of highly effective and brushless DC motors. These motor variants require a motor position or motor speed sensor which is evaluated by the electric control unit and used for commutation and control of the motor (Figs. 10 and 11).

2.5.3 Control Unit

The associated electric control units contain one or more microprocessors which evaluate the sensor signals from the steering components and from the vehicle and then compute the set assistance torque and operate the motor using an appropriate motor control system and the end power stage integrated into the control unit with MOS field effect transistors. Sensors for recording the motor current and the control unit temperature are integrated into the control unit. To increase the steering comfort, further vehicle signals, in particular the vehicle speed and the steering wheel angle, are evaluated by the control unit. Using the information on vehicle speed, a speed-dependent steering assistance can be achieved which permits low actuating forces when the vehicle is stationary or moving slowly, while the steering assistance is continuously scaled back as vehicle speeds increase, improving the haptic feedback from the steering and also the directional stability. With the signals of the steering angle sensor, the return movement of the steering can be set and improved particularly for low and medium vehicle speeds and so adapted to the target vehicle. A multistage safety concept ensures that a reduction of the steering assistance ensues, gradually where possible, in the event of unusual states or errors. In the event of a complete failure of the steering assistance, it is ensured by the electrical and mechanical design that manual steering of the vehicle remains possible. The error memory of the control unit can be output at the diagnostic interface of the control unit, permitting an accurate diagnosis in the event of a fault. Operation of the electric motor by the software in the control unit permits a very sensitive and individualized adjustment of the steering assistance to the target vehicles. By evaluating further sensors from the vehicle or by providing appropriate communication interfaces with the control unit, it is possible with the EPS steering systems to implement efficient and innovative driver assistance functions (Fig. 11).

3 Solutions for Superimposing Torques

If haptic feedback is to be provided to the driver via the steering wheel or by autonomous assistance functions, a steering torque influence that can be activated independently of the driver is required in the steering systems. In the case of hydraulic steering systems, this is not possible without additional actuators, apart from the parameterizable hydraulic steering. Since this system however generally already requires a steering torque to be built up by the driver in order to vary the assistance, this cannot be regarded as a fully fledged solution for self-contained superimposition of assistance torques. If steering torque assistance is nevertheless to be implemented with a hydraulic steering system, then additional actuators are needed to do so.

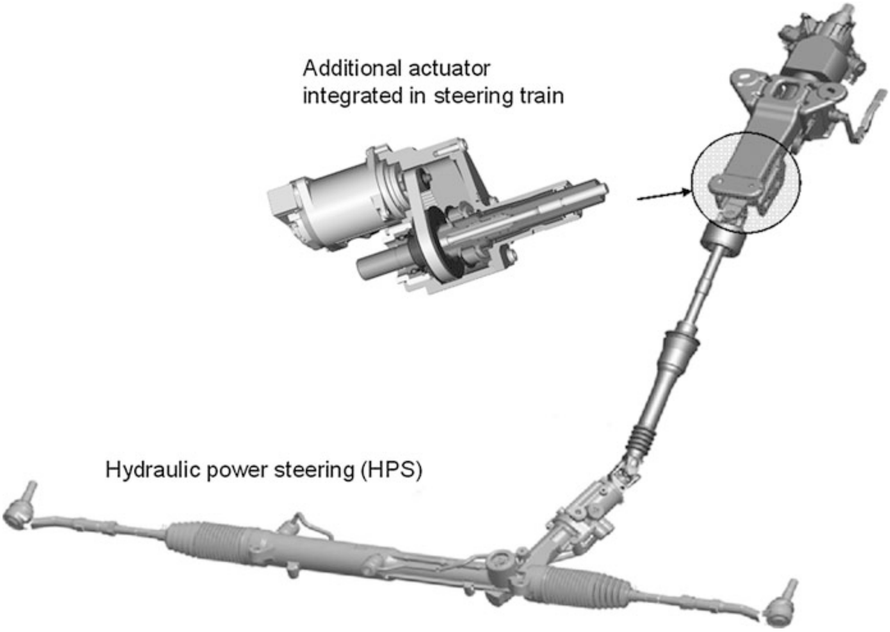


Fig. 12 Additional actuator

3.1 Additional Actuator for Hydraulic Steering Systems

An obvious solution for an additional actuator to implement steering assistance functions with a hydraulic or electrohydraulic basic steering is a steering actuator in which an additional torque, controllable independently of the driver, can be applied to the steering column via a gear stage and an electric motor. The design of an actuator of this type does not differ in principle from steering column EPS (Fig. 12).

Since however only one additional torque, and not the entire steering torque as is the case with EPS, has to be applied by this actuator, the dimensions of the mechanical and electrical components are markedly reduced. If a hydraulic steering system is to be operated with an actuator of this type, an effective torque of 8–10 Nm relative to the steering column is sufficient. Since this also results in lower demands on the loading capacity of the gear stage compared to steering column EPS, alternative and constructive solutions can be applied for the gear stage between the motor and the steering column.

It must however still be ensured that no disruptive torque unsteadiness that might irritate the driver is transmitted into the steering train by the additional actuator.

If this additional actuator is to be used for implementing functions which require a steering torque from the driver, a torque sensor must be installed either in this actuator itself or at another suitable position in the steering train between the driver and the actuator.

Since this is purely an additional system for an already installed power steering system, the safety considerations here are focused on this additional actuator and on the control units connected to it. If the degradation stages in the event of detected faults in the assistance functions are disregarded, the actuator must, whenever an error is detected in the motor or in the sensors connected to it, be put in a state that largely reduces disruptive additional torques and rules out dangerous ones. This means that the motor must be either mechanically disconnected from the steering train using a coupling or the motor must be switched off in such a way that it can on no account build up disruptive braking torques that prevent safe driving of the vehicle.

3.2 Electric Steering Systems

Electric power steering offers ideal conditions as an actuator for steering-based assistance functions, since the electric motor is operated using the software of the control unit. The torque sensor implemented for recording the driver steering torque in the EPS can also be used for the assistance functions to be provided. The electric motor, already firmly connected to the steering train via an appropriate gear stage, can be used not only for providing the servo power but also and at the same time to apply the assistance torque required by the higher-level systems. Since the magnitude of the assistance torque is several times less than the servo torque, it is not necessary as a rule to design the electric motor of the EPS and to increase its power output in order to provide the additional torque (Fig. 13).

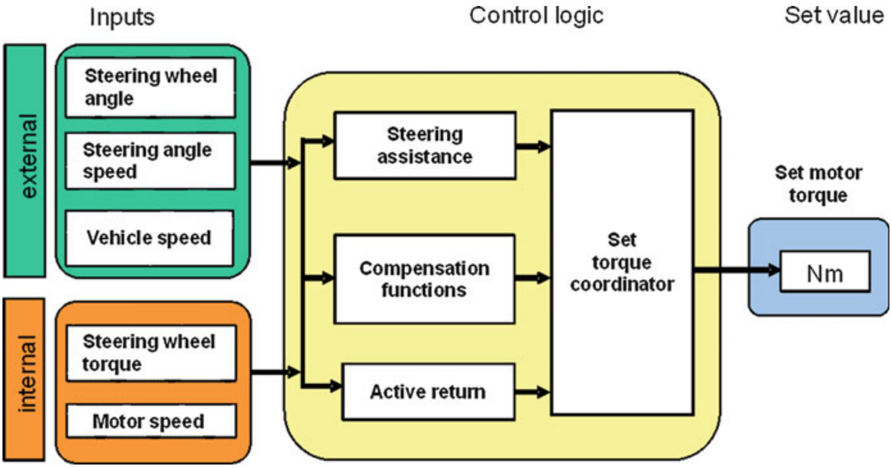


Fig. 13 Control structure of EPS

3.2.1 Torque Superimposition

Even in the case of electric power steering without a connected assistance system, the motor torque is made up of different components and determined as the set torque for the motor control algorithm (direct assistance torque setting or triggering by predefined functions, e.g., vibration). The most important individual components here are the steering assistance varying with vehicle speed and active return of the steering to the straight-line setting of the steering, plus active damping and friction compensation functions. These differing individual set torques for the electric motor are collated by a torque coordinator and added up to a total set torque, if necessary allowing for priorities of the individual functions. The additional torque required by the external assistance system is thus provided via a further input into the torque coordinator and hence considered as an equally valid or appropriately prioritizable individual set torque by the electric power steering.

In view of the possibly limited options for data transmission on the bus system between the assistance control unit and the electric power steering, it is additionally possible, for direct determination and transmission of the assistance torque, to simulate predefined superimposition functions in the software of the steering control unit and to trigger them by a single control instruction via the data bus. This can be useful in the case of functions which trigger an oscillating additional torque, for example. In this way it is possible to trigger a lane-departure warning using only one control instruction containing information on the amplitude and frequency to be set for steering wheel vibration (Fig. 14).

3.2.2 Angle Superimposition

For assistance functions, for example, automatic maneuvering into a parking space, an angle determined by the higher-level control unit is required. However, since the

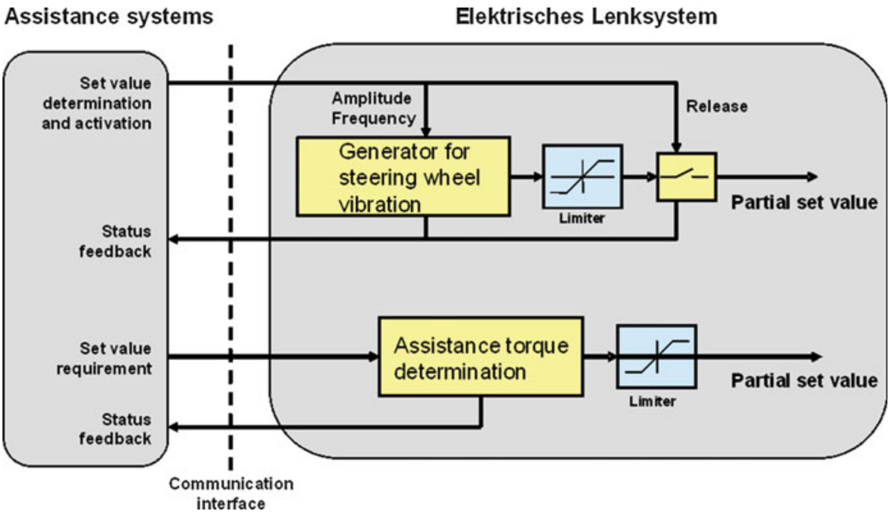


Fig. 14 Predefined superimposition functions transferred to steering control unit

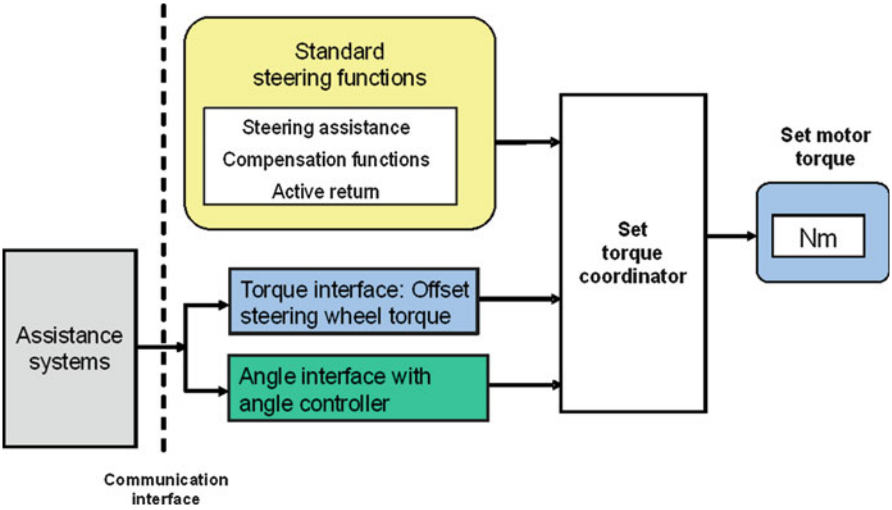


Fig. 15 Superimposition of assistance and standard steering functions

control concept of electric power steering is primarily a matter of torque control, autonomous adherence to a set lane requires a control algorithm that calculates a variable for the electric motor, in the form of a torque requirement, from the set and the actual steering angles of the steering, and stipulates this variable. This angle controller is best integrated into the software of the steering, since the CAN bus mainly used in the vehicle for data transmission at present does not permit time-synchronous transmission. The running time fluctuations inevitable for that reason make it impossible to provide the necessary quality of angle control (Fig. 15).

When an appropriate angle requirement is made to the steering, the actual assistance torque functions are then deactivated and the angle control circuit handles the determination of the set value for the electric motor. For automatic maneuvering into a parking space in particular, it is possible to detect, by evaluation of the torque sensor of the steering, whether the driver intervenes in the steering process and intends to abort the function.

4 Solutions for Superimposing Angles

4.1 Introduction

Conventional steering systems always work with a fixed transmission ratio, for example, 1:18. This is a compromise to ensure that on the one hand minor steering corrections on the motorway do not greatly impact stability and on the other hand that the driver does not have to turn the steering wheel so much in city traffic or when parking. Superimposed steering or active steering by contrast varies the

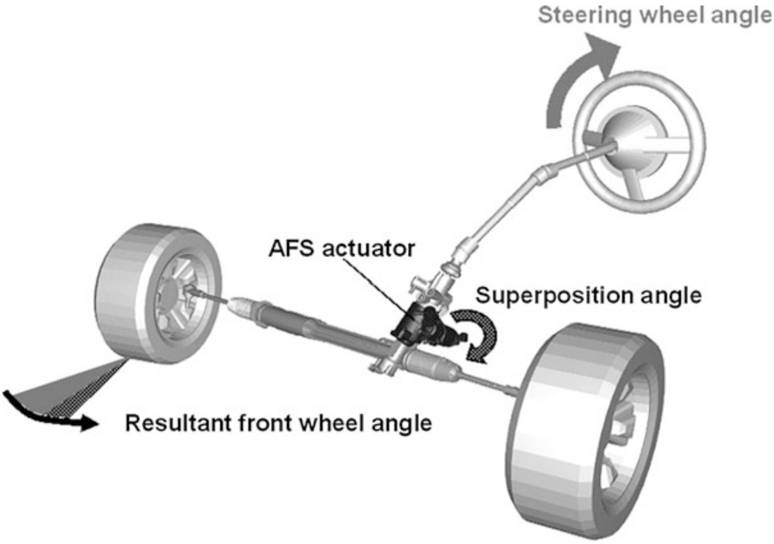


Fig. 16 Principle of superimposed steering (VDI/GMA Fachtagung 2004)

transmission ratio actively and dynamically, from around 1:10 when stationary and up to about 1:20 at high speeds.

Superimposed steering permits both a steering intervention depending on the driver (dynamic) and an active steering intervention at the front axle, without having to disconnect the mechanical coupling between the steering wheel and the front axle (Fig. 17) (VDI/GMA Fachtagung 2004). The additional degree of freedom permits continuous and situation-dependent adaptation of the steering properties. Comfort, steering effort, and steering dynamics are as a result actively adjusted and optimized. Moreover, steering interventions to improve vehicle stabilization are also possible. These are superior to those in existing systems, since the response behavior is faster by an order of magnitude, so that interventions taking place are hardly perceptible. The system limits, the function scope, and the necessary system interface should be defined such that the system is independent of other chassis control systems (Fig. 16).

4.2 Functionality

The active steering system has a complex functionality consisting of kinematic and safety functions (Reinelt et al. 2004; Eckrich et al. 2002).

Based on the signals of the vehicle sensors (steering wheel angle, speed, etc.), the assistance and stabilization functions (e.g., variable steering ratio and yaw rate control) compute a required superimposition angle. This acts as the desired value for the controlled actuator, which emulates as precisely as possible the time response of the required superimposition angle. A safety system monitors and

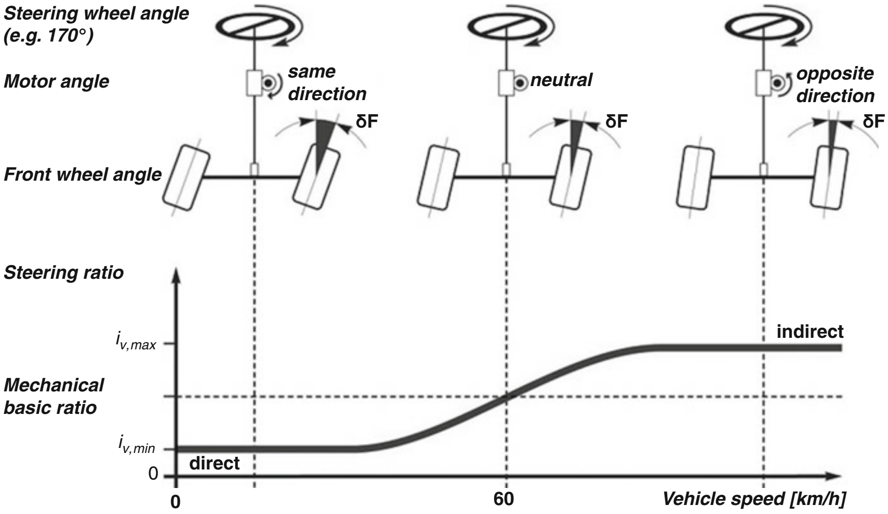


Fig. 17 Principle of variable steering ratio (VDI/GMA Fachtagung 2004)

checks correct functionality of the entire system. The measures range from a differentiated switch-off of part functions to a complete shutdown of the actuator.

Steering assistance functions are preliminary control actions of the steering system with the aim of adapting the static and dynamic steering properties to the driving situation depending on the driver's steering activity. This adaptation is restricted mainly by the actuator dynamics and the steering feel (feedback to the driver).

Figure 2 shows the variable steering ratio as the kinematic steering assistance function. This function $i_V(v_X(t)) = \delta_S(t)/\delta_F(t)$ is used to change the ratio between the steering wheel angle $\delta_S(t)$ and the mean front wheel angle $\delta_F(t)$ depending on suitable vehicle and steering quantities, such as vehicle speed $v_X(t)$ and deflection. The dependence on speeds permits, thanks to a more direct ratio, a reduction in the steering effort in the lower- and medium-speed ranges. Precise lane keeping and safety in the upper speed range are achieved by an indirect ratio. Furthermore, the dependence on deflection optimizes accuracy in the medium range, reduces the steering effort for large steering angles, and permits a modification of the steering behavior in the case of constant steering kinematics (Fig. 17).

4.3 Actuator Variants

The actuator with superimposed gear can be integrated in the steering column or alternatively in the steering gear. Integration in the steering gear is advantageous in terms of the haptic effect, since the friction up to the steering valve is not affected and acoustic propagation in the form of airborne sound in the engine compartment

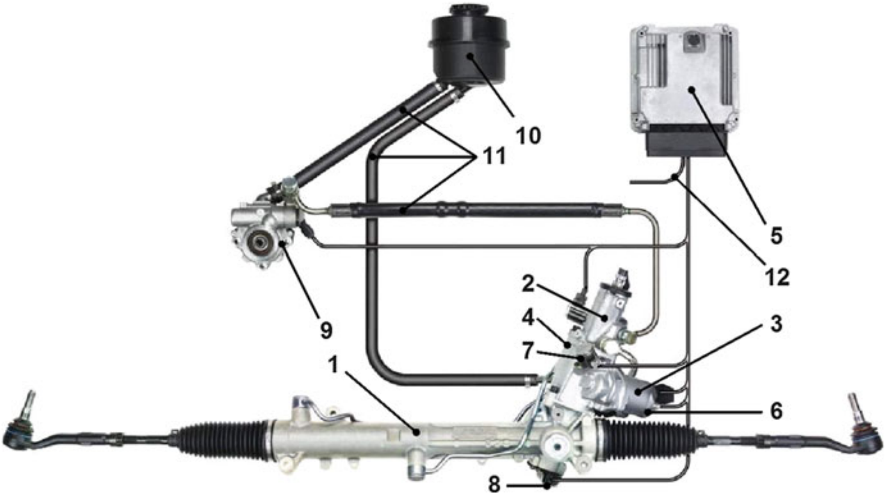


Fig. 18 Components and subsystems of the superimposed steering integrated in steering gear (VDI/GMA Fachtagung 2004)

is less noticeable. The steering column solutions are all designed fixed to the vehicle. The dynamic requirements are comparable in each case.

Variant 1 positions the electric motor transversely to the superimposed gear (cf. Fig. 18). The greatest advantage is the use of a self-locking worm gear to prevent the undesired back-turning in the passive state. For the integrated solution inside the steering gear (in the engine compartment), sufficient installation space must be available and taken into account during the early concept phase of vehicle development. Installation of this variant in the upper part of the steering column is also possible here. Meeting the package conditions and crash requirements for modern vehicles would however appear to be difficult with this actuator version.

Variant 2 involves the coaxial arrangement of the superimposed gear and electric motor (cf. Fig. 22). The use of a hollow-shaft motor in conjunction with a strain wave gear is required here. This combination is very compact and also advantageous with regard to package and crash behavior when installed in the steering column. The haptic effects are insignificant, since the strain wave gears used work almost free of backlash.

Installation of this actuator is also suitable as a fixed-to-steering shaft variant. The upper steering shaft is here firmly connected to the actuator housing and rotates it too (cf. Fig. 26).

The technical criteria determining the development of a superimposed gear are:

- Achievable dynamics
- Pleasant steering feel
- Meeting the radial installation space requirement

- Meeting the axial installation space requirement
- Low noise behavior
- Controlled back-turning behavior
- Suitability for absence of backlash
- Low weight

4.4 Example of Use in BMW E60: Actuator on Steering Gear

The practical implementation of the superimposed variant 1, integrated in the steering gear, is made up of the following parts (Fig. 18):

Rack-and-pinion power steering consisting of a steering gear (1), a Servotronic valve (2), an electronically controlled steering pump (9), an oil tank (10), and appropriate hoses (11).

Actuator consisting of a brushless DC motor with appropriate cables (3), a superimposed gear (4), and an electromagnetic lock with appropriate cables (7).

Control system consisting of a control unit (5), a pinion angle sensor (8), a motor angle sensor (6), appropriate software modules, and cables between the control unit and the sensors and actuators.

Brushless DC motor generates the required electrical torque for a required movement of the actuator. The electrical torque is under field-oriented (FO) control.

Motor angle sensor is based on a magnetoresistive principle and includes a signal booster and temperature compensation in the sensor module.

Pinion angle sensor is based, similarly to the motor angle sensor, also on a magnetoresistive principle and contains a signal booster and temperature compensation. With a CAN interface, the sensor signal can be used by other chassis systems, e.g., ESP.

Electromagnetic lock blocks the worm during system shutdowns: A spring presses the metallic pin of the lock against the locking teeth of the worm (see Fig. 19). This mechanism is opened (unlocked) by a specific current control action from the control unit.

4.4.1 Actuator with Lock and Pinion Angle Sensor

The core of the system is the mechatronic actuator between the steering valve and the steering gear (Fig. 19). This includes the superimposed gear (planetary gear) with two input shafts and one output shaft. One input shaft is connected to the steering wheel via the steering valve and the steering column. The second input is driven by an electric motor via a worm gear as a gear reduction stage. The pinion angle is applied as a weighted sum at the output shaft, which acts on the input of the steering gear, i.e., on the pinion of the rack-and-pinion steering. The steering kinematics determined by the steering gear and the geometry of the steering linkage are effective between the input of the steering gear and the front wheel.

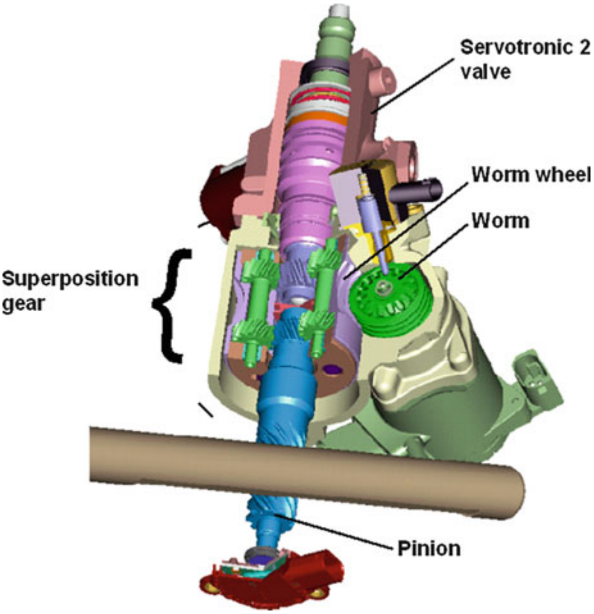


Fig. 19 Section through the superimposed actuator

4.4.2 Control Unit (ECU)

The control unit represents the connection between the vehicle electrical system, the sensors, and the actuators (Brenner 2003). The core components of the control unit are two microcontrollers. These controllers perform all the necessary computations for actuator control and for the utility and safety functions. The electric motor, the electromagnetic lock, the controlled pump, and the Servotronic are operated via the integrated output stages. In addition, the microcontrollers perform redundant computations, thus representing part of the safety concept.

4.4.3 Signal Flow

Figure 20 (VDI/GMA Fachtagung 2004) shows the signal flow: The signals of the steering wheel angle and the vehicle variables (e.g., yaw rate) are processed in the control unit, and the set values of the steering assistance and stabilization functions are computed. This is followed by the coordination of the steering requirements, actuator control, and operation of the electric motor. The actual value of the motor angle is reported back to the controller. All modules are monitored using the safety functions and failure strategy.

4.5 Example of Use in Audi A4: Actuator in the Steering Column

The components are comparable to those in the previous example (see Sect. 4.4). The actuator is in this example integrated behind the steering console in the upper

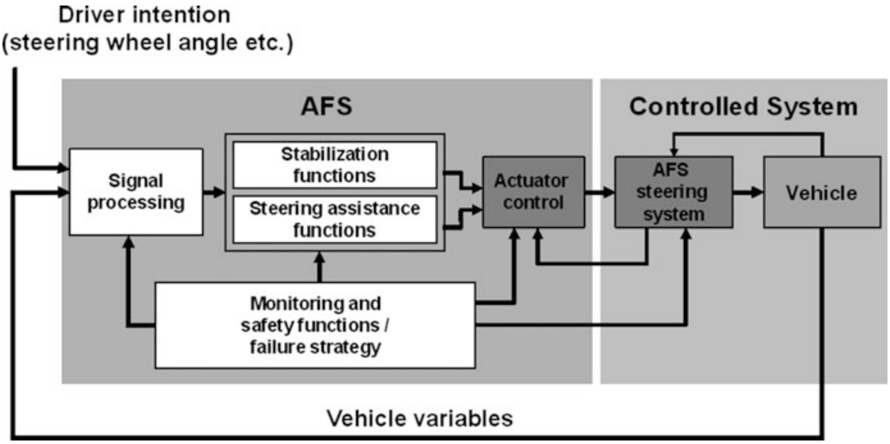


Fig. 20 General signal flow (VDI/GMA Fachtagung 2004)

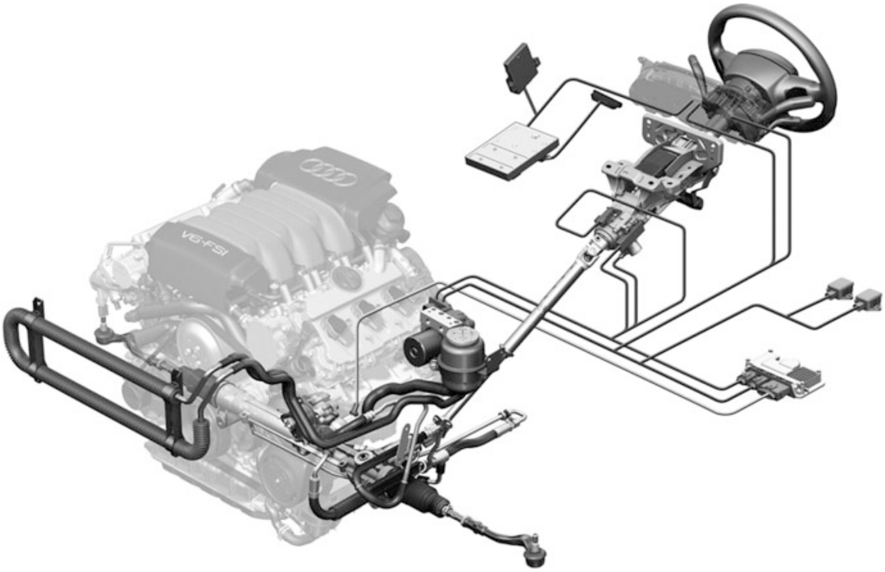


Fig. 21 Components and subsystems of the superimposed steering integrated in the steering column (Schöpfel et al. 2007)

steering column. The compact design of the coaxial arrangement of motor and gear permits positioning above the footwell (Fig. 21).

4.5.1 Actuator with Lock

The high-reduction wave gear is combined with an electronically commutated DC motor and a locking unit that locks the electric motor in the current-free state.

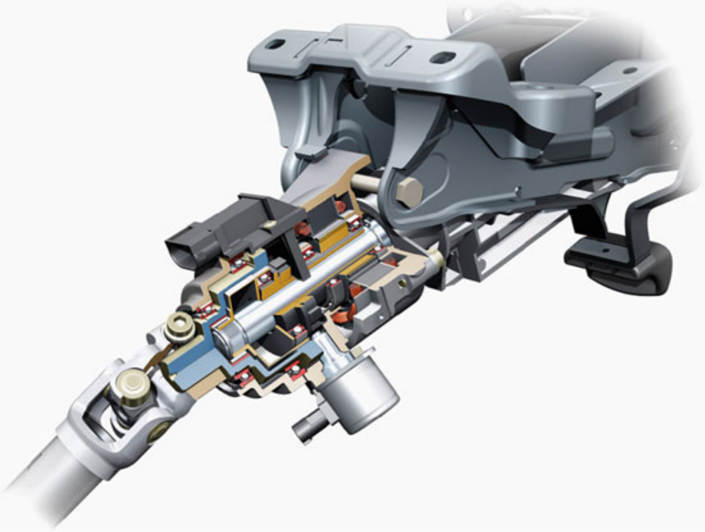


Fig. 22 Sectional view (Schöpfel et al. 2007) and sketch of actuator inside steering column

The motor must be designed with a hollow shaft. The steering-wheel-side shaft is positively connected to the flexible gear cup (flex spline) (Schöpfel et al. 2007). The rotary movement of the steering wheel is transmitted to the output shaft on the steering train side by the outer toothing of the flexible cup via the hollow wheel (circular spline). This force flow also corresponds to the direct mechanical link between the steering wheel and the steering gear in the locked state of the motor (Fig. 22).

4.5.2 Angle Superposition

Angle superposition is achieved by the hollow shaft of the electric motor, formed on the gear-side end as an elliptical internal rotor (wave generator). The generator warps, using a flexible thin-ring ball bearing, the thin-walled flexible cup connected to the steering input shaft. The outer toothing on the flexible cup is, at the high axis of the elliptical rotor, engaged with the hollow gear of the output shaft. Due to the differences in the number of teeth between the flexible cup and the hollow wheel (steering gear side), the result during rotation of the elliptical rotor is superposition (Fig. 23).

4.5.3 Control Unit and Safety Concept

The electronic control unit also meets all the requirements as stated in the example for use 1. The difference is in the 1-processor concept with a smart watchdog (Schöpfel et al. 2007). To meet the safety requirements, all functions must be present in a redundant way (designed independently duplicated).

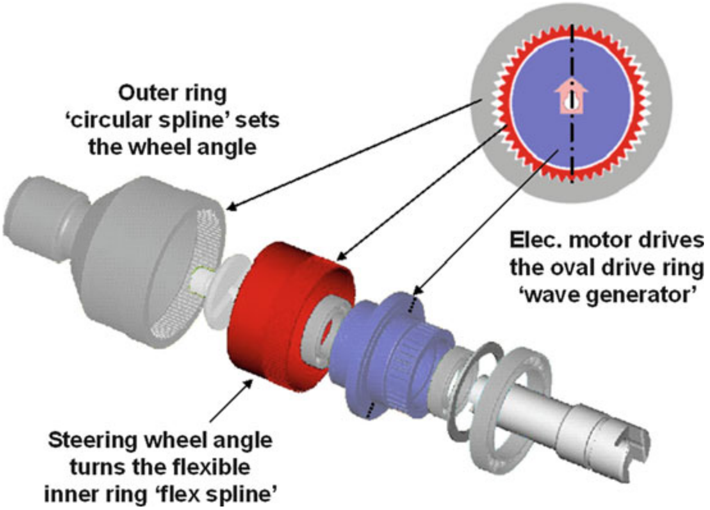


Fig. 23 Superimposition principle, wave gear

At the start are signal processing and a signal plausibility check. In addition, this module computes the vehicle speed. The variable steering ratio function inputs these signals and computes the steering angle correction. As a further task, it synchronizes harmoniously a nonmatching wheel position relative to the steering wheel. This kind of asynchronicity can occur, if in the inactive state, for example, when the combustion engine is switched off, major steering wheel movements have taken place. The sum of these partial angle values is added up together with the processed ESP set partial angle in the coordinator to obtain a total set angle (Fig. 24).

The position control and the motor commutation have the task of passing on the set angle to the output stage driver with the required control quality. The installation position of the superimposed gear between the steering valve and the steering wheel leads to a direct haptic feedback to the driver. This essential condition places heavy demands on the permissible torque ripple of the electric motor.

The control unit must also electronically detect failures and prevent their effects. The derived requirements placed on the control unit are (Schöpfel et al. 2007):

- Avoidance of reversible and irreversible faulty setting requirements that may be caused by the control unit, the electric motor, or the motor position sensor
- Monitoring of externally computed stabilizing interventions and initiation of suitable measures so that the maximum permissible number of faulty setting requirements is not exceeded
- Ensuring that in the event of error, the maximum tolerable jump in the ratio is not exceeded
- Prevention of an uncontrolled steering situation

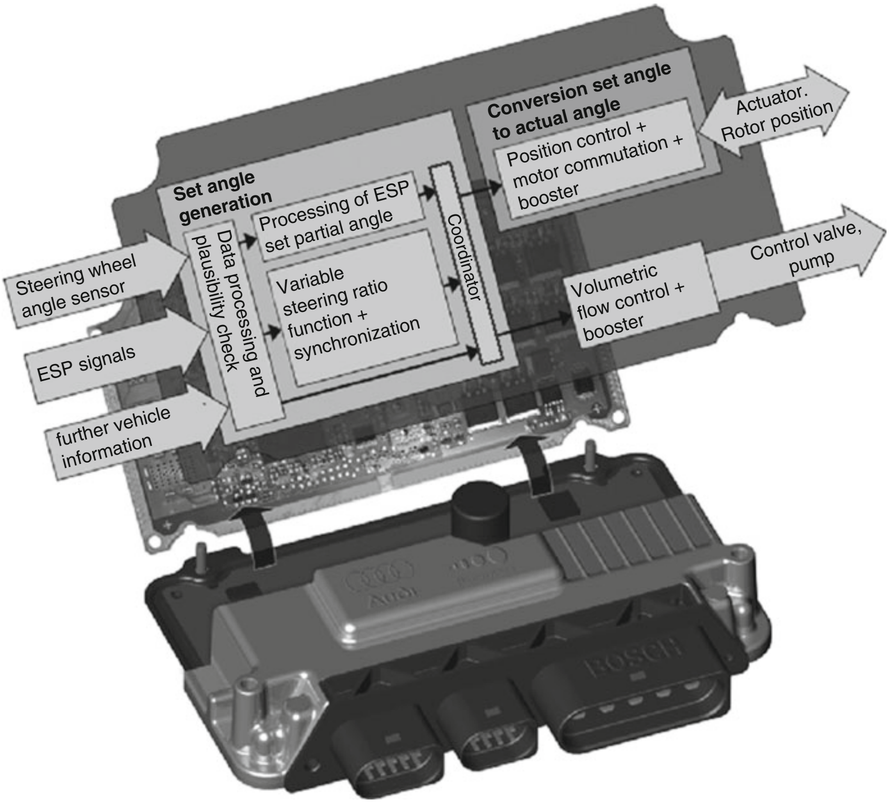


Fig. 24 Control unit with SW architecture (Schöpfel et al. 2007)

Figure 25 shows the three-level safety concept of the control unit (Schöpfel et al. 2007). In the first level, all software modules are integrated which are necessary from the functional viewpoint, including signal plausibility check and failure strategy. All critical paths that can lead to a failure are computed in redundant manner in the second level. This ensures that systematic error causes (e.g., programming errors) cannot lead to a failure. The third level ensures the program sequence and correct performance of the instruction set.

To ensure high availability, a gradual degradation of system functionality is required (Schöpfel et al. 2007) depending on the error which has occurred:

- Setting a constant steering ratio when there is a lack of driving speed information
- Blocking of external stabilizing interventions when low performance is expected, for example, due to fluctuations in the vehicle's electrical system
- System deactivation when steering angle reaches zero if an error is suspected, in order to prevent a misaligned steering wheel
- Complete and immediate system deactivation

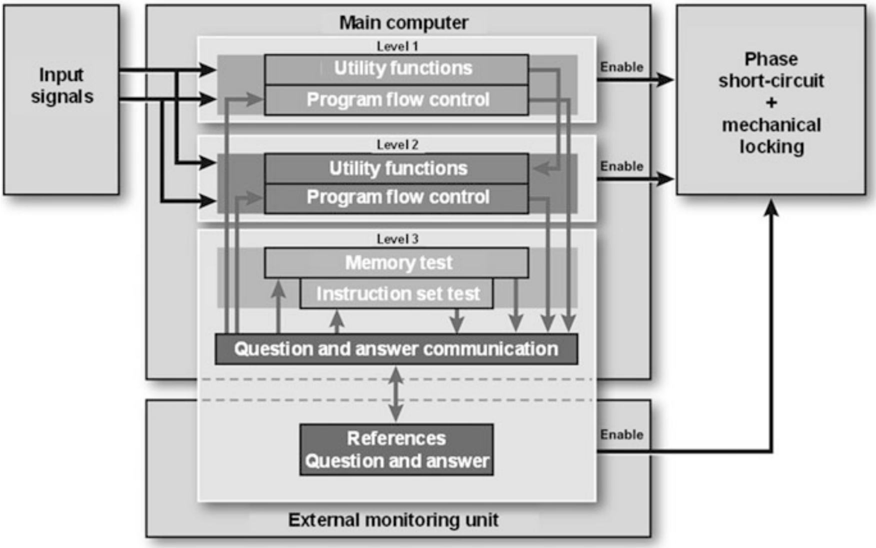


Fig. 25 Three-level safety concept of control unit (Schöpfel et al. 2007)

Furthermore, availability after a deactivation can be restored by an initialization phase, without the need for time at a garage. In addition to preventing failures, the control unit must continue to supply safety-relevant signals for the other vehicle control systems.

5 Steer-by-Wire Steering System and Individual Wheel Steering

All standard steering systems developed to date for cars are based on a dependable mechanical coupling between the steering wheel and the wheels. The driver thus has, in all operating conditions of the vehicle, a direct mechanical link to the steerable wheels, enabling him to follow directly his intended driving route.

The continuing developments in the steering sector made in recent decades by the steering manufacturers and the vehicle industry relate largely to assistance of the steering power or to steering angle superposition. For example, hydraulic or electromechanical power steering systems offer perfectly adjusted steering power for all possible driving states, but remain based on a mechanical transmission mechanism. In the event of errors in particular, i.e., when power systems change to the so-called fail-safe or fail-silent modes, mechanical components take over the task of transmitting the steering command of the driver to the wheels. This aspect

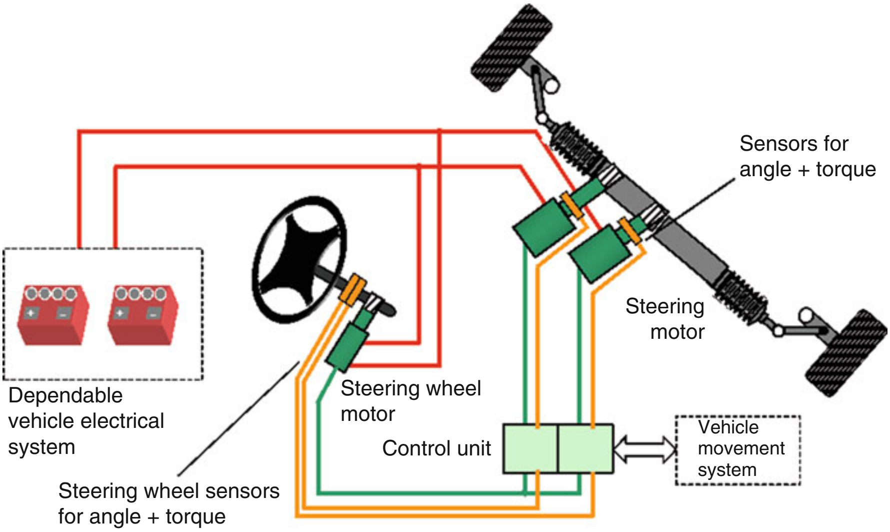


Fig. 26 System structure of steer-by-wire steering system

remains important even in steering systems with angle superposition (active steering).

Steer-by-wire steering systems represent a new approach. This is characterized by a purely electronic transmission of the driver's steering intention or of a complete decoupling of the driver's mechanical steering movement and the steering of the wheels. This obviates the need for conventional mechanical transmission devices. The driver generates at the steering wheel only information about his intended steering movement. This information is fed to an electronic control unit. This control module evaluates the information and converts it into appropriate steering commands. This operates the steering gear which performs the intended steering movement (Fig. 26).

- With the aid of hydraulic, electrical, electronic, and sensor systems, many new comfort and safety functions were developed in the past, to make driving a vehicle much more comfortable and safer.
- Despite all these components, the safety concept of current steering systems is still based on a continuous chain of proven mechanical components.
- Steer-by-wire systems clearly differ in their safety concept from conventional steering systems. In the event of an error, shutdown of the system into the fail-silent mode is not sufficient. Instead a fail-operational mode is needed, using a redundant replacement system with the full range of functions.
- For the market launch of the steer-by-wire system in cars, a classic mechanical or hydraulic fall-back level is probably needed as the safety concept for the first phase of confidence building.

5.1 System Concept and Components

A steer-by-wire concept is made up mainly of two assemblies: a steering wheel actuator and a wheel actuator.

5.1.1 Steering Wheel Actuator

The **steering wheel actuator** in the area of the upper steering column comprises a conventional steering wheel with sensors to record the steering wheel angle and the steering torque and a steering wheel motor to pass the appropriate steering feel on to the driver.

In addition, familiar control elements reduce accident risks, thanks to long years of practice with them, in the event that steering corrections by means of reflex movements are required in critical driving states.

5.1.2 Wheel Actuator

The **wheel actuator** consists mainly of an electromechanical rack-and-pinion steering. For safety reasons, the rack is driven by two redundantly designed electric motors. The high-performance electric motors are usually designed as brushless permanent-solenoid-excited DC motors (BLDC). Sensors are also installed in the wheel actuator for recording the wheel angle.

5.1.3 Electronic Control Unit

An **electronic control unit** processes all information provided by the two assemblies and the data available from other vehicle systems. For safety reasons, a redundant system structure is used consistently. In some cases, this requires up to three sensors independent of one another for a single safety-relevant signal. Only then is a dependable fail-operational mode of the system ensured in the event of an error. Depending on the functional and safety structure, up to eight 32-bit microprocessors are needed in the control unit, which mutually monitor each other for plausibility of the computed set values or rather for failures.

5.2 Technology, Advantages, and Opportunities

On the one hand, the technical latitude for designing steering functions for their comfort, safety, and driver assistance aspects offers excellent opportunities for steer-by-wire concepts. Depending on the available sensor signals and on the integrated network with other vehicle systems, it is possible to make driving the vehicle as safe and easy as possible for the driver in all conceivable operating conditions.

As the previously mentioned experience with electromechanical steering and active steering has shown, it must be ensured that newly developed functions and design principles are regarded as supportive and helpful by all drivers. Stabilization functions in particular which are based on automatic driver-independent steering interventions should not be perceived by the driver as a loss of responsibility for the respective driving situation.

A further important point in steer-by-wire systems relates to the haptic information to be imparted in real time when handling the steering and which must describe the tire/roadway frictional connection as precisely as possible. This information is highly valued by the driver, since he can use it to assess the right driving speed and the available acceleration and deceleration capacities of the vehicle. It is usually also the only information source which supplies him quickly enough with knowledge of abruptly changing roadway friction coefficients, so that he can reflexively get a dangerous situation back under control based on practiced behavior patterns.

This so-called feedback information that imparts a familiar steering feel to the driver must be generated artificially by the steering wheel motor in the steering wheel module in the case of steer-by-wire. Depending on the available sensor data, the electronic control unit computes a setting value for the steering wheel motor which thus simulates a steering resistance at the steering wheel. This should ideally reproduce the tire/roadway frictional connection conditions at a suitable force level.

Resetting forces during cornering can also be simulated in this way. When the steering wheel is moved, the steering wheel motor counters the movement direction and the movement torque to a level that can be fixed as required, regardless of whether the axle resetting forces of the vehicle achieve ideal values or not. Even an end stop can be simulated with a blocking torque in the steering wheel motor, without a mechanical stop in the upper steering column being needed.

Disturbance forces acting on the steered wheels, for example, tire imbalance, pothole effects, etc., can simply be selectively faded out or simulated at the steering wheel with any required intensity. This can be scaled in any way required as part of the design of the control software and would in the case of traditional steering systems have required at least design measures for the mechanics or hydraulics.

In the same way, the steering system can be adapted optimally to any vehicle using the parameterizable software. Even self-steering behavior, such as oversteering or understeering, can be influenced in this way, to impose on every vehicle model the required brand character, also known as “blend-by-wire.” It is even conceivable to accommodate to the personal driving style of the individual driver by individually controlling his preferred steering parameters.

As regards driver assistance and stabilization functions, it is of course possible to implement all the solutions already described and practiced in electromechanical power steering and in active/superposition steering, such as variable speed-dependent ratio, steering lead, yaw rate control, yaw moment compensation, side wind compensation, automated parking, etc. To that extent, it is possible with this combination to represent most of the steer-by-wire functions.

Thanks to the complete mechanical decoupling of steering wheel and steering gear, these functions will doubtless be of even higher quality in the long term. Fully automatic lane keeping and fully automated evasive maneuvers without the participation of the driver in conjunction with all other vehicle systems in the braking and driving fields can be achieved. In the final analysis, autonomous driving is indeed possible.

With the aid of single-wheel steering (each front wheel is individually steered by an electrically operated actuator, and the rigid connection using a tie rod is dispensed with), the wheel angle can be designed, using only the control algorithms filed in the software of the control unit, so individually that today's mechanical multi-link axles could be replaced by simple and inexpensive wheel suspensions.

But until this technology is introduced, the latest statutory regulations need to be changed and the cost/benefit ratio must evolve toward an acceptable and profitable range.

The opportunities for introduction in vehicle concepts newly designed from scratch, such as electric and hybrid vehicles in which electric motors are used directly as the wheel drives, are certainly greater than in classic vehicles with combustion engines.

6 Rear-Axle Steering Systems

The use of rear-axle steering could avoid many of the compromises resulting from the design of passive axles. The resultant adjustment range opens up the potential to do so (see Fig. 27):

For the end customer, this results in considerable improvements in the properties influenced by the chassis. Depending on the driving situation, driving stability, agility, or maneuverability is optimized. Driving pleasure, and the feeling of comfort and safety, is improved, while the properties of driving dynamics can be experienced with greater awareness.

6.1 Basic Functions and Customer Benefits

The following basic functions and objectives for the use of rear-axle steering can be distinguished (see Fig. 28):

- Turning circle reduction/parking assistance: improved maneuvering and parking
- Agility function at low and medium speeds: more driving pleasure, improved handling, sportier driving characteristics, less steering effort, and so greater comfort
- Stability function at high speeds: considerably increased driving stability and safety and improved subjective feeling of safety

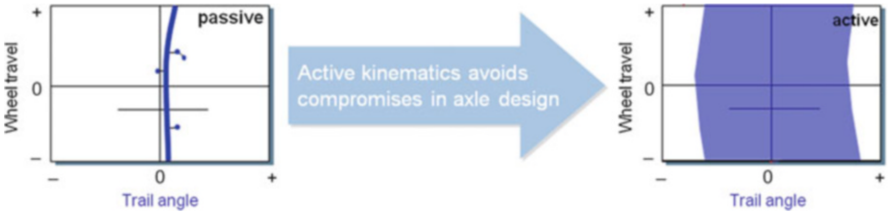
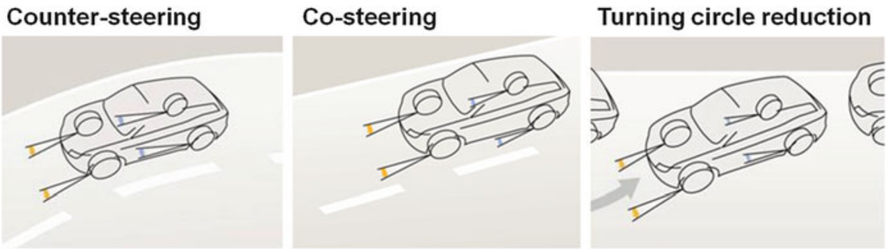


Fig. 27 Cf. passive and active kinematics



Counter-steering

Improved agility

Greater steering and yaw reaction

Co-steering

Higher driving stability

Improved transmission behaviour Lateral acceleration, increased yaw damping

Turning circle reduction

Improved parking

Shifted center of rotation

Fig. 28 Functional advantages of rear-axle steering

6.2 Function Principle

Rear-axle steering offers in principle two intervention options (see also Fig. 28):

6.2.1 Counter-Steering

Counter-steering reduces the turning circle, and the result is a reduction in the necessary steering wheel angle. In driving physics terms, this is the result of "virtual wheelbase shortening." From the viewpoint of driving dynamics, counter-steering initially increases the effective yaw moment. The driver feels an improved maneuverability and agility.

6.2.2 Co-steering

In the case of co-steering, a definite improvement in driving stability can be experienced. The reason for this is the synchronous buildup of the cornering forces at both axles, so that the time until a stationary lateral dynamic state is achieved is shortened. Furthermore, the yaw moment is reduced and limited in its dynamics (less overshooting), which directly improves driving safety. From the driving physics viewpoint, this is a "virtual wheelbase extension."

6.2.3 Driving Dynamics Borderline Area

The maximum potential when the rear wheels are steered can be used in the borderline area in the event of understeering. The rear axle has not yet reached its gripping limit here and can generate additional cornering forces. During oversteering, by contrast, intervention is not reasonable, since the rear axle is already at its grip limit and there is no potential for an increase in cornering forces. Below the physical borderline area, both driving-dynamics situations can be equally well corrected.

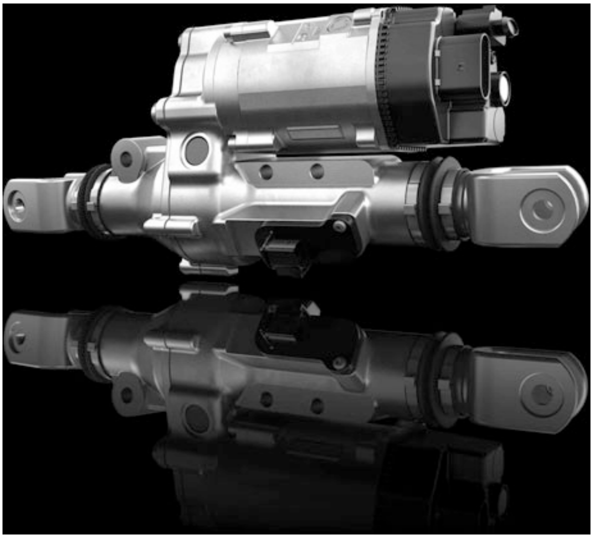


Fig. 29 Central actuator system (ZF Friedrichshafen AG)

6.3 System Design/Structure of System

The systems on the market can be classified into two basic types:

6.3.1 Central Actuator Systems

Structure similar to that for front-axle steering with a centrally arranged actuator (see Fig. 29). The rear wheels are here "mechanically coupled."

6.3.2 Dual Actuator Systems

Structure with two wheel actuators installed in the axles instead of tie rods/links. There is no "mechanical coupling" of the rear wheels here.

6.3.3 Subsystems

- Mechanical:
  Mechanical housing assembly, transmission stage (e.g., toothed belt), transmission gear (e.g., ball screw or trapezoid screw), etc.
- Mechatronic:
  electric motor, sensors, cable harnesses/plug connections
- Electric/Electronic:
  control unit with power electronics
- Software:
  operating software (low level), steering function (high level)

6.4 Interlinking/Expanded Functionality

Due to the increasing number of active driving-dynamics systems, intelligently interlinking them is becoming a necessity. This results in further functional potential. This is illustrated in the following by way of example:

Electric Power Steering The parking assistance systems already available today can be improved with the aid of the increased maneuverability due to rear-axle steering systems.

Active Steering By a functional interlinking of active front-axle steering (variable ratio) and rear-axle steering, the overall steering behavior of the vehicles can be variably determined by both axles.

Electronic Stability Systems (Brake) The function of stability control systems can also be expanded by rear-axle steering. Steering interventions can be made well before the borderline area (and hence also before engagement of the brake) and barely perceptible for the driver. This is referred to as the so-called soft stabilization.

References

Brenner P (2003) Die elektrischen Komponenten der Aktivlenkung von ZF Lenksysteme GmbH, Tagung PKW-Lenksysteme – Vorbereitung auf die Technik von morgen. Haus der Technik e.V., Essen

Eckrich M, Pischinger M, Krenn M, Bartz R, Munnix P (2002) Aktivlenkung – Anforderungen an Sicherheitstechnik und Entwicklungsprozess. Tagungsband Aachener Kolloquium Fahrzeug- und Motorentechnik 2002:1169–1183

Reinelt W, Klier W, Reimann G, Schuster W, Großheim R (2004) Active front steering for passenger cars: safety and functionality. In: SAE World Congress, steering & suspension technology symposium, Detroit, Mar 2004

Schöpfel A, Stingl H, Schwarz Ralf, Dick W, Biesalski A (2007) Audi drive select. ATZ und MTZ Sonderausgabe – Der neue Audi A4, Vieweg Verlag, Sept 2007

VDI/GMA Fachtagung (2004) “Steuerung und Regelung von Fahrzeugen und Motoren AUTOREG 2004”. 2 and 3 Mar 2004, Wiesloch, Germany. VDI Report No. 1828, pp 569–584

Part VII

Human-Machine-Interface for DAS

Guidelines for User-Centered Development of DAS

32

Winfried König

Contents

Winfried König is retired.

W. König (✉)
Offenburg, Germany
e-mail: winfried.koenig@gmail.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_33

Abstract

Above all, a driver assistance system (DAS) should be transparent to the driver and perform predictably and in accordance with the driver's expectations. A DAS should also be simple to use and learn and have limits which are clear and well communicated to the driver. Other requirements of a DAS include comfort, safety of use, and acceptance by drivers and the wider community. The development of a DAS requires the cooperation of experts from engineering, science, and humanities. During the development process, research is very important and effective measurement methods must be developed and applied, by a team with extensive knowledge and experience. The issue of driver responsibility is of crucial importance as DAS plays an even greater role in the control of the vehicles. There are many HMI factors associated with a DAS. Although these are well known, they are not yet fully described in standards and requirements documents. Also standardized measurement methods are not well developed.

1 Introduction

During many years of research by car manufacturers, suppliers, and research institutes, broad but still incomplete knowledge has been gained about the interaction between driver information systems (DISs), driver assistance systems (DASs), and their users. Through German and international projects such as PROMETHEUS, DRIVE, MOTIV, INVENT, RESPONSE, and ACTIV, car manufacturers, suppliers, and governmental and private research institutes have come together to advance precompetitive research on such systems. This chapter presents some of the acquired knowledge to facilitate the development of a DAS. The first section explains the interaction between human, vehicle and environment, and discusses the most significant issues. The second chapter will present some problems which occur in different forms and intensity in all DASs and which therefore can be considered together. The third chapter will consider an approved method for the development of a DAS and the integration of HMI factors. Finally, in the last chapter, the evaluation of the HMI of an existing and planned DAS will be examined in more detail.

2 Relevant Factors

The driver, the DAS-equipped vehicle, and the environment of the vehicle act closely together in time and space (Fig. 1). Therefore the design of these systems from a purely technical point of view is not sufficient, and the habits, the abilities and the deficiencies of drivers, as well as other factors must be taken into account. Only in this way can an improvement be made in the safety, comfort, and commercial demand of such systems.

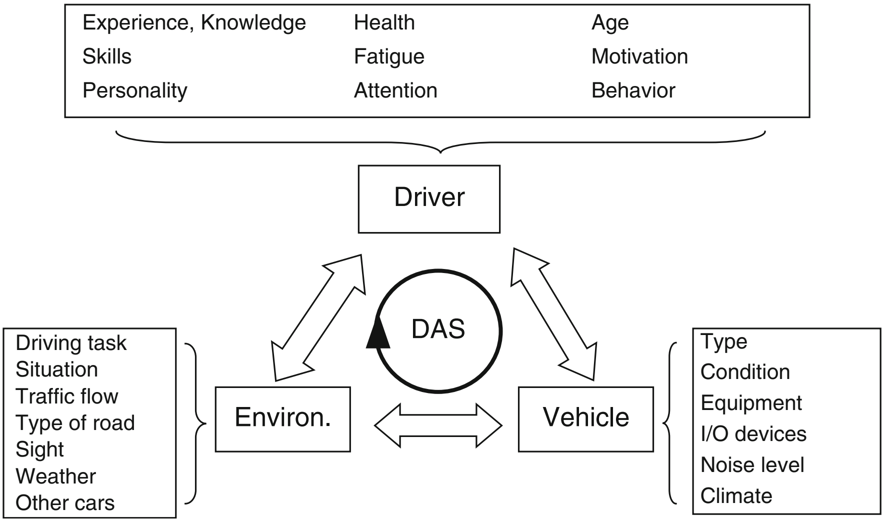


Fig. 1 Cooperation of the driver, vehicle with DAS, and environment

2.1 Support by DAS

Driver assistance systems can assist throughout the driving process: stabilization, lane keeping, navigation, and secondary activities. They can assist in or take over different subtasks from the driver. Their contribution can reach from simply providing information, through the analysis and evaluation of a situation, to the selection and execution of an action. At the same time, compliance with this basic demand of the "Vienna Convention on Road Traffic" must be ensured: "Every driver shall at all times be able to control his vehicle or to guide his animals" (UNEC 1968). To investigate the demands and the possibilities of supporting the driver, it is necessary to develop a thorough understanding of the behavior of drivers in road traffic in different driving situations. This includes not only the extreme case of an accident but also "normal" driving, when drivers sometimes act outside traffic regulations. Additionally, drivers perform extended secondary tasks but must nevertheless successfully master difficult traffic situations. The sequence and the actions of drivers during accidents in Germany are captured in the GIDAS database (German in-depth accident study) (GIDAS 1999) in which the data sets of more than 20,000 accidents are stored [as at July 2012]. Less knowledge exists about "normal driving," although the first projects concerning the collection of such data have been finished in the USA and Europe (Neale et al. 2005; Lietz 2010).

2.2 Performance and Limits of DAS

While designing a DAS, the relevant parameters of the drivers, the vehicle, and the environment must be identified, quantified, and described. It must be clearly defined

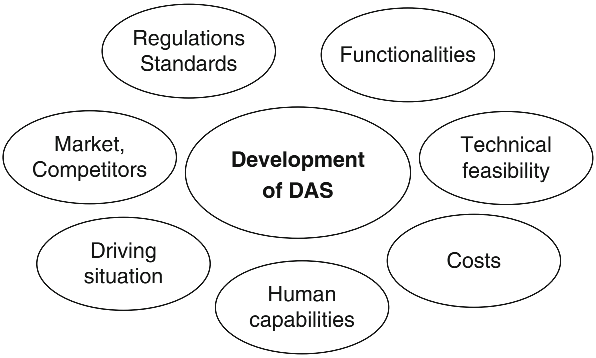


Fig. 2 Factors with influence on the development of DAS

which features the DAS is to be able to deliver, in which situations, and within which limits. The knowledge and assimilation of these limits is an essential part of the process in which the driver "learns" the DAS.

2.3 Skills and Facilities Needed

For the development of an HMI of a DAS, the knowledge and skills of engineers must be applied. Additionally, a thorough understanding of physiology and traffic psychology is necessary in order to take into account the demands and the behavior of drivers. A proven development strategy is to use an interdisciplinary team (human engineering team), the permanent members of which must at the very least include engineers and psychologists. Further specialist groups should be included as appropriate on a case-by-case basis.

2.4 Influencing Factors

In addition to the different functions of the DAS, which must be described systematically (Fig. 2) and in detail, further influencing factors have to be considered: for example, a specific function may have to be designed differently, depending on whether it is to be used by the driver only while stationary or also while driving. Other factors are the risk of driver distraction and the requirement that any dialog between the driver and DAS must be able to be interrupted at any time. Also the broad spectrum of abilities of different user groups is significant. Examples are possible physiological and cognitive deficiencies in older drivers and an increased risk propensity and less developed anticipation of danger in younger drivers. National and international regulations, guidelines, and standards must also be taken into account. They may for example describe minimum standards of usability. Also a minimum level of standardization is necessary so that drivers can use a minimum set of functions without excessive learning. This standardization must be weighed against the drive for innovation that a competitive marketplace demands.

2.5 Channels for the Interaction Between Driver, DAS and Vehicle

Human beings mainly use sight to recognize their environment, i.e., a visual system (Fig. 3). Other road users, their position, their estimated behavior, their track and lane, and objects in the road space are all detected and selected by the visual system and the extremely powerful image processing of the brain. Images are evaluated by different regions of the brain to determine their relevance, both immediate and predicted. The road infrastructure is designed for a visual system. Traffic signs convey regulations, and road markings separate lanes. Flashing indicator lights warn of changes of vehicle direction, and rear brake lights signal deceleration. Thus the visual channel of a DAS is also of great importance. A DAS collects information from cameras and image processing and from other optical sensors in the visible spectrum as well as in the near- and far-infrared and ultraviolet ranges. The acoustic channel is mainly used by humans and the DAS for communication with other participants and especially for warning of danger. The acoustic channel includes the input of commands using speech input systems as well as the output of warning and information by the DAS to the driver via sound signals, noise, and speech output. The haptic channel is used for the input of commands via the hand and foot. Conversely, the DAS uses this channel for force feedback from pedals, the steering wheel, seats, and other special haptic devices.

2.6 Changes in the Interaction Between Driver and Vehicle by DAS

When a driver uses a DAS which directly intervenes in the driving process (like ACC which assists in longitudinal control or a stop-and-go function), there is a

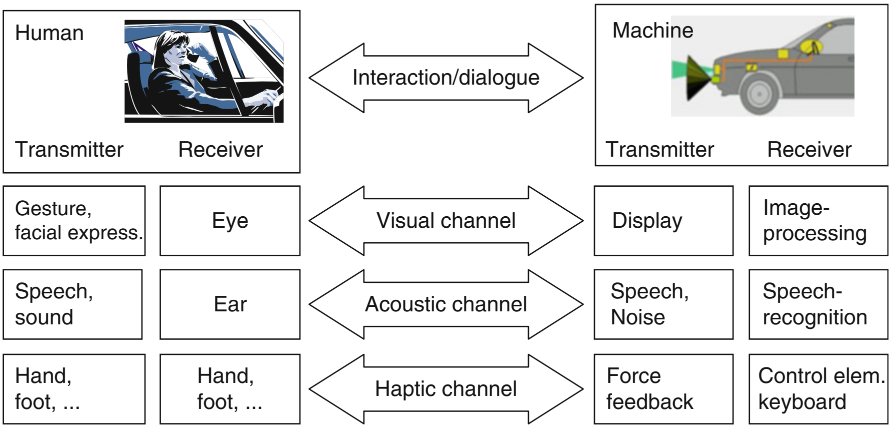


Fig. 3 Channels for the interaction between the driver and DAS

fundamental change in the task of driving the vehicle. Parts of the former driving task may be delegated to the DAS, as a means of reducing demand on the driver and potentially improving traffic safety. The modified task will then contain less controlling and more supervising parts. It may, however, become difficult for the driver to respond well in difficult situations in which the DAS may reach its functional limits and the driver must take full control again. The risk is most apparent when the driver has not been part of the control loop for an extended period and so loses his skills in one or more functions. A deterioration in situational awareness may also occur if the driver does not need to permanently track the details of the driving situation which are important for a given function. The use of a DAS may result in an overall driving behavior which might differ from the actual behavior of the driver. Driving a DAS-equipped car may sometimes feel like being or having a co-driver depending on the level of automation. The quality of this cooperation between the driver and DAS defines the acceptance of the system to a large degree.

2.7 Situational Awareness and the Driver's Intent

For the evaluation of the driving situation, the system uses sensors whose range of detection is normally not the same as that of the human senses. The limits of the sensors and of the associated signal processing are essential for the functionality of the DAS. However, if these limits are not apparent to the driver, it will be difficult for him to use the system as intended by the manufacturer. The intentions of other road users are important in developing a suitable strategy for a driver's own behavior in a given traffic situation. This includes the assumption that the other road users will normally act in accordance with traffic regulations and norms. Experienced drivers however are more able to anticipate unusual, incorrect, or illegal actions by other road users before a dangerous situation has developed. This ability may be called “situational awareness” and is perhaps most important in a driving situation which includes secondary tasks such as operating a sound system. Drivers with good situation awareness only divert themselves to secondary tasks if their estimation of the traffic situation allows it. They will check the development of the driving situation with short glances and interrupt a secondary activity if the difficulty of the situation demands it. In the process of evaluating a driving situation, it is often difficult to identify the most important information and events (“cues”). It was found that while engaged in a secondary task, a driver constantly observes the development of those cues assessed as the most important. Other factors are often filtered out. Situational awareness can be developed only to a limited degree through technical means and is therefore not normally included in the planning process for a DAS system action. A DAS which acts automatically, such as initiating emergency braking, needs its own situational awareness, as otherwise the DAS' actions might conflict with those of the driver. This situational awareness includes an awareness of the intentions of the driver himself, as well as the environment of the vehicle and actions of other road users. It could happen that the driver intends to reinforce or interrupt an action which the DAS has initiated: for example, emergency braking might be initiated, although the driver was planning to change lane to avoid an obstacle rather than brake (see ▶ Chap. 3, “Framework Conditions for the Development of Driver Assistance Systems”).

2.8 Mental Model

With the increasing functionality of the DAS and the resulting decrease in driving demands, the complexity of the system increases. There is an associated increase in the risk that the system will not be understood by a driver. For example, it might be that a driver has understood the functions of a speed control system or can at least use the system without problem. But in more advanced systems such as an adaptive cruise control (ACC) system, the driver may need to learn how to use the system and to appreciate its functional limits. This problem is likely to increase as ACC systems are further developed with stop-and-go functionality and with lateral control support. It is important to ensure that a driver can build a mental model of a system, whether through product information or other means such as a “demonstration mode.” Such a model need not be a physically accurate representation of the system functionality. It might instead consist of images and metaphors from the experience of the user. It is crucial though that the model contain important warnings, as well as messages and functional limits. A driver will need particular support in the use of rarely used functions and in understanding infrequent messages and warnings. This additional support will allow a driver to learn and integrate little used but potentially important DAS functions and messages into the mental model. System behavior in dangerous situations is more difficult to learn and can often only be learned through the use of simulators.

2.9 Relief or Additional Load by DIS and DAS?

In the process of designing a human-machine system, a basic rule is to avoid overloading or “under-challenging” the human component. It should be borne in mind that any interaction with DIS and DAS will absorb a certain amount of a driver’s mental capacity. This is an additional load which must be offset by the load reduction offered by the system. In several projects (e.g., SANTOS (König et al. 2003), COMUNICAR (1999)), an aim was to design human-machine interaction so that the total mental load on the driver, caused by the driving process and possible secondary tasks, did not exceed a safe level. For that purpose, estimates of the mental load caused by the complexity of the traffic situation and by secondary tasks like talking with a passenger were combined with an estimate of the actual capability of a driver. The adaptation of a DAS to the individual capability and preferences of a specific driver (personalization) has also been a theme in several projects. If the reduction in load on a driver is too great, then there may be an increased risk of poor attention or sleepiness. As part of a DAS development process, research must be carried out to investigate such problems as well as other effects such as the possibility of the driver adopting a more dangerous style of driving.

2.10 Responsibility of the Driver

In a current debate between DAS experts, it is widely accepted that a driver must maintain responsibility for vehicle control irrespective of the use of a DAS. This is already stipulated in the Vienna Convention (UNEC 1968). The implications of this requirement for the design of a DAS are a subject of debate among DAS professionals. Some experts consider that systems which cannot be overridden by the driver are generally not acceptable. This includes the emergency braking system (EBS) as well as speed limiting systems. Others think that the Vienna Convention is less proscriptive and that systems such as the EBS do comply with the Vienna Convention standards if appropriately designed.

The question of responsibility also exists with the so-called cooperative DAS, a system which is influenced by data derived from other road users and the road infrastructure. A relatively simple example is a speed limiting system. The questions that arise from the use of such a system are whether the external influences are authorized, whether the data on which this interference is based are reliable, and of course who has the responsibility for the changed behavior of the vehicle. Changing the Vienna Convention would require considerable effort due to its global validity, and any changes that result in even a partial transfer of responsibility from driver to manufacturer or supplier would be controversial. As a result of this background, a DAS should be designed so that its actions can be overruled at any time by the driver. This requires a design that shows the driver the actual state of the system, so that he or she can build up and update a good mental model of the system behavior.

2.11 Strength of Human and Machine

An opinion held by some DAS experts is that a DAS should be primarily used for those tasks for which a human being is less suitable because of natural limits in human capability in some areas. Such tasks include:

- Routine tasks
- “Simple” but time-critical tasks
- Vision at night and in bad weather conditions
- Estimation of distance and speed differences
- Maintaining a safe and appropriate distance from other road users

This subset of tasks however results in the significant problem that a DAS will be able to handle even more tasks in increasingly difficult situations and therefore the driver will have to interfere less often, possibly resulting in skills fade. Despite this he or she may still have to take control of the vehicle again in the most difficult of situations.

3 Systematic Development of the HMI of DAS

3.1 The Development of the HMI in the DAS Development Process

To ensure that the abilities and limits of drivers are considered in every phase of DAS HMI development, HMI experts must be employed in addition to technical experts. From the very start of the development process, the creation of concepts and the demands of HMI users are crucial factors. The next step in the design process is to develop a precise and structured description of the features of the system and the conditions in which these can be delivered. The RESPONSE questionnaires or other similar documents are a good basis to find out and document the features of a specific system. Several questions concern the conditions in which a system can be successfully used and those which cannot be handled by the same system. An example of this is a system that can only be used on major roads but not in urban areas where traffic is much more chaotic.

The next step in testing is to run driving simulator tests with representative user groups in the safe environment of a laboratory. Often at this early stage, a realistic HMI of the DAS is not yet available, but rather a limited computer simulation or a virtual prototype. With the increasing maturity of a system and a growing knowledge of its effects on the user, live driving tests will become possible, first on a test track and later in real traffic. These tests would initially be run with DAS experts as drivers for reasons of safety and economy, and then later with selected user groups. As soon as the system is introduced to the market, additional experience is generated which is collected and evaluated by HMI experts. All of these process steps contain iterations if and when modifications and improvements to the system are necessary.

3.2 Need of Support of the Driver

The inspiration for DAS features which are sensible and likely to achieve market success can be derived systematically from different sources.

These include:

- Explicit wishes of customers, as collected and evaluated by manufacturers' sales divisions.
- The evaluation of accident data. Sources such as the German GIDAS database (GIDAS 1999) can deliver relevant ideas and specifications.
- Field observations of traffic situations.
- Interviews with user groups.

To reduce the possible variations of user groups and situations, it has proved beneficial to define specific user groups and situations. One type of user might be a “mother with child” and the accompanying situation “driving into an underground garage in an SUV.” The evaluation of a sequence of situations described more broadly such as “family holiday trip to Spanish hotel” may give hints on as yet unidentified DAS applications.

3.3 Guidelines for the Development of DIS and DAS

In the European project RESPONSE (RESPONSE 3, 2009), the responsibility of car manufacturers, suppliers, users, and legislators was investigated by a group of experts drawn from manufacturers, suppliers, authorities, research institutes and legal professionals. The results led to guidelines which have since been either applied by many manufacturers or included as additions to or modifications of existing company proprietary procedures. The ability to control a DAS and override its actions is a critical issue.

In the RESPONSE project systems, the DAS was divided into the following categories:

- Information and warning systems
- Active systems that the driver can override at any time
- Active systems that the driver cannot override due to their design or his or her psychomotoric limits.

The focus in RESPONSE was mainly on active systems (advanced driver assistance system, ADAS), which are characterized by an intensive and safety critical interaction between the driver, system, and environment. During the ADAS development process, it is necessary to consider not only system failure during the specification, production, and integration phases but also failure resulting from mistakes during system use or misuse by the driver. In the RESPONSE project, it was recognized that a DAS can only be handled from the point of view of the legislator and the user if it can be controlled and overruled by the user at any time. The allocation of responsibility has to be investigated and defined precisely in individual cases.

The manufacturer must consider how users can use or misuse the system and must establish whether users will be able to perceive and understand system warnings and limits. This can be done by tests with user groups. The manufacturer must also consider possible malfunctions and the implications of these for vehicle and driver (e.g., an accident). Such investigations must be documented.

It is difficult to make an assessment of the risks that could result from incorrect use or misuse of a DAS by the user. To make such an assessment, the user’s expectations of the system must be known, as well the possibilities for system misuse. For example, consider what happens when a driver counteracts the steering torque of the DAS in an attempt to avoid an obstacle. Perhaps this counteraction is a skilled reaction to an emergency situation in which the driver considers that his own action is better than the current behavior of the DAS. Alternatively, the driver may just be reacting through shock, a “knee-jerk” reaction. An example of foreseeable misuse is that of a driver using a lane keeping system to allow him to perform secondary tasks to an unacceptable degree. When using only a DAS information and warning system, a driver retains complete responsibility for vehicle control. It is however possible that information or warnings delivered by the system are incorrect or imprecise, resulting in driver error. In this case, the responsibility of the manufacturer or information supplier should be considered. In the RESPONSE project, a detailed checklist for the specification of DAS was developed (Checklist A). It contains questions on the task which the DAS should solve and questions about the user group, the vehicle type, and the market for which the DAS should be designed. Additionally, precise questions are given to address the sensors, the driving situation, possible risks during usage, and the planned usage information and documentation, as well as secondary issues such as maintenance and repair. A second questionnaire (Checklist B) deals with the effects of the DAS on the driver and on road traffic.

3.4 Guidelines for Driver Information System (DIS)

The increasing number of vehicles equipped with a DIS and telematic systems has raised the question in the EC of whether there is a need for design guidelines for a DIS. An EC expert commission developed and published the “European Statement of Principles on Human-Machine Interface for In-Vehicle Information and Communication Systems” (EsoP) on December 22, 2006 (European Union 2006). These guidelines are relevant for all partners of the DIS supply chain, including the producer of the hardware or software, the data supplier, the car manufacturer, and the consumer himself. In after-sales products, the importers and dealers with their individual responsibility were included. The EC intended that the 2006 Statement of Principles be a voluntary agreement in the different states. The guidelines first were limited to the DIS, but many of the principles can be equally well applied to the DAS. The general goal is that the driver should not be distracted, strained, or disturbed. The guidelines were designed not to hinder future development and are therefore formulated without reference to specific technologies. The basic rules and recommendations in each case are described by an explanation, reinforced with positive and negative examples. The general development goals are presented in the guidelines as follows:

- The system should be designed to support the driver and should not give rise to potentially hazardous behavior by the driver or other road users.
- The system should be designed in such a way that the allocation of driver attention to the system displays or controls remains compatible with the attentional demand of the driving situation.
- The system should be designed so as not to distract or visually entertain the driver.

Other guidelines in the 2006 Statement of Principles require a safe installation in which all optical displays can be seen clearly and do not obstruct vehicle controls. There are also suggestions given for interaction with displays and input devices, the system behavior, the system user information (e.g., user manual), and the safe use of the system. The Statement of Principles even includes guidance for sales organizations, car rentals, employers of professional drivers, and the driver himself.

3.5 Standards for the Design of DIS and DAS

The EsoP guidelines contain requirements and solutions, but not detailed numerical values or measurement procedures. The guidelines therefore refer to existing standards or those in development, which in turn refer to single DAS or general concepts like the design of displays, warnings and dialogs. Standards normally define minimum demands. In order to offer a superior product, each manufacturer will try to meet and exceed the requirements of a given standard. Standards are not laws per se but are strong guidelines, which are often obligatory. In court, standards are referred as state of the art. Standards should not obstruct technical progress but they should specify the performance required of a given system (performance standard). Additionally, standards should not preclude a brand-specific design, unless such a design would make a change of a car hazardous for the driver and therefore other road users.

3.6 Development of Standards

International standards are developed by the International Organization for Standardization (ISO). National standards are set by organizations such as DIN in Germany, the SAE in the USA, and JAMA in Japan. Due to the international nature of the automotive market, it is generally desirable that national standards be converted to ISO standards.

3.7 ISO Standards for Human-Machine Interfaces (HMIs) in Vehicles

The ISO working group TC22/SC13/WG8 develops standards which are important for the interaction between a driver and a DIS. These standards address such issues as the design of the dialog between driver and system, the design of auditive information, and the design of interaction devices and visual information. The TC22/SC13/WG8 standards are not only concerned with DIS but may be applied to many different systems within a vehicle. As an example, the ISO standard 15008 (ISO 2013a) contains requirements concerning the presentation of information on optical displays. This includes for example the visual range and the illumination conditions in which a driver must be able to read a display. To define this, the standards include such detail as a minimum acceptable contrast for the display and a minimum size for any alphanumeric characters. Also addressed are glare and specular reflection. Measurement procedures are also defined where appropriate. Other new and planned standards documents concern the management of dialog between driver and systems (ISO 2013b), the design of acoustical signals in the vehicle (ISO 2011), and the measurement of the visual behavior of the driver (ISO 2002).

4 Evaluation of DAS Design

4.1 Evaluation Methods

During the different phases of DAS design, compliance with principles and standards must be verified. As a system progresses through the development process and the implementation of the HMI advances, different evaluation methods can be applied (Fig. 4), from the earliest stages of development, when researching the demand for new DIS features. The creation of an evaluation system, which gives a clear understanding of the features and limits of a DAS, is difficult. Even when this is done well, the comments of test users can only be considered as hints, especially if the use of the system should be intuitive. As an example, even true HMI experts find it impossible to fully evaluate an ACC system without having had real "driving experience" with the system.

4.2 Quantifying Driving Behavior During Evaluation

As soon as a simulation or a prototype of a DAS is available, the use of the system and its possible effects on a driver's behavior can be evaluated. This includes the measurement and evaluation of significant driving dynamics parameters, for example, longitudinal or lateral movements and forces. Measurements of this type can be easily gained in a driving simulator, but other measurements can be more difficult to obtain. Quantifying the lateral position of a vehicle in a lane during actual driving is an example case. Tracking eye and other movements by the driver is also particularly challenging. Eye tracking is of great interest because changes in the driver's normal "scanning" of the interior and unusually long glances at displays can indicate visual overload. Obviously, it is important to consider if any overload is the result of interaction with the DIS. Physiological parameters give clues about mental or physical loads. Questionnaires and interviews can provide more subjective information about attitudes and experiences.

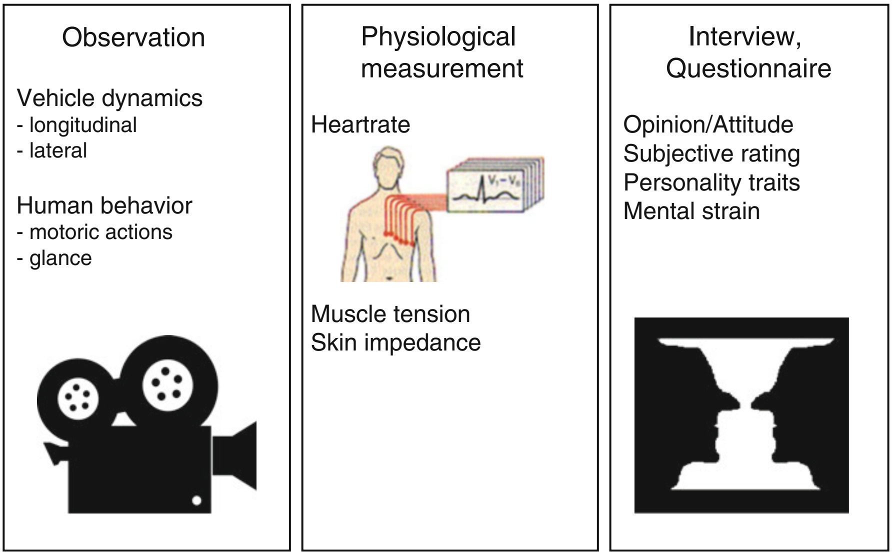


Fig. 4 Instruments for the observation of the behavior of the driver

4.3 Test Environment

DAS usage tests cannot be performed exclusively in a laboratory, since even the most experienced HMI professional cannot avoid introducing some artificial simplification and abstraction. Whereas it may be possible to effectively test the input dialog of a navigation system on a desktop computer, a similar test of ACC modeling would be missing the essential information and feedback from driving dynamics. If the aim of a test is to evaluate the influence of a DIS on the primary driving task, a suitable driving simulator has to be used (see ▶ Chap. 9, “Dynamic Driving Simulators”). The simulator requirements derive from the object of the evaluation. For example, when evaluating a visual DIS, the quality of the visual presentation is likely to be of prime importance, while for testing a DAS which influences the longitudinal and lateral stabilization of the vehicle, the critical factor will be the quality of the dynamic simulation. To establish the effectiveness of a simulator test, the simulator must be “calibrated” against a field test. Simulators do have important advantages, including safety and the ability to accurately repeat test scenarios, but real driving field tests are essential, first on a test track and later in the complex environment of real traffic. During such field testing, on track or road, the safety of the driver and others must be ensured, and this is of course especially important in live traffic conditions. On the public road, a safe test can be assured by using a dual control vehicle, in which an accompanying driver can intervene if necessary.

To reduce the effort involved in creating a realistic driving environment on a test track, good results can be achieved by creating a computer model for use with augmented reality equipment (e.g., semitransparent display goggles). The test driver may then be presented with a road traffic simulation, while actually driving in a safe environment (see ▶ Chap. 10, "Vehicle in the Loop"). Field tests are costly but an indispensable part of the product development process. Long-time tests are necessary for studying how drivers learn to use a DIS/DAS and the effects that a system has on their driving behavior (see ▶ Chap. 12, "User-Oriented Evaluation of Driver Assistance Systems").

4.4 Application of Evaluation Methods and Error Probability

The application of evaluation methods requires a thorough subject knowledge, developed through both education and extensive experimental work experience. Such a knowledge reaches from the development of a research plan, through the selection of research subjects and the execution of tests and trials, to the evaluation and interpretation of results. Besides normal scientific experimental errors, vehicle dynamic sensors and physiological sensors are possible sources of error, and evaluating drivers' mental processes is particularly error prone. Additionally, the presence of a researcher can influence results, especially if suggestive questions are posed or assistance given. When making physiological measurements, there is often a significant delay between stimulus and reaction. There is also a risk that intrusive or uncomfortable sensors can restrain or even frighten the subject. Furthermore, a car interior is an environment with significant audible and electromagnetic noise, and the resulting interference makes the measurement of extremely low-power physiological signals very difficult. Finally, physiological signals vary greatly between subjects and even for the same subject in different situations. Other sources of errors exist during the design of questionnaires and interviews. Suggestive questions should of course not be used, but there are other less obvious issues. Answers may reflect social norms and a subject's ideals or may be influenced by a subject's perception of what is expected of him or her. Subjects may also feel the need to justify actions or opinions. Researchers should assume that subjects will have gaps in their memories, although this can be tackled to some extent by supporting subjects with video recordings of their tests.

References

COMUNICAR (1999) Communication multimedia unit inside car. http://www.cordis.europa.eu/project/rcn/56893_en.html. Accessed Feb 2015

European Union (2006) Commission recommendation of 22 December 2006 on safe and efficient in-vehicle information and communication systems: update of the European statements of principles on human machine interface. Off J Eur Union 6.2.2007, L32/200. ftp://ftp.cordis.europa.eu/pub/telematics/docs/tap-transport/hmi.pdf. Accessed Feb 2015

GIDAS (1999) German in-depth accident study. http://www.vufo.de/forschung-und-entwicklung/gidas/. Accessed Feb 2015 http://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf. Accessed Feb 2015

ISO (2002) ISO15007, Road vehicles – Ergonomic aspects of transport information and control systems – Measurement of driver visual behaviour with respect to transport information and control systems, ISO/TC22/SC13/WG8, ISO Central Secretariat, Geneva. www.iso.org/iso/catalogue_detail.htm?csnumber=26194. Accessed Feb 2015

ISO (2011) ISO15006, Road vehicles – Ergonomic aspects of transport information and control systems – Ergonomic aspects of in-vehicle auditory presentation for transport information and control systems, Specifications and compliance procedures, ISO/TC22/SC13/WG8, ISO Central Secretariat, Geneva. www.iso.org/iso/catalogue_detail.htm?csnumber=55322. Accessed Feb 2015

ISO (2013a) ISO15008, Road vehicles – Ergonomic aspects of transport information and control systems – Specifications and compliance procedures for in-vehicle visual presentation, ISO/TC22/SC13/WG8, ISO Central Secretariat, Geneva. www.iso.org/iso/catalogue_detail.htm?csnumber=50805. Accessed Feb 2015

ISO (2013b) ISO15005, Road vehicles – Ergonomic aspects of transport information and control systems – Dialogue management principles and compliance procedures, ISO/TC22/SC13/WG8, ISO Central Secretariat, Geneva. www.iso.org/iso/catalogue_detail.htm?csnumber=34085. Accessed Feb 2015

König W, Weiß KE, Mayser C (2003) S.A.N.T.O.S – An advanced concept for integrated driver assistance systems. In: Conference electronic systems for vehicles, VDI-Berichte 1789, Baden Baden, ISBN 3-18-091789-X

Lietz H (2010) Methodische und technische Aspekte einer naturalistic driving study (methodic and technical aspects of a naturalistic driving study), Forschungsvereinigung Automobiltechnik FAT-Schriftenreihe 229, VDA

Neale VJ, Dingus TA, Klauner SG, Sudweek J, Goodman M (2005) An Overview of the 100-car naturalistic driving study and findings. Paper Number 05–0400, national highway traffic safety administration

RESPONSE 3 (2009), Code of Practice for the design and Evaluation of ADAS, V5.0, Aug. 2009, http://www.acea.be/uploads/publications/20090831_Code_of_Practice_ADAS.pdf. Accessed Feb 2015

UNEC (1968) Convention on road traffic, Vienna, 8.11.1968, consolidated version, pp 11–15, UN Secretary-General. http://www.unece.org/fileadmin/DAM/trans/conventn/Conv_road_traffic_EN.pdf. Accessed Feb 2015

Design of Human-Machine-Interfaces for DAS

33

Ralph Bruder and Muriel Didier

Contents

Abstract

Interaction between human and machine occurs via interfaces that provide the driver with information and are meant to assist the driver in the task of driving safely, effectively, and efficiently. The following is an explanation of how displays and controls must be designed in the course of product development and which aspects deserve the most attention from the standpoint of the interaction between human and machine. Firstly, a working model is presented in order to explain human information processing and the action process. This model can be regarded as the foundation for designing an HMI. This is followed by various methods of systematizing displays and controls best suited for

---

R. Bruder (✉)
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: bruder@iad.tu-darmstadt.de

M. Didier
Novartis Pharma AG, Basel, Switzerland
e-mail: muriel.didierpro@googlemail.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_34

approaching the issues involved in driving a vehicle. However, the human being should be at the center of the design process, which is the reason why design guidelines and principles are listed, in order to explain the underpinnings of the approach while focusing on user-oriented implementation.

1 A Working Model of Human–Machine Interfaces

As the foundation for a working model of HMIs, we use the so-called stimulus–organism-response (SOR) model, which is also known as the stimulus–response or input–output model. It is a model of human information processing originally used in the field of psychology that explains how stimulus and response are linked. It is based on the idea that a stimulus, e.g., a buzzer sounding in a car, is processed in the organism and triggers certain reactions in the form of motivational decision-making or learning processes, for example, a physical reaction such as moving a lever or twisting a knob. On so acting, the organism receives a feedback signal, e.g., a tone sound, which confirms the success of the act performed (see also ► Chap. 1, "Capabilities of Humans for Vehicle Guidance").

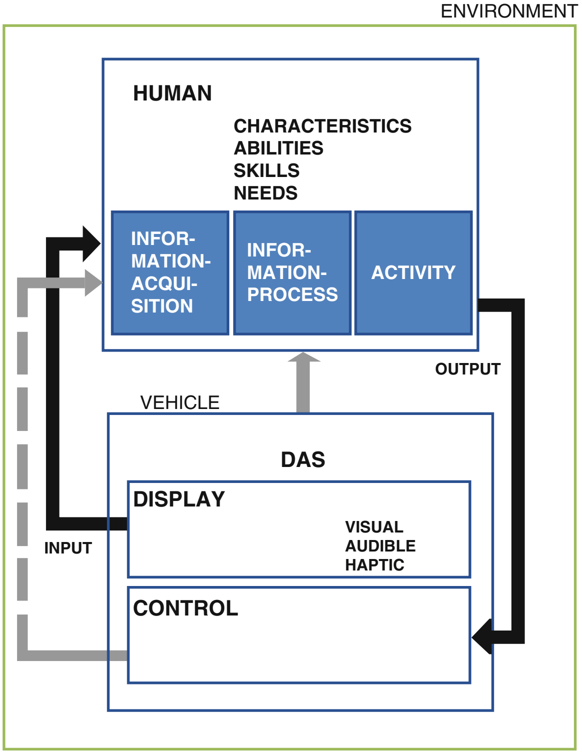
In today's vehicle, information is transmitted by means of visual displays, audible alarms, and signals or through haptic feedback via the steering wheel or the seat. Then, this information is processed, which usually results in motion applying force to the steering wheel, pedals, switches, or levers. Both the human factors and the environmental parameters influence the three stages of this process: information acquisition, information processing, and action. Since the interface performs the function of a "mediator" throughout this entire process, the interface must be designed to allow for all relevant determining factors and magnitudes (see Fig. 1).

Contemporary motor vehicles incorporate progressively more devices meant both to assist the driver in acquiring information and to support his output activity. In the course of the current trend toward multiplication of DAS – especially when they assume some of the driving tasks (e.g., ACC) and thereby modify the way information is processed – during development, special attention must be paid to the human aspects of design. Interfaces should be adapted to humans in order to optimize this information processing.

2 Primary Classification of Interfaces

The first basic classification of interfaces, one that is very commonly used both in research and application, differentiates between displays and controls. A display, meaning any kind of information acquisition, thus constitutes the trigger of human information processing. On the other hand, the controls constitute the operational part, namely, what the driver eventually does, "activates," or "operates" after receiving and processing the information. Thus, these two groups represent two completely different kinds of devices. Accordingly, they should be considered

Fig. 1 Working model of human interaction with interfaces in vehicles



separately. In the following chapters dealing with specific design recommendations, warnings, which belong to the information acquisition process, are dealt with separately (in ▶ Chap. 36, "Driver Warning Elements"), since the factors involved in them are very specific. Based on the subdivision of human–machine interfaces into two main categories – controls and displays – different approaches have been tried to characterize interfaces. We intend to simplify the process of optimization among human needs, human ability to perform as well as interface capacity. Beneath, we present the approaches most readily applicable to the task of driving.

2.1 Control Devices

Rühmann (1993) developed a broad general characterization of controls: He orders the distinguishing features using five different classification systems. An approach to characterizing controls that specifically addresses vehicles was developed by Eckstein (2001), who also lists therein five features that specifically cater to the field of motor vehicle operation.

Rühmann's (1993) method of classifying the wide range of commonly used control devices presents the following five ordering systems:

- *Mode of operation*: Control devices can be classified by the human limb used to operate them or act on them, such as fingers (light switch), hand (gear shift), foot (accelerator), or leg (brake pedal). Further subdivision is possible according to the manner of gripping or treading.
- *Type of movement*: According to the direction in which the controls move, we can distinguish among rotation, linear motion, and quasi-linear motion.
- *Mode of action*: With regard to the mode of action, we distinguish between analogue (continuous) and digital (discrete) controls.
- *Dimensionality*: Dimensionality is the number of degrees of freedom available to the control device.
- *Integration*: If a single control device affects several operating functions, it is called an "integrated control." Thus, a single integrated control can perform various partial or parallel tasks by activating it in a certain sequence or simultaneously.

Eckstein (2001) classifies the concepts of control manipulation in motor vehicle controls according to five characteristics:

- *Number of controls*: For example, three levers for three functions, blinkers, windscreen wipers, and ACC.
- *Degrees of freedom*: In the "steering wheel and automatic transmission" operating concept, there are three degrees of freedom, namely, steering wheel, accelerator, and brake pedal.
- *Set point*: The set points are the angle (steering wheel angle), the path (throttle valve), and the amount of force exerted (brake pressure).
- *Feedback*: The forces and paths acting on the steering wheel constitute feedback information.
- *Type of control* (isomorphic, isotonic, and isometric): For example, the accelerator pedal and the brake pedal are instances of an isomorphic control type.

In addition, it is also quite feasible to make a classification based on which dimension of the motion is being determined, namely, "crosswise" and "longitudinal" motion or one based on the three categories, "steer," "accelerate," and "brake."

If we apply Rühmann's (1993) and Eckstein's (2001) classification methods, it becomes evident that – apart from the mode of operation criterion – controls are the focus of attention. A human-centered perspective on the transmission of commands from the driver to the vehicle results in a classification according to different input modes. When optimizing interface design, it is important to combine both perspectives so as to address both the hardware and the human being.

In motor vehicles, input from the upper and lower limbs predominates. This mode of input can be designated **hand–arm movement** for the upper limbs, including fingers, and **leg–foot movement** for the lower limbs. For example, hands and arms are used in motor vehicles to control sideway motion via the steering wheel, while backward and forward movement is controlled by the lower limbs (leg and foot). However, use of the limbs is not limited to stabilizing functions. All buttons, rotating knobs, and touch screens also require the use of at least parts of the hand–arm system. Accordingly, they are also of great importance for operation of driver assistance systems. Beneath, input modes are distinguished suited for inputting control signals. The last three are typical no-contact input options.

Hand–arm input: In addition to the aforesaid steering wheel, there are other arm- and hand-operated controls for secondary and tertiary driving tasks such as manipulating the direction indicator light or infotainment devices.

Leg and foot movements: The leg–foot system is used for pressing pedals and in some cases for the parking brake.

Weight shifting: Shifting the human body's center of gravity can also be used as an input medium, as occurs indirectly in motorcycles, for example.

Voice input: Inputting commands by uttering keywords is known from the field of human–computer interaction. The transmission of such verbal commands is used in motor vehicles, for example, for mobile communication devices or operating infotainment systems.

Eye movement: Interactive systems of this type have long been available for use with computers. However, such systems are not yet suitable for standard use in vehicles.

Facial expressions and gestures: The use of gestures as an input option, e.g., for shifting gears, is still in the experimental stage. Further development toward recognition of emotions and facial expressions is conceivable, although difficulties persist in transferring these systems in vehicles to different types of riders.

2.2 Display

The primary purpose of a display is communicating information to humans for use as an input in production processes. When designing displays, three main issues arise (Schmidtke 1993):

- *What* information must be conveyed ("information content")?
- *How* should this information be transmitted ("embodiment")?
- *Where* is the information presented ("display location")?

Display classification systems are often based on representations that show different degrees of detail.

By contrast, when describing the ways in which displays report information, Schmidtke (1993) distinguishes among three signal types (optical, acoustic, and haptic) and moreover describes in which form the information may appear in terms of their technical basis (digital or analogue). Furthermore, displays are classified according to other characteristics, such as the shape of the scale (e.g., continuous, in steps, circular) or the display's dimensionality (fixed scale and moving pointer or vice versa).

For their part, Timpe et al. (2000) write of visual, auditory, and haptic interfaces. In addition, for each category, the authors distinguish the types of information it can transmit to humans: For example, sound signals can consist of verbal or nonverbal information. Nonverbal information, in turn, is subdivided by making a distinction between tones and other sounds.

Display classification systems are based mainly on the technical characteristics of the display elements. Systematization according to the senses from a human perspective provides information on human capacity, especially the capacity to absorb information (see ▶ Chap. 1, "Capabilities of Humans for Vehicle Guidance").

A classic scheme for deploying forms of sensory perception in vehicles can be described as follows: The visual channel enables the driver to perceive other drivers and displays in the vehicle (e.g., fuel gauge). Acoustic signals are often used for warnings and more rarely for indicating status (e.g., relay noise from an inversion indicator). Vestibular perception informs the driver of the various forces acting on him. Sensory faculties are used in the case of tactile perception, for example, for pushing a button. Kinesthetic perception is mainly used for larger movements and vibrations such as turning the steering wheel.

How are DASs integrated into this scheme? For the DAS currently being marketed, visual, tactile, and kinesthetic perception is used, as with torque forces acting on the steering wheel of a lane-keeping system or vibration devices when switching lanes within a lane-following system. Even vestibular perception can serve to some extent as a display. This occurs in DAS that perform stabilizing functions in the vehicle, such as adaptive cruise control. Deceleration of the vehicle by cruise control informs the driver that the vehicle ahead has been detected, with no need to take his eyes off the road. Acoustic signals are often used by DAS as warnings, since sounds are not linked to specific factors such as the direction of the driver's gaze.

During operation of a motor vehicle, the optical channel is exposed to many stimuli and should, whenever possible, not be burdened by additional information from driver assistance systems. However, development of DASs that use tactile and kinesthetic perception is still in its cradle and is usually supplemented by additional visual or audible signals.

Comparison of the two approaches, namely, technical characteristics and human qualities, is intended to show which senses are available and which displays are most suited for transmitting information to humans. This optimization process is particularly important whenever several messages must be transmitted simultaneously or information must be processed quickly.

In principle, it should be noted that classification of displays cannot yield any information on the content and display location of messages. These two attributes of a display depend on many factors, according to the complexity of the information processing. These issues form part of the design process and must be reviewed for each task or development and adapted to the specific operations. The next section presents guidelines and principles that support the design process.

3 Guidelines and Principles of Design

3.1 Design Guidelines

An **overriding principle** for developing human–machine interfaces is that the machine and its associated elements such as displays and controls must be suited both to the user and to the task that the user wishes to perform. To comply with this general principle, the system must be designed allowing for **human characteristics or abilities** in terms of their physical, psychological, and social aspects. In ▶ Chap. 1, "Capabilities of Humans for Vehicle Guidance," they are classified into three categories: characteristics, abilities, and skills.

A certain interface corresponds to a specific set of users. Clearly defining the composition of this set allows precise identification of the capabilities that are available when selecting an interface.

Other higher-level guidelines for designing human–machine interfaces are set forth in standards, policies, and guidelines.

ISO standard 9241-110 (2006) describes six ergonomic principles or requirements that must be considered when designing interfaces (displays and controls). The fulfillment of the six basic principles (suitability for the task, self-descriptiveness, controllability, compliance with user expectations, error tolerance, customizability, and encouragement of learning) serves as the basis for creating successful human–machine interfaces, as well as for designing DASs. Basically, these guidelines and principles should be deemed a sine qua non throughout the design process.

3.1.1 Suitability for the Task

An interface is suited for a task to the extent that it supports the user in performing a task safely, effectively, and efficiently. Here, the suitability for the task can be differentiated in function allocation, divide the complexity, grouping, distinctiveness, and the functional relationship of a task.

Function allocation designates a sensible distribution of functions between human and machine determined after considering the requirements of the task and the human user's characteristics, skills, and abilities. Complexity should be minimized. Speed and accuracy of human action are variables that should be considered in this context. In particular, the complexity of the task structure and the nature and scope of the information to be processed by the user must be kept in mind. Grouping of the displays and controls should be such that they can be used easily in combination. Also important are the distinctiveness of the various displays and controls and their arrangement by function, since they must be readily identifiable to assure safe use.

3.1.2 Self-Descriptiveness

An interface is self-explanatory if the user can easily and unambiguously recognize the displays and controls and understand the process. In addition to being understandable, the principle of availability of information is also important.

In accordance with this principle, information on the state of the system is made available to the driver immediately in request, without thereby interfering with or neglecting other activities. The system must confirm to the operator without undue delay that it has accepted the operator's action.

3.1.3 Controllability

An interface is controllable if the user is able to determine how he or she wishes to perform the entire task facing him. In the process, the user must control the system and not vice versa. To illustrate the concept of controllability, three basic principles can be distinguished: redundancy, accessibility, and leeway.

Additional indicators and controls must be provided when such redundancy can improve the security of the overall system, since in certain situations the system's capability and security depend on its ability to provide additional information to the user. Furthermore, information meant for the driver should be easy to retrieve and access. This means that the movements of individual parts and limbs of the driver's body and his body movements should not be perceived as uncomfortable.

3.1.4 Conformity with User Expectations

The user has certain expectations of how the human–machine interface works, resulting from the user's knowledge of prior operation, his training, and experience as well as from generally accepted conventions. To prevent inappropriate use or occurrence of a preprogrammed error, function, movement, and location of the interface should comply with expectations.

Compliance with user expectations can be divided into learned stereotypes, such as turning clockwise; stereotypes from practice, such as braking while driving; and consistency between similar interfaces and similar functions.

3.1.5 Error Tolerance

An interface is robust against errors if the intended result of work performed is achieved despite evident errors in input, entailing minimal correction costs or none at all. Systems should be able to check errors and afford the user tools for coping with such errors, while distinguishing between error detection and treatment time.

In terms of DASs, the following errors occur: lack of information, failure to perceive, misinterpretation, mistaken decisions, and faulty execution. When designing the interface, these errors should naturally be avoided or at most have only negligible effects.

3.1.6 Suitability for Learning

An interface is adaptable when it is flexible enough to adapt to the needs and abilities of individual users. In terms of vehicle or DASs, additional parameters should be considered, such as style of driving or driving situations. In this context, sensible and useful supplements are adaptability, enabling users to make changes to the system, and adaptation, in which the system itself makes changes based on user behavior.

Learning how to use an interface should be simplified and supported with instructions. This means that principally during driving – especially if DASs are being used – the time between first use of the system and mastering it should be kept as short as possible. The selection of the interface has a strong bearing on fulfillment of this goal.

3.2 Principles of Design

Design principles to fulfill these six overarching goals should be considered as early as possible during the design process. Testing the selected solutions under realistic conditions is an important aspect of the development process.

With respect to DAS, the design principles of compatibility, consistency, grouping of displays, controls, and balance between tedium and overload should be followed. On the other hand, when designing DAS interfaces, the vehicle system must be considered as a whole, instead of considering each individual item by itself. Other principles, such as comfort, satisfaction, and the joy of using interfaces – although not easy to apply – should also be taken into account.

Compatibility: The application of this principle when designing human–machine systems supports mainly the factors of information processing, namely, perception, memory, problem-solving skills, and action. When designing, spatial compatibility, movement compatibility, and conceptual compatibility should be distinguished.

Examples: To set a higher value, the control dial must be turned clockwise. Moving a lever forward means "more."

Consistency: Uniform design of interfaces in a vehicle expedites information processing and action, thus hastening the learning process and allowing fewer errors and swifter processing. An action should always generate the same effect. In this respect, design should be consistent across systems.

For example: If operating a DAS requires an image of the host vehicle, the perspectives used should be consistent, for example, "rear view."

Spatial arrangement: Optimal grouping of the controls and displays for the DAS and for basic driving allows greater speed of information processing and reduces the error rate. This should foster unity of content and of function, while the frequency with which controls are manipulated and in what sequence should be considered adequately.

For example, the introduction of new technology such as head display offers new ways to arrange information; however, the cockpit should not be cluttered with devices. For this reason, the way in which information is distributed between the additional display and the existing display must be analyzed anew.

Balance between the tedium of idleness and overload: On the one hand, DASs are supposed to relieve the driver while performing his actual driving task. On the other hand, the driver must activate, set, and operate the system. This is added to the main task of driving. A further and important aspect of DAS is the driver's demand to be able to override the system if need be. When some DASs take over the part of the driving tasks, this greatly influences interface design, in order to achieve the goal of assuring the driver a balance between the tedium of idleness and information overload.

For example, if part of the driving task is performed by a DAS, the driver should nonetheless be kept informed of its actions.

The degree of detail with which this is done is a critical issue: If there is too much information, the expected relief provided by the DAS is lost. On the other hand, if there is too little information, the driver will face difficulty when taking over the functions of the DAS.

To generate comfort, satisfaction, and joy of use: When DASs are introduced, especially if they are voluntarily activated by the driver, the design of their interface should generate feelings such as joy of use, satisfaction, and comfort. Use of DAS is influenced by the features offered, but also by the interface itself.

Example: When a DAS interface is considered satisfactory, it encourages use and consequently indirectly encourages learning and acceptance. Conversely, successful use improves the user's opinion of the interface. An unpleasant sensation on touching a surface can discourage the driver from touching it. As a result, the driver may be reluctant to alter DAS parameters as often as possible or as often as necessary.

Considering the overall system: The principle of considering the overall system when designing individual DAS is reinforced by the current trend toward proliferation of DAS in vehicles. A successful design of a system interface can become a failure if it interferes with other interfaces or DAS. When designing several DAS interfaces simultaneously, priority criteria must be established and complied with.

Example: It is technically possible to design "integrated DAS interfaces" that group many functions in a single display or control. With such integrated approaches, the spatial principle is fully satisfied, but the driver needs a complex mental model in order to manipulate the controls safely and quickly. It must be verified whether the ensuing growth of complexity is still compatible with the driver's capabilities.

4 Design Process

When describing interfaces (displays and controls), a separate analysis is possible. However, this would hinder achieving the goal of ergonomic design, especially when designing controls and indicators, since ergonomic principles govern the

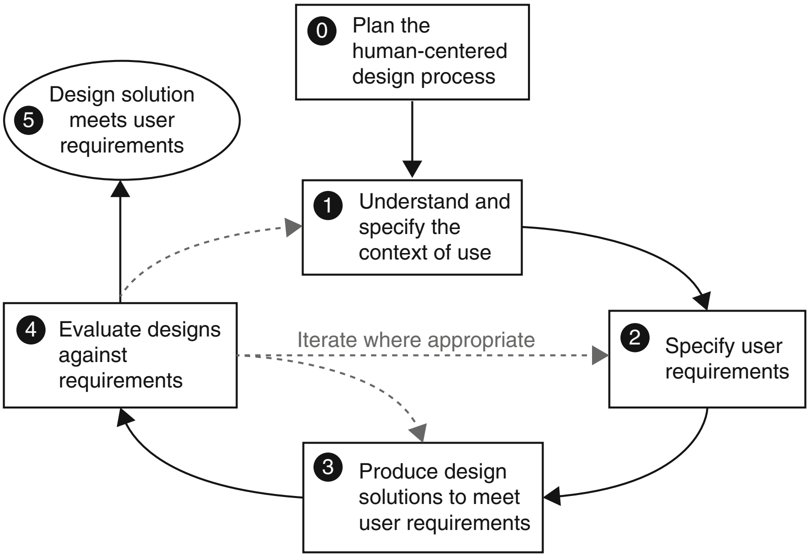


Fig. 2 User-oriented design process (According to ISO 9241-210 (2006))

entire human–machine interface. Successful design for humans can only be assured when the system is considered as a whole.

When developing human–machine interfaces, in order to fulfill the ergonomic requirements, the designer must keep the user in mind at every stage of the design process. For this purpose, EN ISO 9241-210 (2006) provides user-oriented guidance for designing interactive systems, which can be integrated into a multidisciplinary design and development process. This design process comprises four main steps that must be performed iteratively (Fig. 2):

- Understand and specify the context of use
- Specify the user's needs and the specified requirements
- Create design solutions
- Evaluation of solutions according to user-oriented criteria

Regarding the selection of design solutions for controls and indicators, Kirchner and Baum (1986) suggest methods (Figs. 3 and 4) that support the identification of solutions step by step. These steps should be performed as part of an iterative process. The solutions selected are tested under realistic conditions, so as to ensure compliance with the requirements.

Both selection processes are described beneath.

Selecting Controls

- *Task and requirements*: At the beginning of the design process, requirements must be drawn up for the control in terms of the task to be performed. It may be

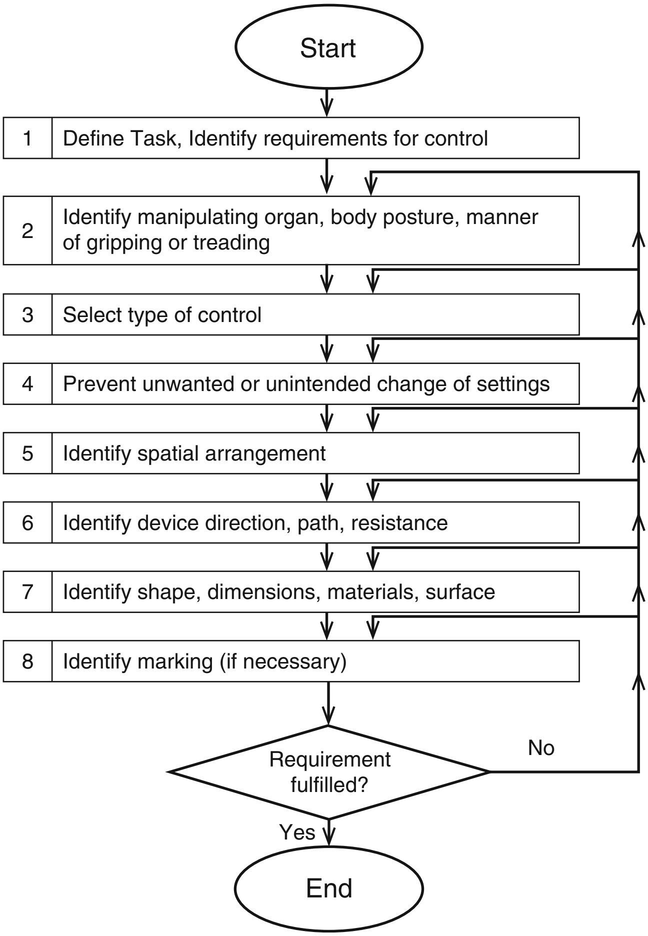


Fig. 3 Method for selecting controls (After Kirchner and Baum 1986)

necessary to assign the requirements, for example, when a contradiction appears or the state of the art does not suffice for complete fulfillment (Fig. 3).

- *Manipulating organ, posture, and manner of gripping or treading*: Review relevance of the different input options. Body posture and any necessary movements, although limited inside the vehicle, should be taken into account, particularly in terms of the frequency and duration of manipulation of use. When choosing the manner of gripping or treading, the amount of force that must be applied plays a key role.
- *Type of control device*: The choice of the type of control device to use depends directly and immediately on the requirements for operating it and especially on the accuracy and speed with which it is activated. With respect to DAS, the relevance of a multifunctional control device is being checked.

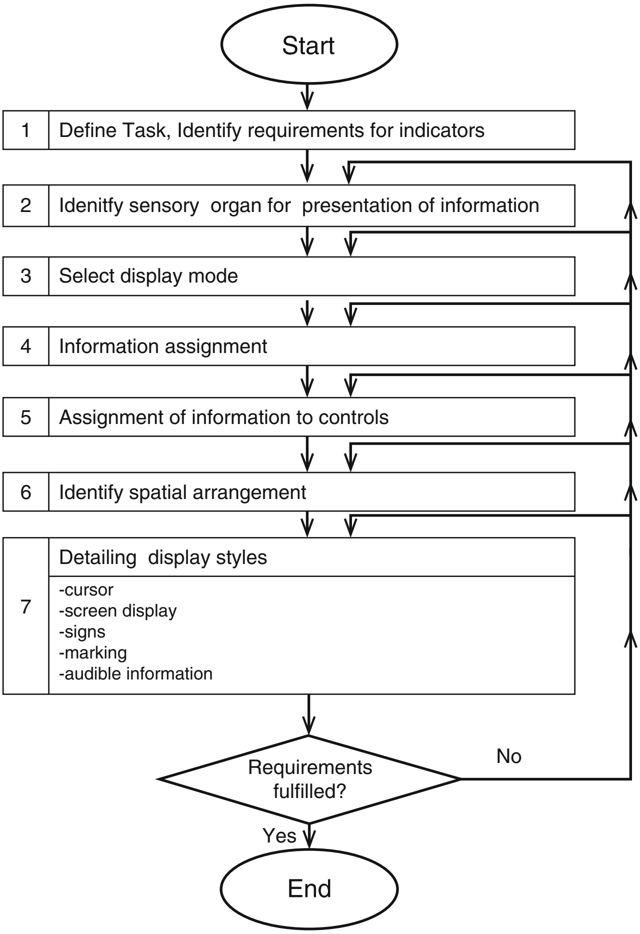


Fig. 4 Procedure for selecting indicators (After Kirchner and Baum 1986)

- *Prevent unwanted change of settings*: Reference to the issue of whether the control device should be protected from inadvertent activation. In this respect, system functions critical to safety must be protected, and the occurrence of undesirable consequences must be avoided.
- *Spatial arrangement*: An item's geometric location in the vehicle and its allocation with respect to other controls should be established, which includes allowing for its functions and the timing of the task. This raises again the issue of selecting a combined control device. The assignment is intended to create a uniform operating model and to take existing stereotypes into account.
- *Device direction, device path, and device resistance*: Determine the technical details of how the control device moves.
- *Shape, dimensions, materials, and surface*: Definition of the "appearance" of a control device.

- *Marking*: To assure visual and/or tactile distinctiveness and operating safety, and to facilitate learning, markings should be selected by determining layout, shape, size, labeling, color, and materials.

Selection of Indicators

- *Information task and requirements*: At the beginning of the design process, the designer must specify the requirements demanded of displays. This information supports the information processing required for performing the task. The purpose of the information (monitoring a variable condition, surveillance of a setting, etc.), the accuracy of the information acquisition (reading, guiding perception, etc.), and the information content (actual value, target value, difference values, etc.) must be defined. Special attention must be paid to the amount of information to be transmitted, which must not overwhelm the user (Fig. 3).
- *The sensory organ*: The choice of the information to be presented should be determined by the sensory organs best suited for the task. Load on the sensory organ (in motor vehicles, the visual channel already bears a heavy perceptive load), the need for speedy response, the need for distinctiveness, and availability of the information and its acceptance are the main constraints that must be considered.
- *Display mode*: The type of information is determined by the sensory organ to which it is addressed. The information task with its content and the attributes of the human user are the basis for selecting display methods (analogue or digital display, a tone or spoken message, etc.).
- *Information assignment*: Connections (functional or physical) among individual items of information must be identified and used as a basis for allocating displays (e.g., set point value and actual value appear close together), while the complexity of information processing should be taken into account.
- *Spatial arrangement*: The display's location and position can be set according to the requirements noted in the prior steps. Distinctiveness and error prevention are important goals. A successful decision is made, when the display is viewed as embedded in the overall system environment, instead of in isolation; the issue of combined displays again arises.
- *Detailing display styles*: Among others, parameters such as contrast, font size, scale, rate of change, color, sound, and signal frequency must be determined. Among other things, the requirements demanded of the display, ergonomic rules, distinctiveness, and the technical possibilities are considered.

5 Practice and the Design Process

With the expansion of an ACC system's functions at low speeds, the issue of the interface arises once more. The new system expands the range of situations in which the system can operate. In other words, the frequency with which such situations occur has changed: Urban traffic at low speed, 90° turns, and right of way are just a few of the applications one could name.

As a result of these changes, the information load on the driver is modified, for example, in city traffic the visual channel is exposed to a higher load than on a highway due to the complexity of the urban driving environment. The amount of information relevant to ACC that must be transmitted is thus significantly increased by the proliferation of ACC status changes.

When transmitting ACC information, several questions arise: Can the visual channel support a larger information load? Are there any places for presenting information other than the area of the speedometer? Is the information presented relevant to the driving situation?

The situation and the surroundings make greater demands on the driver than on a motorway and provide more distractions, so the driver has little time for looking at displays. Consequently, it is advisable to place the display in the middle of the driver's field of vision, namely, on the windscreen, to avoid looking away from the road. The use of other input channels should also be explored further, e.g., vibration in foot area. Information displayed about the driving situation could be steadily updated. Such a change would also reduce the amount of information needed.

This example is intended to show that every time a DAS is changed and every time the vehicle in which the DAS is incorporated is modified (e.g., by installing additional DASs), the interface design processes listed above must be performed once more in order to attain the goals of a successful human–machine interface.

References

Eckstein L (2001) Entwicklung und Überprüfung eines Bedienkonzepts und von Algorithmen zum Fahren eines Kraftfahrzeugs mit aktiven Sidesticks, vol series 12, vol 471. VDI-Verlag, Düsseldorf

ISO 9241-110 (2006) Ergonomics of human-system interaction. Part 110: dialogue principles

ISO 9241-210 (2010) Ergonomics of human-system interaction. Part 210: human-centred design for interactive systems

Kirchner JH, Baum E (1986) Mensch-Maschine-Umwelt (Human-machine-environment). Beuth Verlag, Berlin/Cologne

Rühmann H (1993) Schnittstellen in Mensch-Maschine-Systemen (Interfaces in human-machine-systems). In: Schmidtke H (ed) Ergonomie, 3rd edn. Hanser, Munich/Vienna

Schmidtke H (1993) Ergonomie (ergonomics). Hanser-Verlag, Munich/Vienna

Timpe KP, Jürgensohn T, Kolrep H (2000) Mensch-Maschine- Systemtechnik – Konzepte, Modellierung, Gestaltung, Evaluation (Human-machine-systems engineering – concepts, modeling, design, evaluation). Symposion Publishing, Düsseldorf

Input Devices for DAS

34

Klaus Bengler, Matthias Pfromm, and Ralph Bruder

Contents

Abstract

This chapter gives an overview of the requirements for the input devices for driver assistance functions and the resulting design options: The reader is provided with a systematic procedure for designing input devices according to (Kircher JH, Baum E (Hrsg), Mensch-Maschine-Umwelt. Ergonomie für Konstrukteure, Designer, Planer und Arbeitsgestalter (Man–machine

---

K. Bengler (✉)
Fakultät für Maschinenwesen, Technische Universität München, Garching, Germany
e-mail: bengler@lfe.mw.tum.de

M. Pfromm • R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: m.pfromm@iad.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_35

environment. Ergonomics for engineers, designers, planners and human factors engineers). Beuth Verlag GmbH, Berlin/Köln, 1986). This procedure will begin with identifying the requirements for driver assistance system (DAS) input devices, then follows an explanation of how which body part (ex. finger, hand, etc.), posture and grip type, for system interactions are determined. Additionally, it supports the selection of input devices and provides guidance of how to avoid accidental and unauthorized input. Finally it helps with the design and geometric integration of the arrangement, the definition of feedback and use direction, travel and resistance and the identification of controls. General recommendations are illustrated through specific examples of hardware, demonstrating how manifold input controls can be. In the last part of the chapter an overview of novel operation concepts is given, most of which are not currently implemented in vehicles, however, are estimated to gain importance in the future.

Keywords

Input devices • Input element • Human-machine interface • Manual control • Acting extremity • Body posture • Grip type • Grasp type • Coupling type • Accidental input • Placement • Grouping • Reaching range • Position coding • Gestures • Gaze • Eye-tracking system • Brain computer interface • Speech recognitions

1 Introduction

An input element can be understood as the realization of a technical, human-machine interface that enables the control and regulation of some technical process or sequence to be executed. This term is analogous to “manual control”.

Automotive input devices are interfaces controlled by the finger(s), hand(s) or pedal(s) (Rühmann 1993). Increasingly so, vehicles can recognize many inputs based on gesture or voice recognition technologies. So far, these recognizable inputs are related to driver information functions, but are not directly relevant for driver assistance systems.

A distinction must be made between controls that are used to activate or deactivate assistance functions by discrete inputs and those that are used to adjust a particular driver assistance function. Additionally, there are control elements that allow the driver to continuously control the vehicle in cooperation with a driver assistance system (Eckstein 2000). Through control elements, the driver performs driving tasks that can be divided into primary, secondary and tertiary tasks (Bubb 2003). In the primary task of driving, the driving process, especially maintaining a safe lane position, is of upmost importance. Secondary driving tasks are those relevant to environmental conditions and traffic regulations. Gear shifting, turn signal and assistance system operation (e.g., adaptive cruise control), are also considered secondary driving tasks. Tertiary driving tasks are not directly connected with the driving task; rather, these tasks are relevant to the comfort of the driver (e.g., radio).

As can be seen in Fig. 1 in ▶ Chap. 33, “Design of Human-Machine-Interfaces for DAS”,” control elements, in addition to their input functionalities, also give feedback to the driver such as system state and or entry confirmations.

In ▶ Chap. 33, “Design of Human-Machine-Interfaces for DAS,” a procedure is given on how to approach designing control elements, according to Kircher and Baum (1986). The outline of this chapter is based on this approach.

2 Requirements for DAS Input Devices

As a rule, similar fundamental design requirements are valid for both control elements for driver assistance systems and other driving tasks.

Since input devices are typically used while driving, interactions should be able to occur accurately and error-free without distraction. This is especially true for those input elements related to driver assistance. The main design goals are:

- Quick
- Safe
- Intuitive
- Precise use
- Compatible with the corresponding functions, such as setting display parameters (e.g., so that increasing values will be displayed as either “going up” or clockwise)

Measures that serve to achieve these design goals are:

- The determination of the interacting body part (hereafter, acting extremity), body posture, and grip type
- The selection of input device type
- The avoidance of accidental and unauthorized positions
- The arrangement in the interior
- The definition of feedback and usage direction, travel, and resistance
- The labeling of controls

3 The Determination of the Acting Extremity, Body Posture, and Grip Type

It is important that the placement of interior controls ensures and allows the user to precisely input information and to do so maintaining a comfortable position, especially for frequently used elements (e.g., control actuators for vehicle guidance). No displeasing and or uncomfortable positions should be needed due to poor placement of interior controls.

Selection criteria for the acting extremity (e.g., foot, hand or finger) and grasp type provide the required operating force, accuracy, and speed. In general, for

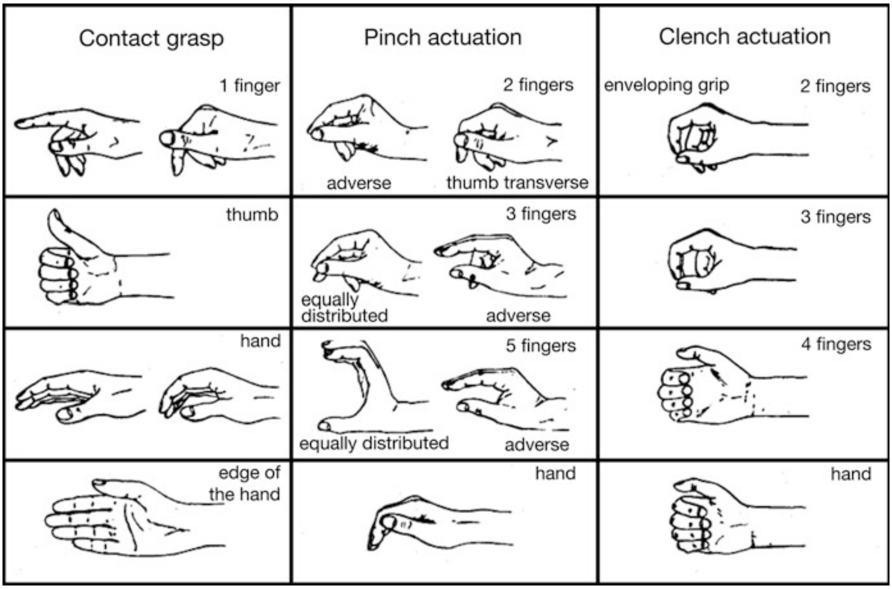


Fig. 1 Grip types (acc. to Schmidtke 1989)

strong and or fast movements, the foot or hand forces (arm movements) are required and for those actions requiring more exactness, the finger is preferred (contact grasp or pinch actuation, see Fig. 1). Therefore for driver assistance systems, hand or finger operation is generally preferred. Furthermore, operation is characterized by the coupling between the body part and control element. It is distinguished between form and force closure. A suitable assignment of grasp type, coupling type, and control element is provided in Kircher and Baum (1986) according to DIN EN 894-3.

4 Selection of Input Device Type

As stated in the introduction, fundamental task types associated with driver assistance functions are distinct:

- Turning a function on/off
- The parameterization and setting of values
- The permanent input in terms of cooperative vehicle guidance, managing, and defining the responsibilities shared by the driver and automation

Depending on the control task, inputs via control elements can either be performed in discrete steps or continuous operation. The definition of input device

Fig. 2 Control panel driver assistance systems in BMW 7 series (BMW AG)

type also depends on the operation purpose, required precision, control speed, distance of the movement, as well as how the vehicle cockpit is arranged. Discrete control tasks are often operated by push buttons and switches, or “rocker buttons.” Rotary knobs and levers are suitable for both discrete and continuous inputs. For example, the light switch in a vehicle is often executed as a discrete input and the temperature of the air conditioner as a continuous rotation. The ACC system is frequently turned on and off by a lever with discrete locking positions. Time gap adjustments are also often executed by a lever. Actuators are integrated in order to be able to provide the user with haptic feedback about the system’s current state, i.e., active control elements. These elements are suitable for a continuous cooperative vehicle operation (Damböck et al. 2010).

Originally, each control element was connected to a specific function. These “hard keys” allow the driver to directly access their function at any time (Fig. 2). However, as the number of controls increases, in order to avoid confusion and conserve the already limited in-vehicle space, multiple keypads are often combined in an integrated control element (Lindberg et al. 2010). The lever for operating the ACC functions is a suitable example for this (Fig. 3). For the “soft keys,” the assigned function depends on the context in a specified area of the screen (Fig. 4), i.e., various functions can be performed with a single key. Menu systems with a display and central control element (Fig. 5) as well as touch screens (Fig. 6) also allow the use of a large number of functions. In terms of operational safety and fast access to specific functions, this increase in functionality access may prove disadvantageous, e.g., causing longer glances and higher cognitive workload, in comparison to direct key operation at longer glance times. Blind operation of touch screens is also virtually impossible because buttons cannot be identified by tactile and haptic feedback.

Fig. 3 ACC-lever (Audi)

Fig. 4 Softkeys (Softkey)

5 The Avoidance of Accidental and Unauthorized Input

Accidental input should be avoided especially for safety-related functions. For this purpose, the following measures can be applied: Buttons should be placed in locations with low accidental contact probability; they should be easily distinguishable based on various criteria (shape, size, location, and color). By necessary

Fig. 5 Menu system with display and central control element: BMW iDrive (Source: BMW AG)

confirmations in dialogue sequences, or unlocking mechanisms for special functions accidental use can also be prevented. A good example of this can be seen in BMW vehicles (Fig. 2) where the control panel for driver assistance systems is usually located on the instrument panel to the left of the steering wheel. This way, accidental operation by the driver as well as unauthorized operation by the passenger is avoided.

Frequently, ACC controls are realized by levers or buttons on the steering wheel; careful attention needs to be directed at not confusing these controls for direction indicators, windshield wipers, or infotainment buttons. Since the driver does not ordinarily look at these control elements to use them, they should be spaced out appropriately and tactilely discernible based on shape, size, and surface configuration of the keys (for example, as in Fig. 7), to allow blind operation. Additionally, these controls should also be visually discernible.

Fig. 6 Touchscreen in Tesla Model S (Tesla 2013)

Fig. 7 Control element on steering wheel in AUDI TTS (Audi AG)

6 Design of the Arrangement and Geometric Integration

According to DIN EN 894-3, control importance, frequency, and execution sequence, should be taken into consideration when determining their placement and grouping.

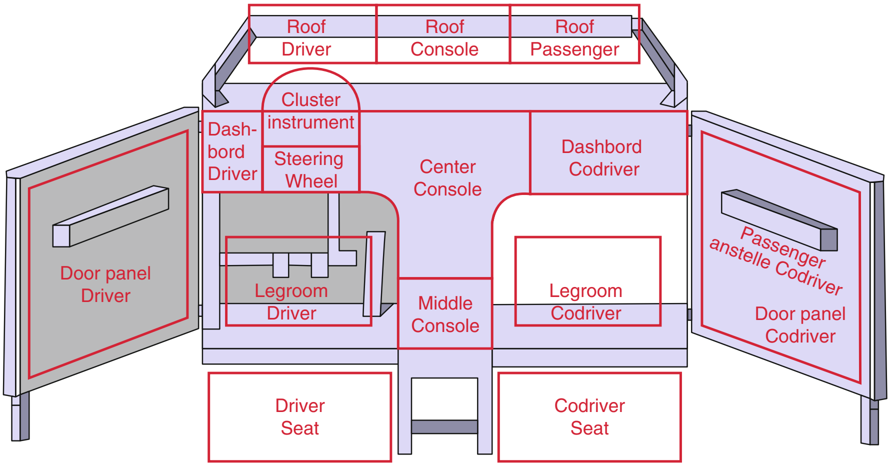


Fig. 8 Definition of placement position/cockpit areas in vehicle interior

The more important a control is for driving safety, the more central it must be placed within the optimum visual and reaching range. Elements that are associated with emergency functions or the shutdown of functions must always be easily accessible.

Other miscellaneous elements with a display element should be able to be reached and seen given a sensible ergonomic seat and steering wheel arrangement, based on the 95th percentile of the male body size and the 5th percentile of the female body size. Evaluating and confirming that the situation is ergonomically sound, can be done through the analysis of a digital human model (e.g., RAMSIS, Human Solutions).

The size of the control element and the spacing between controls is also to be based on anthropometric data where the values for finger sizes and grasping widths are given in tables, e.g., DIN EN 894-3 or (Flügel et al. 1986).

In addition to geometric places, location relative to other keypads is to be considered. In driver assistance operations in the case of operator sequences (e.g., 1. activate, 2. configure the ACC, etc.) changing of one’s grip is avoided by the close proximity of the individual controls.

The significant increase and heterogeneity of driver assistance functions has led to an increase of typical in-vehicle placement positions (Fig. 8) that follow the aforementioned logic (Lindberg et al. 2010). These include:

- To be placed left of the console and to the right behind the steering wheel
- In the steering wheel spokes
- In the middle console area in front of the armrest
- In the central area of the middle console by the instrument panel

7 The Definition of Feedback and Use Direction, Travel and Resistance

The operator should be shown the state of the executed function (e.g., ACC on/off) or active parameters (e.g., size of the desired time gap at ACC). This may take the form of an integrated lighting element (LED) or through some sort of position coding (inclination angle, locking position). Alternatively, the information can be displayed in a proper display and the control element that requires attention, must be indicated. In general, the basic rule is that multimodal, redundant coding is recommended to access more sensory channels of the driver.

The operating direction of controls should be obvious, making them easy to learn and reducing the likelihood of errors, especially in difficult situations. Using stereotypical movements is a suitable way to reduce the ambiguousness of operating directions (Fig. 9).

Additionally, the movement direction of the operating element and the intended direction of the movement of the system should match (motion effect stereotypy).

With the ACC control (Fig. 3), according to the stereotypical movements, increases of speed should be controlled as a lever up movement and to lower the speed, lever down.

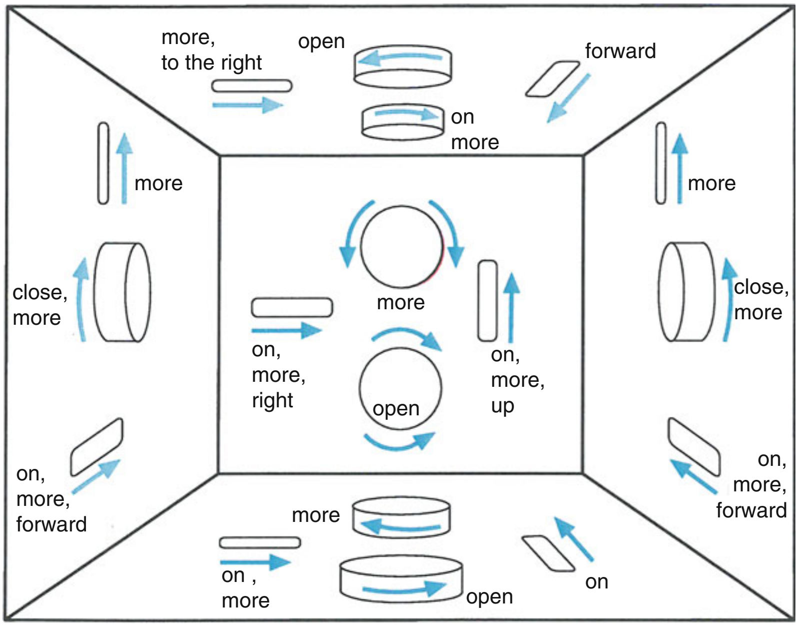


Fig. 9 Spatial alignment of rotary knobs and sliders with associated stereotypical movements (Götz 2007 acc. to Woodson and Conover 1964)

The control path and angle traveled when operating the control element; not all controls have one (e.g., touch sensors, isometric – pathless – controls). Important design rules for conventional controls are:

- That the control element is easily recognizable through the control path
- If stepwise operation is required, secure switch positions through notches (DIN EN 894-3 2010; Reisinger 2006)
- Keep slackness as small as possible, since otherwise control performance can be negatively influenced

In the context of driver assistance systems, the operating resistance of a control element plays a minor role. This resistance, however, is to be considered for the implementation of a tactile feedback and to protect against accidental actuation.

8 The Identification of Controls

Labeling controls is necessary when these are rarely used or if their design is not the same in different vehicles. One way to do this is to establish a form-coding system. Form-coding can be detected both visually and tactilely and allows blind operation. Easily and safely distinguishable forms should be chosen and sharp corners and edges are to be avoided. As an additional possibility, labels or symbols on or next to the operation elements could also be used.

Internationally standardized symbols (ISO 2575) provide guidance on how to use/select forms. When several functions are present, it is important to ensure that symbols for various functions differ enough from each other. Trigams, for example, representing ESC, DSC, ACC, etc., are not intuitive. Additionally, they differ to some extent between vehicle manufacturers and must be learned by the users.

9 Alternative Operation Concepts

The number of vehicle operation functions is steadily increasing. Along with this, driver cognitive and visual workload also increase and issues with the in-vehicle design become more prominent; precisely why human-machine interaction control concepts are being researched.

9.1 Gesture Control

Various consumer electronic devices, such as tablet PCs or game consoles, can be operated by hand, arm, or finger gestures. Therefore, it is logical to consider the possibility of also controlling vehicle functions with gestures.

It has been shown that operating secondary and tertiary driving tasks via gestures requires less eyes off the road time than conventional operation (Geiger et al. 2012;

Fig. 10 Gestures (Kreifeldt et al. 2012)

Kreifeldt et al. 2012). In terms of interior design, fewer controls means less need to be installed, which is needless to say, very advantageous.

In the context of human-vehicle interaction, the following gestures can be distinguished: direction inducing (kinematic) gestures (e.g., pointing or waving left/right), facial gestures (e.g., imitation pick up/hang up of a telephone), deictic gestures (e.g., pointing to a display), and symbolic gestures (e.g., a horizontal swipe for “cancel”) (Geiger et al. 2012) (Fig. 10).

Gestures can be executed on a flat surface, such as a touchpad (touch gestures) or in the air (free-hand gestures) (Pickering et al. 2007).

Free-hand gestures can be carried out between the driver and the windshield, around center of the windshield, and in the area around the center console, below the windshield. For the two latter options, both hands would not be able to execute gestures (Pickering et al. 2007).

To detect the non-contact gestures, infrared sensors or cameras, similar to those found in the Microsoft Xbox Kinect, can be used. For automotive applications, infrared sensor technology has proven to be the more appropriate of the options. Already standard movements, such as touching, pushing, pulling, and wiping can be distinguished. However, non-contact gestures still need to be improved in terms of recognition accuracy and latency between recognition and obstacle interaction (Kreifeldt et al. 2012).

In the automotive sector, free-hand gestures have only been considered for the operation of infotainment systems. For example, gestures such as “move” are used to transfer content from the center console screen to a mobile phone or instrument cluster (Continental 2011). A study (Landrover) suggests controlling turn signals and headlights, so secondary driving tasks, via gesture.

Gestures performed on a touchpad have already been implemented in some vehicles (Fig. 11). The input device is usually integrated into the center console, where finger gestures can be used to navigate through menus and handwriting, in

Fig. 11 Touchpad in Mercedes C-Class (Weckerle et al. 2014)

the form of symbolic gestures, can also be detected (Opel 2013; Weckerle et al. 2014). Haptic and or acoustic feedback is typically given per entry.

In general, it is important to keep things as simple as possible and to utilize gestures already established in consumer electronics. Overly complicated "interaction language" could potentially overwhelm the driver, and is to be avoided (Geiger et al. 2012).

In the "Conduct-by-Wire" project (▶ Chap. 60, "Conduct-by-Wire"), an operation concept for maneuver-based vehicle guidance was implemented where the driver was able to enter a maneuver behavior through gestures on a touchpad, which then automatically executed the input action through the vehicle.

In (Pfeiffer et al. 2010), a steering wheel with a gesture recognition interface was implemented. Unlike conventional steering wheel buttons, many more features can be accommodated and unlike conventional touchscreens, the driver is less distracted during periods of operation. A similar approach was taken in (Ruck and Stottan 2014) where hard and soft keys, as well as a touchscreen with haptic feedback, were integrated into a steering wheel.

9.2 Eye-Tracking Control

In a research project (FU Berlin 2011), vehicle control was performed by the drivers' gaze. Drivers wore an eye-tracking system, which determined the desired direction via recordings of both the pupil and environment. Two modes were implemented: the "free ride" mode, where the viewing directions were directly linked to the steering actuator, i.e., the vehicle is traveling in the direction of the driver's gaze; and the "routing" mode, where the vehicle is mostly automatic and only at certain decision points, such as traffic jams, does the driver select the desired direction via glance direction (FU Berlin 2011).

Fig. 12 EPOC neuroheadset (Göhring et al. 2013)

The widespread introduction of such technology is still in the distant future, but eventually this technology would even make it possible for people with disabilities to use semi or highly automated vehicles.

9.3 Brain Computer Interface

(Göhring et al. 2013) studied vehicle control via brain computer interface (BCI) (Fig. 12) via EEG (16 sensors were used). The driver was to think of different movement patterns (e.g., left, right, push, pull), which were then detected and classified. Similar to that of gaze control, free ride and control of semi-automated system scenarios were investigated. A free ride, meaning that steering, throttle, and brake activities are directly controlled by the BCI, was possible but simply not accurate enough for practical use. Control of a semi-automated system with the BCI was, however, more promising. Here, the driver communicated the desired direction via the BCI at given decision points. The recognition accuracy was 90 %.

9.4 Voice Control

Recently, speech recognition has progressed in the automotive sector such that various functions can already be controlled through free speech. Combinations of on-board and off-board controls are frequently used: if no mobile internet is available, then local speech recognitions are to be used; if internet connection is available, a more powerful cloud technology could be used (Haag 2012). The control of infotainment functions via voice is implemented according to vehicle series. Secondary function control, such as vehicle settings, are usually not possible by voice control (Weckerle et al. 2014).

For people with disabilities, it is possible to retrofit a system so that voice control can be used to control secondary and comfort functions. For example, it is possible to control the direction signal, or to turn lights on or regulate the air conditioning with up to 50 voice commands under any driving condition. The manufacturer specifies a maximum response time of one second with a detection rate of 95 % (Zawatzky 2012).

References

Audi. http://www.audi.co.uk/content/audi/new-cars/a4/a4-saloon/driver-aids/cruise-control/_jcr_content/main-stage/tab_content/container1/image.img.png. Accessed 5 June 2014

Bubb H (2003) Fahrerassistenz primär ein Beitrag zum Komfort oder für die Sicherheit? (Driver assistance primarily to contribute to the comfort or safety?). In: Der Fahrer im 21. Jahrhundert: Anforderungen, Anwendungen, Aspekte für Mensch-Maschine-Systeme. VDI-Berichte Nr. 1768. VDI Verlag, Düsseldorf, pp 25–43

Continental (2011) Magic user interface. http://www.continental-corporation.com/www/presseportal_com_de/themen/pressemitteilungen/3_automotive_group/interior/press_releases/pr_2011_10_19_magic_user_interface_de.html. Accessed 5 June 2014

Damböck D, Kienle M, Bengler K (2010) Die Zügel fest in der Hand halten – Automationsgradumschaltung durch Griffkraftmessung (Hold the reins firmly in hand – automation degree switching by grip strength measurement). In: VDI-Bericht 2099, Useware 2010. VDI Verlag, Düsseldorf

DIN EN 894-3 (2010) Sicherheit von Maschinen – Ergonomische Anforderungen an die Gestaltung von Anzeigen und Stellteilen – Teil 3: Stellteile (Safety of machinery – ergonomics requirements for the design of displays and control actuators – part 3: control actuators). Beuth Verlag GmbH, Berlin

Eckstein L (2000) Sidesticks im Kraftfahrzeug – ein alternatives Bedienkonzept oder Spielerei? (Sidesticks in the Vehicle – an Alternative Operating Concept or Gimmick?) In: Bubb H (ed) Ergonomie und Verkehrssicherheit, Beiträge der Herbstkonferenz der Gesellschaft für Arbeitswissenschaft. Herbert Utz Verlag, München

Flügel B, Greil H, Sommer K (1986) Anthropologischer Atlas Alters- und Geschlechtsvariabilität des Menschen (Anthropological atlas age and sex variability of the human). Minerva – Edition Wissen, München

FU Berlin (2011) http://www.fu-berlin.de/en/presse/informationen/fup/2010/fup_10_106/. Accessed 5 June 2014

Geiger M, Nieschulz R, Zobl M, Lang M (2012) Bedienkonzept zur Gestenbasierten Interaktion mit Geräten im Automobil (Operating concept for gesture-based interaction with devices in vehicles). In: VDI-Bericht 1678, Useware 2012. VDI Verlag, Düsseldorf

Göhring D, Latotzky D, Wang M, Rojas R (2013) Semi-autonomous car control using brain computer interfaces. In: Intelligent autonomous systems 12. Springer, pp 393–408

Götz M (2007) Die Gestaltung von Bedienelementen unter dem Aspekt ihrer kommunikativen Funktion (The design of controls in terms of its communicative function). Dissertation, Technische Universität München, Garching

Haag A (2012) Sprachsteuerung mit semantischer Spracherkennung (Voice control with semantic speech recognition). In: ATZelektronik, No. 06/2012, pp 466–471

ISO 2575 Straßenfahrzeuge – Symbole für Bedienungselemente, Kontrollleuchten und sonstige Anzeiger (Road vehicles – symbols for controls, indicators, and other indicators)

Kircher JH, Baum E (Hrsg) (1986) Mensch-Maschine-Umwelt. Ergonomie für Konstrukteure, Designer, Planer und Arbeitsgestalter (Man–machine environment. Ergonomics for engineers, designers, planners and human factors engineers). Beuth Verlag GmbH, Berlin/Köln

Kreifeldt R, Roth H, Preissner O (2012) Bediensysteme mit Gestensteuerung Stand der Technik und Zukunft (Operating systems with gesture control state of the art and future). In: ATZelektronik, No. 04/2012, pp 248–253

Human solutions. http://www.human-solutions.com/mobility/front_content.php. Accessed 5 June 2014

Landrover (2014) Discovery vision concept. http://newsroom.jaguarlandrover.com/de-de/land-rover/neuigkeiten/2014/04/lr_dvc_lf/. Accessed 5 June 2014

Lindberg T, Tönert L, Bengler K (2009) Integration von aktiven Sicherheitssystemen für die Querführung (Integration of active safety systems for lateral guidance) VDI-Berichte Nr. 2085. VDI-Verlag, Düsseldorf, pp 79–94

Lindberg T, Tönert L, Rötting M, Bengler K (2010) Integration aktueller und zukünftiger Fahrerassistenzsysteme – wie lässt sich der Lösungsraum für die HMI-Entwicklung strukturieren? (Integration of current and future driver assistance systems – how can the solution space for the HMI development be structured?). In: Lichtenstein A, Stößel C, Clemens C (eds) Der Mensch im Mittelpunkt technischer Systeme. 8. Berliner Werkstatt – Mensch-Maschine-Systeme, 7.–9. Okt. 2009. VDI Verlag, Düsseldorf, pp 187–188

Opel (2013). Intuitiv und individualisierbar: neues Infotainmentsystem für den Opel Insignia. http://www.springerprofessional.de/4596298. Accessed 5 June 2014

Pfeiffer M, Kern D, Schöning J, Döring T, Krüger A, Schmidt A (2010) A multi-touch enabled steering wheel: exploring the design space. In: CHI'10 extended abstracts on human factors in computing systems, pp 3355–3360

Pickering CA, Burnham KJ, Richardson MJ (2007) A research study of hand gesture recognition technologies and applications for human vehicle interaction. In: 2007 3rd Institution of Engineering and Technology conference on automotive electronics, pp 1–15

Reisinger J (2006) Haptik von Bedienelementen (Haptic of control elements). In: Ergonomie aktuell, Ausgabe 7, Lehrstuhl für Ergonomie. Technische Universität München, pp 29–33

Ruck H, Stottan T (2014) Interaktives Lenkrad für eine bessere Bedienbarkeit (Interactive steering wheel for better usability). In: ATZ-Automob. Z., Issue 116, No. 5, pp 56–61

Rühmann HP (1993) Schnittstelle in Mensch-Maschine-Systemen (Interface in human-machine-systems). In: Schmidtke H (ed) Ergonomie. Carl Hanser Verlag, München/Wien

Schmidtke H (1989) Handbuch der Ergonomie (Ergonomics handbook). Carl Hanser Verlag, München/Wien

Softkey. http://i1.ytimg.com/vi/-oSzDcJFCHg/maxresdefault.jpg. Accessed 5 June 2014

Tesla (2013) http://image.motortrend.com/f/oftheyear/car/1301_2013_motor_trend_car_of_the_year_tesla_model_s/41007851/2013-tesla-model-s-cockpit.jpg. Accessed 5 June 2014

Weckerle T, Böhle C, Gärtner U, Boll M, Fuhrmann D (2014) Mehr Klasse denn je (More class than ever). In ATZextra, No. 08/2014, pp 28–35

Woodson WE, Conover DW (1964) Human engineering guide for equipment designers, 2nd edn. University of California Press, Berkeley/Los Angeles

Zawatzky (2012) http://www.zawatzky.de/cms/upload/b_fahrzeugumbauten/fernbedienungen/Sprachsteuerung/Sprachsteuerung_deutsch_Bildschirm.pdf. Accessed 13 June 2014

Information Visualization for DAS

35

Peter M. Knoll

Contents

Abstract

In modern vehicles, we are faced with a rapidly increasing flood of information to the driver coming from the own vehicle and neighboring vehicles, from the road, and from telecommunication equipment. In addition to established information systems, driver assistance, collision mitigation, and collision avoidance systems are being integrated more and more in vehicles. The information coming from all those systems must be presented to the driver with appropriate

---

P.M. Knoll (✉)
Institute for Theoretical Electrical Engineering and System Optimization, KIT (Karlsruhe Institute of Technology), Karlsruhe, Germany
e-mail: ifuf.ingb.knoll@t-online.de; peter.barbara.knoll@t-online.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_36

displays taking into consideration the ergonomic requirements of the human/machine interface.

Keeping to the former standard practice, i.e., providing each new information component with its own display and an individual keypad, inevitably would have led to an overloaded cockpit, similar to that in aircraft, on which displays and indicators have to be relocated to unfavorable read-off positions and input elements have to be relocated to areas not easy to reach. Current and future vehicle information systems provide this huge amount of information mainly in three information centers: a more or less reconfigurable instrument cluster and a head-up display, both with driver-relevant information, and a center console display with driver and passenger relevant information. For these systems an appropriate bundling of the information, in conjunction with menu-prompted operating techniques, is essential both technically and ergonomically.

Keywords
Human factors (HMI) • Interaction channels • Code of practice (CoP) • Advanced Driver Assistance Systems (ADAS) • Display technologies • Active displays • Passive displays • Graphic displays • Active LCD • Head-up display • Night vision system • Thin-film transistor (TFT) • Navigation

1 Human/Machine Interface for Vehicles

The human/machine interface (HMI) consists of displays to show the driver of a vehicle its current status and operating elements to interact with it.

1.1 Interaction Channels

There are three channels for the human to interact with the vehicle:

1.1.1 Visual Channel: Seeing

Humans recognize their environment mainly with the aid of the visual sense. Other road users, their position, their estimated behavior, the lane markings, and objects within the road area are recognized with the eyes and interpreted with the highly efficient human picture processing capability. They are selected and evaluated by the human brain with regard to their relevance.

Also the road infrastructure postulates the visual channel: Road signs provide rules, road markings separate lanes, indicator lamps show the intention of a change of driving direction, and braking lights warn from vehicle deceleration.

The visual channel is therefore of highest importance for the driving task. This applies to the conscious seeing with the driver pointing his viewing direction explicitly on objects and focusing on them as well as the peripheral vision which is important for the positioning process within the lane. Therefore, additional views on the vehicle displays during the interaction with driver information and driver assistance systems (DIS, DAS) and their surveillance must be evaluated with regard to possible consequences on the roadworthiness.

1.1.2 Auditive Channel: Speaking and Hearing

For the communication with other road users, mainly for signalizing dangerous situations, the human and advanced driver assistance systems (ADAS) use frequently the auditive channel. It is used for warnings and information from driver assistance systems via speech output as well as commands to speech recognition systems.

1.1.3 Haptic/Kinesthetic Channel: Operating and Feeling

The haptic channel gives a feedback to the driver during all motor activities, when operating switches and while braking and steering. A safety system being in series since a couple of years warns the driver haptically by tightening the safety belt in hazardous situations.

Also the kinesthetic channel, being used by the human for the detection of the vehicles movements, is used in series cars to alert the driver, e.g., with a short braking jerk.

1.2 European Statement of Principles on Human/Machine Interface

Multifunctional displays in vehicles, which can be arranged with displays and messages in accordance with the respective situation or with the driver's wishes, must be structured taking into consideration the regularities of human perception (Bouis et al. 1983).

In the last decade the European Commission published an updated version of the European statement of principles on human/machine interface (ESoP) for in-vehicle information and communication systems (European Commission 2007, 2008). It summarizes essential safety aspects to be considered for the HMI of in-vehicle information and communication.

The principles are not a substitute for any current regulation or standards, which should always be taken into consideration. They can be reinforced by national legislation or by individual companies. These principles constitute the minimum set of requirements to be applied.

The rules and recommendations for information and communication systems apply also for the new group of ADAS, being introduced stepwise in the market. They are based on surround sensors scanning the vehicles' environment and calculating dangerous situations based on objects in the own vehicles' vicinity. Thus, warnings can be given to the driver or the vehicle's computer or may directly interact with the vehicle. Examples are ultrasonic-based parking assistance systems and RADAR-based adaptive cruise control with the further step, the predictive safety systems. They interact with the vehicle's brake and stop the vehicle automatically in the case of an imminent crash. Video-based systems are a third group with major impact on the functionality of ADAS.

1.3 Code of Practice for the Design and Evaluation of ADAS

Within the EU-funded project PReVENT, a code of practice (CoP) for the design of the human/machine interface (HMI) of ADAS was developed and published by the European Commission, Directorate Information Society (European Commission 2008). It comprises a suitable ADAS description concept including ADAS-specific requirements for system development. It summarizes best practices and proposes methods for risk assessment and controllability evaluation. The code of practice has been produced by a group of experts within the RESPONSE 3 project, a subproject of the EU-funded integrated project PReVENT.

The CoP applies to ADAS. It is not specifically intended to be applied to systems providing vehicle stabilization (such as ABS and ESP; those systems do not require an HMI) or mere information and communication systems (such as navigation systems and telephones). It may also be applicable to future systems including vehicle to vehicle communication.

The CoP contains 50 single postulations to the HMI. They mainly refer to human factors, safety of operation, integration into existing vehicle equipment, distraction of and strain to the driver, as well as instructions and information about the use of the systems (Knoll 1986).

1.4 DIS/ADAS Architecture

Different driver information systems (DIS) and advanced driver assistance systems (ADAS) are interacting between each other and with the driver in a non-coordinated manner. There is, therefore, a high risk for the driver to be flooded by a huge amount of information and not being able to distinguish between information with high or lesser importance.

To overcome this problem, an information manager must be introduced to coordinate the information given to the driver in such a way that only information with high priority is presented while information with lesser importance is suppressed in the case that multiple information occurs at the same time (Knoll 2007). As an example, an incoming call is suppressed when the sensors detect a stress situation for the driver and is released only when the dangerous situation is over (Knoll 1986).

2 Display Concepts for Vehicles

We are faced with a rapidly increasing flood of information to the driver. In addition to established information systems (car radio, vehicle monitoring, mobile phones), high-class vehicles feature navigation systems almost as a standard. In the current decade, driver assistance and collision mitigation systems are appearing in vehicles.

Hence, there is an increasing demand for supplying the driver with more information that helps him to drive safer and more economical. In parallel, the price decline in the computer market and the availability of powerful graphic hardware and software concepts make it possible to enhance the classical functions of the instrument board to an interactive multifunctional information panel, with the dashboard being the main interface between the car and driver. These new systems are putting new ergonomic requirements on the HMI. The risk of distraction by this flood of additional information must be counteracted by carefully matching the ergonomic requirements of the human/machine interface and interaction to the system's technical degrees of freedom.

New driver assistance systems must be ergonomically matched with the main communication zones of the vehicle consisting of instrument cluster, head-up display, center console, and rear passenger compartment for mobile office and entertainment. The most stringent requirements as regards optical parameters of a graphic display relate to the presentation of information coming from a video camera.

As drivers are changing their physiological and psychological conditions during a longer journey, there will be the need to adapt the warning threshold to the degree of attention and of distraction and to a possible drowsiness of the driver. Once these human parameters are being measured, it will also be possible to predict the driver's intention in certain situations.

Vehicle information systems provide this information in a more or less reconfigurable instrument cluster, a head-up display, and a center console display. For these systems the question for additional visual and cognitive stress and a possible distraction of the driver by the huge amount of information and its complexity becomes predominant.

2.1 Communication Zones in the Vehicle

If we visualize the diverse range of information available and consider what information is necessary, appropriate, or desirable for the driver and other passengers, we necessarily obtain four different display zones characterized by differing requirements as regards the performance of the relevant display medium. These four communication centers in the vehicle are:

- The instrument cluster for information relevant to the driver at the lower edge of the driver's primary field of view
- The center console for information relevant to driver and passenger

- The windshield for information relevant to the driver, in the primary field of view, and for information presentation without the driver having to take the eyes off the road and without the need for visual accommodation
- The rear passenger compartment as a mobile office or simply as an entertainment zone for the children

Many discussions with European car makers have led to the following recommendations:

Dynamic information to which the driver must respond should be displayed as close as possible to the primary field of view, i.e., in the area of the instrument cluster. If it is intended to achieve a high level of attention, e.g., in the case of warnings from a distance-warning RADAR unit, or if the information in question is guidance information which must be followed at short notice, the obvious choice would be a head-up display (HUD) on which the information is reflected into the windshield and/or through the auditory sensory channel by voice output or other meaningful displays. Guidance information, in this case, is understood as arrows, indicating how to turn at the next intersection and/or simple intersection symbols, *not* maps.

Statistical information, i.e., status information or operator dialog in the form of prompts, should be displayed near to the operator control unit in the center console.

Information aimed at entertaining should be kept away from the primary field of view. This information belongs in the remote zone of the dashboard and should be aimed at the passenger or it belongs in the rear passenger compartment. This is also the right location for the mobile office. The passenger seat backrest is an ideal location for the laptop's display and keyboard.

Driver warnings for lane change or side-crash avoidance should be presented in an area where the driver is most likely to look during a maneuver, i.e., the side mirrors or their housings (Knoll 1997).

2.2 Development of Automotive Instrumentation

The development of the instrumentation over the years shows a clear trend; see Fig. 1 (Knoll 1997): Starting from the simple instrumentation of the 1950s of the last century, visual output of information for the driver has focused almost exclusively on the conventional instrument cluster. More and more information was accommodated in the available installation space.

2.2.1 Instrument Cluster and Center Display

At the beginning of automotive instrumentation, there were only a speedometer and some warning lamps to survey the most important functions. Later, single gauges like revolution indicator, fuel gauge, and cooler temperature have been added to the instrumentation ①.

Until the early 1960s, single instruments have been the order of the day in the cockpit. They were superseded by instrument clusters, i.e., clusters of several

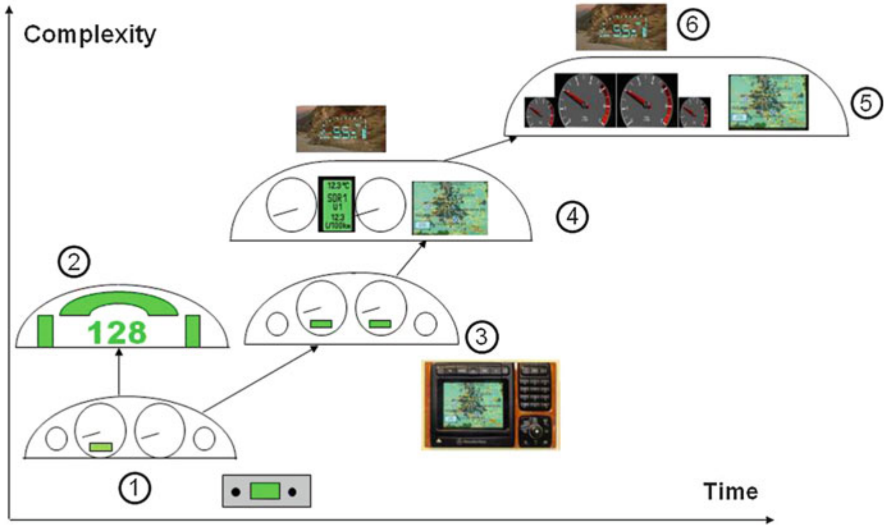


Fig. 1 Evolution of the instrument cluster toward a driver information system

information units in one housing. This approach is, overall, less costly, more reliable, and easier to manufacture and represents an ergonomically better approach, not least owing to the good illumination and antireflection options (Knoll 1997). In the center console region the standard car radio was dominant.

Due to the development of new electronic components and to the necessity of their surveillance, more and more information units have been placed within or onto the available cockpit area leading to a confusing appearance of the information. Even though the appearance of the instrument cluster did not change significantly from the early days, there was a steady development of the intelligence behind the panel. Today, data is provided by a local bus (CAN) or as electrical signals and is collected and interpreted by microcontrollers. Since customers anticipate a broad variety of instrument appearances, manufacturers today offer some ten different dashboard variations per model which requires enormous logistic efforts from the instrument provider.

From this basis originated two development paths:

– Digital instruments ②. This kind of information representation has been established in the USA and Japan mainly in upper-class vehicles. Some attempts to place digital instruments on the European market (e.g., Fiat Tipo and Audi Quattro in the early 1980s) have not been successful, and digital instrument clusters have largely disappeared and play only a niche role. By this development unfortunately the advantages of a digital speed representation, i.e., a quicker and more precise readability, compared to analog representation, have been lost.

– Analog instruments ③. This has remained the most common gauge representation but behind the dial a significant technical change happened.

When the first navigation systems appeared on the market, an additional graphic display was necessary to show the relevant information. Due to space restrictions in the upper dashboard area, it was placed in the center console area because it was easy to integrate it there by using the space of the radio compartment and the ashtray. These displays have been color capable from the early beginning and have a screen diagonal between 4″ and 8″.

Over the course of time, the modern instrument cluster grew in size with additional areas for accommodating also graphic-capable displays ④. Graphic modules in the instrument cluster are better suited for displaying functions relevant to driver and vehicle, e.g., service intervals and check functions for the vehicle's operating condition and vehicle diagnosis for the workshop. Likewise however, they may also be used to display route guidance information from the navigation system. In this connection, we do not mean displaying digitized map excerpts in the instrument cluster of a moving vehicle but, rather, route guidance symbols, i.e., arrows, indicating where to turn off at the next intersection or junction symbols with corresponding turn-off information. One advantage of arranging the graphic display near to the car driver's primary field of view is the fact that this information can be read off more quickly without the driver having to take the eyes off the road for a long period than if placed elsewhere (Knoll and Vollmer 1994).

The next consequent step is the replacement of mechanical gauges by graphic displays ⑤. Reconfigurable instrument clusters, based on a microprocessor-controlled active-matrix color display, provide a powerful alternative to the usual mechanical/electromechanical cluster instruments in vehicles. They help to strengthen passive safety, they adapt to user and situation requirements, and they are easy to install, configure, and maintain. In future cars, reconfigurable instrument clusters will have a huge impact on traffic as they canalize much more information than ever provided to the driver. The functions are manifold and reach from classical driver information like speed to navigation, to video and multimedia access, and to a new class of driver assistance systems with safety functions.

2.2.2 Head-Up Display (HUD)

Conventional instrument clusters are arranged at a viewing distance of between 0.8 and 1.2 m. In order to read off information in the area of the instrument cluster, the driver must accommodate the eyes from infinity (observing what is happening on the road) to a short viewing distance for the instrument. This accommodation process usually takes between 0.3 and 0.5 s. For older car drivers, this accommodation process is a strain and, depending on age, is sometimes almost impossible.

This situation may be improved by using a projection technique as used on the head-up display ⑥. The virtual image of the HUD should never overlay what is happening on the road, in order to avoid distracting the driver from paying attention to the road scenery. The image should be displayed within an area with low information content, for example, floating on the engine hood. In order to avoid excessive stimulation in the primary field of view, the HUD may not be overloaded with information. Consequently, it can never replace the conventional instrument cluster. Also type of information visualized with an HUD should carefully be considered (Roscoe 1981). Information relevant to safety, such as warning indications or an indication of the safety clearance or route guidance information, is, by contrast, well suited to HUD display (Knoll et al. 1993).

Head-up displays with low information content (digital speed indicator and two to three warning icons) are offered mainly by Japanese and some American car makers as an option.

2.2.3 The Rear Passenger Compartment Area

While the information units of the instrument panel are reserved for informing driver and passenger of vehicle and road conditions and for navigation/route guidance and communication with infrastructure facilities, the focus of information presentation in the rear passenger compartment of the vehicle is on entertainment and office communication.

Data transfer and multimedia transmission to the moving vehicle with excellent image and sound quality can be made with the digital audio broadcast (DAB) channel. It can be expected that, as the result of this development, television in the motor vehicle and, thus, the use of TV-capable flat screens in the rear passenger compartment of vehicles will further increase in popularity. With respect to drivers distraction, TV sound should not be audible to the driver.

The rear end of the center console projecting into the rear passenger compartment is an obvious choice for a suitable location for such a monitor. The space conditions allow installation of screen sizes up to approximately 6–8″. As the result of the abovementioned internetworking of the information components, it is also possible to display information from the center console zone, such as information from the navigation system with a travel guide function, in the rear passenger compartment area.

Spacious passenger compartment areas allow installation of the mobile office with all essential information components such as PC, fax, and phone. The mobile phone network provides the capability of data communication between the mobile office and the office at work or at home.

The backrest of the passenger’s or driver’s seat is an obvious choice as a suitable location for the flat PC screen and keyboard. The requirements applicable to the screen are not based on the typical data for in-car use in this application. Rather, these requirements are based on the temperature range conventional for normal office communication, but extended somewhat. Even the screen size and resolution will be based on the office standard, i.e., 10.4″ LCD monitors can be used. Crash-safe keyboards must be available for this application.

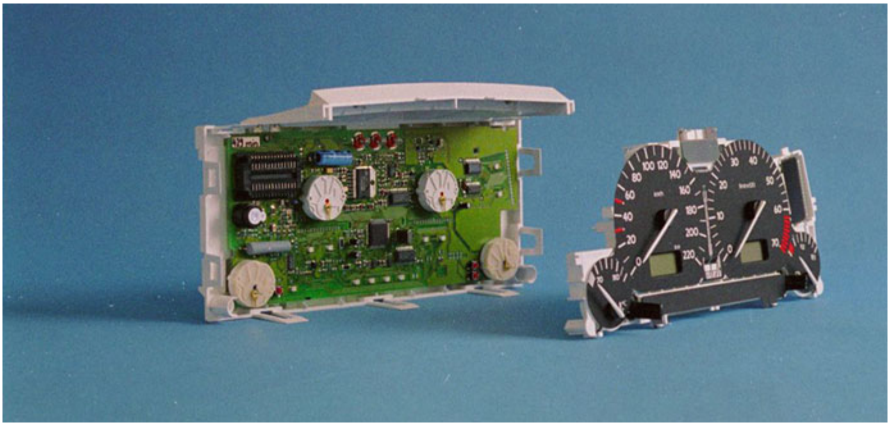


Fig. 2 Exploded view of the mechanical instrument cluster of the Golf 3. *Left*, ECU with stepper motors; *right*, dials and displays

2.3 Display Concepts for the Instrument Cluster

Depending on the installation location and information content of the relevant display, we obtain greatly differing requirements applicable to the relevant display.

Until today the plurality of automotive instruments are still using mechanical instruments with dials and pointers. The bulky eddy-current speedometer has disappeared and was replaced by much smaller and electronically controllable cross-coil moving-magnet movements. Meanwhile robust stepper motors are dominating with a very small thickness. Figure 2 shows the instrument cluster of the Golf 3, introduced firstly in the early 1990s. It incorporates four stepper motors and three LCD devices for odometer and trip computer as well as a gear indicator in the upper right.

2.3.1 Dial Illumination

Formerly, instrument clusters had thin aluminum dials being illuminated from the front with incandescent lamps and small plastic light pipes. Meanwhile the transmissive illumination is predominant due to its appealing appearance. Later, the incandescent lamps were replaced by light-emitting diodes (LEDs). LEDs are not only well suited for warning lamps but also for the rear illumination of dials, displays, as well as pointers in plastic light-pipe technology.

For special, sophisticated geometries in some instrument clusters, cold cathode fluorescent lamps (CCFL) have been used. This technology allowed the setup of an instrument cluster with a tinted cover with exceptional brilliance. Also active-matrix LCDs need a powerful backlight due to their low transmission in the range of 6 %.

Electroluminescence (EL) foils for dial illumination have a very homogeneous light distribution, but they did not meet automotive requirements before the year

Fig. 3 Digital instrument cluster of the Audi Quattro in VFD technology (*top*) and in LCD technology (*bottom*)

2000. They are still more expensive than LEDs and have therefore largely disappeared.

2.3.2 Alphanumeric Displays

Alphanumeric displays in liquid crystal (LC) or vacuum fluorescent (VF) technology belong meanwhile to the standard of instrument clusters as odometer or trip computer. These technologies also allow realizing a complete instrument cluster. Due to limited sizes in former times and due to limited multiplexing capability, these instruments have been composed of plural single display modules. The upper picture in Fig. 3 shows the digital instrument of the Audi Quattro in VFD technology (1984). The lower picture shows a view on the LCD version (1986). In this instrument the chip-on-glass technology has been introduced for the first time in an automobile.

With these segmented displays a realistic appearance of pointers cannot be achieved. The speed must therefore be displayed digitally.

Instrument clusters in VF technology are in widespread use in Japanese and also in some US American vehicles.

2.3.3 Small Graphic Display Modules in the Instrument Cluster

The increase of information in the instrument cluster requires graphic-capable modules which can display any information in a flexible way and arranged according to their priorities. This tendency leads to a mechanical instrument with an integrated graphic display. Figure 4 shows an example.

Besides the previously mentioned driver- and vehicle-relevant information, these graphic modules allow to display information from a navigation system (e.g., arrows) in the instrument cluster.

Fig. 4 Graphic module integrated into a mechanical instrument showing the information from a semiautonomous parking system

Automotive modules of this type have initially been monochrome versions. With the time and with cheaper prices, they are being stepwise replaced by color displays as the advantages of the color display in relation to read-off speed are preferably if properly implemented.

2.3.4 Graphic Display Screens in the Instrument Cluster

Since the early 2000s larger and graphic-capable display screens in active-matrix LCD (AMLCD) technology are used in instrument clusters for replacement of the mechanical gauges and to increase the information content of the instrument. In 2005, a reconfigurable LCD instrument cluster was introduced in the Mercedes S-Class. It is used to show a large analog pointer instrument and alternatively a night vision picture from a video camera. This technology allows the flexible definition of the appearance on downloading a program at the back end of the manufacturing line and avoiding screen printed dials. For the car maker there is the chance of having only one type of instrument for all cars of a model line (Knoll and Vollmer 1994).

Figure 5 (top) shows the instrument cluster of the Mercedes S-Class 2005, consisting of a combination of an 8″ LCD and three mechanical side gauges with their dials matched perfectly to the appearance of the LCD screen in the normal operating condition. When the night vision system is activated by the driver, the video picture is shown on the graphic screen, while the speedometer appears as a bar graph below the video picture; see Fig. 5 (bottom) (Knoll and Reppich 2006).

The next step, a complete substitution of all mechanical instruments with a full-size graphic display has been done with the instrument cluster of the Mercedes S-Class 2013 allowing to display alternatively the information of quasi-analog information of speed and information from other in-car components like adaptive cruise control (ACC), radio, ESP, reversing aid, etc., as well as the picture of a night vision camera.

Fig. 5 Instrument cluster with large graphic display (8″ diagonal); normal mode (*top*) and in night vision mode (*bottom*)

2.3.5 Large Graphic Displays in the Center Console

The first navigation systems have been developed in the 1980s as nomadic devices. Due to the limited temperature range of AMLCDs at that time, the first systems in Europe used a 4″ CRT on a gooseneck. They have been superseded by AMLCDs in the early 1990s. Starting with the navigation function, the systems became quickly sophisticated, combining all additional information from components in a compact display and operator control unit. It was also practical to have telephone and the heating and air-conditioning systems operated via the central keypad. Best results have been obtained with not more than ten multifunction keys and not more than three menu planes (Heintz and Knoll 1982) (Knoll and Elke 1990). Touch screens are less suited due to unavoidable fingerprints on the screen and the lack of haptic aids. The components are networked via bus systems and are dialog capable. Since this information is required by both driver and passenger, it was obvious from the point of readability and operability to arrange this all-purpose terminal in the center console. While the first graphic displays had been placed in the lower console area due to better space conditions, today they are mostly placed in an ergonomically more favorable place on the same level as the instrument cluster.

Fig. 6 Dashboard of the Mercedes S-Class 2013 (*top*) and operating element for all driver information functions (*bottom*)

This information is output visually using a graphic-capable AMLC display whose requirements as regards resolution and color performance are determined by the most demanding function, navigation with map display, or image display for TV. Route guidance information is additionally provided with voice output as realized in most systems available on the market.

Figure 6 (top) shows the dashboard of the new Mercedes S-Class 2013 with the appropriate operating element, arranged in an ergonomically favorable position in the center console (Fig. 6, bottom).

2.3.6 Dual-View Display

New addressing methods allow the realization of a so-called dual-view display. This method allows showing two different pictures at the same time, one with driver-relevant information with a viewing direction to the driver and the other with a different information content (e.g., movies) with the viewing direction to the front passenger.

2.3.7 Head-Up Display (HUD)

The optical system of the HUD generates a virtual image at such a large viewing distance that the human eye no longer needs to accommodate, i.e., it can remain

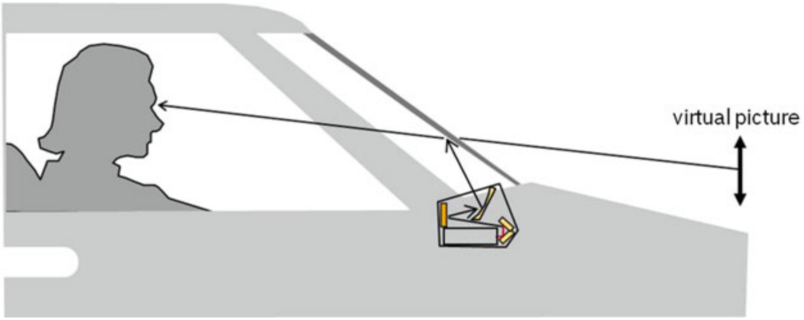


Fig. 7 Principle of a head-up display

accommodated at infinity. As a further advantage, the information can be read off without averting one's view to the instrument cluster, i.e., it increases eyes-on-the-road time.

The principle of the HUD is shown in Fig. 7. It contains the display with its drive unit for generating the image, a lighting system, an imaging optic, and the windshield as combiner. The combiner reflects the picture to the driver's eyes. Optical elements in the optical path (usually concave mirrors) enlarge the viewing distance. The windshield may be covered with a reflective layer to improve the reflectance of the window glass.

To avoid double reflections caused by the inner and the outer glass surfaces, a wedge-shaped plastic layer is integrated into the multilayer glass. For the driver's view the two reflections are covering each other.

Figure 8 shows the appearance of an HUD in a series vehicle (BMW) showing digital speed and the information from an adaptive cruise control (ACC) system containing set speed, "object ahead detected," and "system operating in the ACC mode."

Particularly good results may be achieved in future with holographic combiners. In this case, the hologram of a plane mirror is embedded as a film in the windshield's laminated glass structure.

2.3.8 Displays for Night Vision Systems

Due to limited space in the dashboard area and also due to cost, an additional display for a night vision system cannot be considered.

Using the large graphic display in the instrument cluster (see Fig. 5) has the advantage that the driver, while cruising at night, must concentrate only on two supervision areas, the road and the instrument cluster.

Own (unpublished) investigations have shown that with a brilliant picture of a near-infrared (NIR) system, a very fast interpretation of the picture can be done by the driver, lasting not much longer than the usual look at the speedometer. Investigations with test persons accomplished by the universities of Berlin and Chemnitz,

Fig. 8 HUD in a series vehicle

Germany (Mahlke et al. 2007), showed the best acceptance for this form of information presentation.

Other vehicle suppliers who do not use large AMLCD screens in their instrument clusters to show the night vision picture must use the AMLC display in the center console. To reduce the distraction from this third supervision area, the warning from a detected pedestrian is shown with an HUD. Figure 9 shows such an example of an NV system making use of the passive, far-infrared (FIR) detection technology.

2.3.9 Nomadic Devices

Due to the widespread use of cellular phones with the navigation app, today very often these nomadic devices are put on the windscreen with sucker bowls.

These systems are autonomous and may be connected via Bluetooth with the audio system. They are also using AMLCD displays or AMOLED displays (AMOLED = active-matrix organic light-emitting diode).

Unfortunately, these devices have very small operating elements and menu structures not being in convergence with the operating philosophy of the car maker with the risk of a higher distraction compared to the integrated systems. The EU has taken these systems under supervision and is considering specific regulations for this kind of HMI.

Fig. 9 FIR night vision display with warning from a pedestrian (Photo courtesy of BMW)

3 Display Technologies for the Vehicle Instrumentation

From the large number of available display technologies, only a few have remained for automotive applications due to the high requirements of this rough environment.

At first, purely mechanical gauges dominated, combined with just a few warning lamps. With the development of new display technologies, these bulky and heavy units have been almost completely replaced by electronic displays.

3.1 Evaluation of Displays

The evaluation of the suitability of a display for a certain application can be made on criteria, which can be summarized in three groups.

- The most important *optical* criteria are contrast, luminance, and readability at high ambient light and color capability. They must always be evaluated with respect to the intended application.
- The *technical-economic* criteria are mainly influenced by the physical properties of the respective technology. They contain criteria like operating voltage, power consumption, switching times, addressability, multiplex operation, and cost.
- Finally, displays can be evaluated from the *application* point of view. For automotive application there must be mentioned mainly: temperature range,

space requirements for cockpit integration, failure rate, resistance against humidity, pressure changes, and mechanical shock.

A display technology meeting optimally all these requirements does not exist and requires making a compromise between effort and capabilities of the display system.

One of the most important requirements of a display is a good readability over a large range of ambient luminance. Here, the contrast is the decisive feature. It is defined as the ratio between the luminance of the character L_C and its background L_B:

$$C = \frac{L_C}{L_B}$$

Besides this, two kinds of contrast are defined: The contrast of dark characters on light background is called “positive contrast,” while light characters on dark background are called “negative contrast.”

Usually, the numerator or the denominator is set to 1 (e.g., C = 1:8 or 10:1) depending on positive or negative contrast.

The readability of a display usually is being evaluated subjectively: The contrast of a typical permanent display like a newspaper usually is evaluated as good. It is in a region of 1:7. The subjectively best contrast – following experiences from color TV – is in a region between 10:1 and 5:1, depending on the current adaptation condition of the observing person (Knoll 1986).

3.2 Electromechanical Gauges

In Europe, there is a continued trend toward the classic instrument with mechanical pointers and dials even though major changes have occurred behind the dial. In the early beginning the voluminous eddy-current speedometers dominated, consisting of a permanent magnet driven by a mechanical axle. The mileage counter was mechanically driven too. Due to their bulkiness they were firstly superseded by cross-coil moving-magnet pointer movements.

With this movement (Fig. 10, top) a permanent magnet is enclosed by two crossed coils. By appropriate addressing the pointer can be moved within the four quadrants. A circular spring (not shown) brings the pointer back to the pin at the zero pointer position without voltages applied.

Nowadays, the far more compact stepper motor based on the Lavet principle (Fig. 10, bottom) leads to a greatly reduced overall depth and a far higher indicating accuracy. It is a stronger version of the stepper motor commonly used for watches.

In a Lavet stepper motor, a very small permanent magnet is embedded between yokes of ferroelectric material. Agitated by an alternating current, the magnet is turned and drives the pointer axle via a bi-level gearbox.

Gearbox stepper motors have a depth between 5 and 8 mm. They have a low power consumption of only 100 mW, and they allow a very quick and accurate pointer positioning with a high momentum.

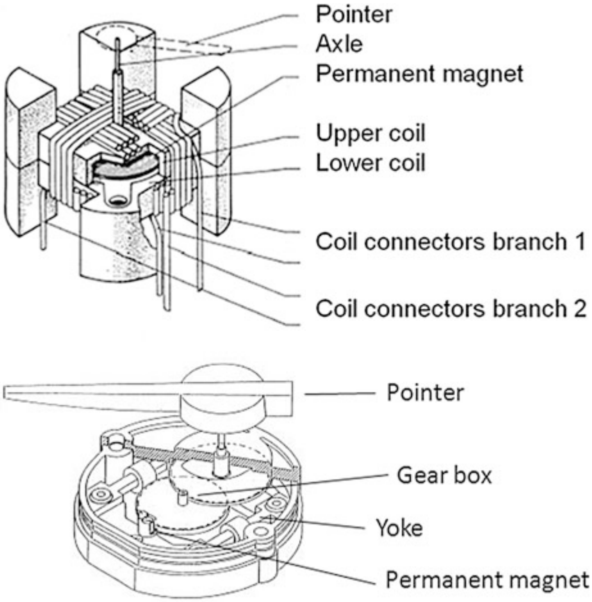


Fig. 10 Cross-coil moving-magnet movement (*top*) and stepper motor (Lavet principle) (*bottom*)

Also the voluminous, mechanical counting mechanism for indicating total and daily mileage is meanwhile completely superseded by the conventional LCD.

3.3 Active Displays

It is common use to distinguish displays based on their physical phenomena into active and passive displays.

Active displays emit light and consume therefore much energy if they are bright enough to be readable with good contrast at high ambient luminance which happens frequently in automobiles. The main active display representatives for automotive application are incandescent lamps, light-emitting diodes (LEDs), organic light-emitting diodes (OLEDs), and vacuum fluorescent displays (VFDs). Gas discharge displays or plasma displays (PLD) and electroluminescent displays (ELD) until now do not play a role in automotive applications. The bulky cathode ray tubes (CRT) have disappeared from automobiles.

3.3.1 Incandescent Lamps

Besides illumination, incandescent lamps are used in automobiles as warning lamps or for the illumination of dials and passive displays. As a mass product, they have very low prices but their electro-optical efficiency is very low with less than 4 %. The standard lifetime of an incandescent lamp lies between 1,000 and 5,000 h when operated at rated power.

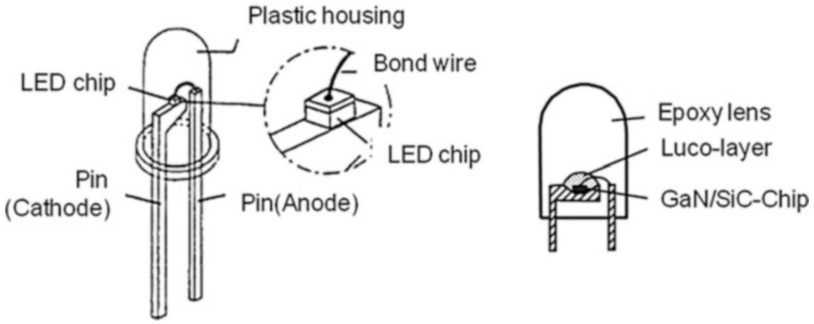


Fig. 11 Setup of a standard LED (*left*) and of a LucoLED (*right*)

3.3.2 Light-Emitting Diode (LED)

The generation of light in LEDs is based on the recombination of electrons from the conduction band with holes from the valence band. As a consequence, the generated energy is emitted in the form of photons. Before this happens, the semiconductor must be agitated by a voltage.

With the standard materials gallium arsenide (GaAs, red), indium phosphide (InP, red to yellow), silicon carbide (SiC, blue), and gallium nitride (GaN, blue), all colors of the visible spectrum can be generated. In spite of their initially limited brightness and higher cost, they have been used as warning lamps because they used to have a much lower failure rate and higher lifetime than incandescent lamps.

White light was initially generated by additive color mixture by using multiple LED chips of different primary colors in one housing. Later, with the invention of the LucoLED (Light Conversion LED), the breakthrough to high-brightness white LEDs was accomplished. With this new technology a bright LED-based backlight for dials and graphic display modules has been realized, even with tinted cover glasses of instrument clusters. In LucoLEDs a part of the blue light of a gallium nitride chip is converted in red and green colors by a conversion layer placed on top of the LED chip. The result is white light.

The schematic setup of standard LED is sketched in Fig. 11, left. The right picture shows the basic setup of a LucoLED.

The mostly very small semiconductor chip with the dimensions of $0.35 \times 0.35 \times 0.2\ \text{mm}^3$ and evaporated with metallic contacts is glued with a conductive epoxy on a metallic pin. The anode is connected to the second pin with a gold wire by wire bonding. After this, the system is embedded into epoxy resin. This cover leads to a good mechanical stability. It also influences the radiation characteristics.

Typical technical data of LEDs:

- Addressing voltage (depending on semiconductor material): 2–10 V
- Operating current: 1–100 mA

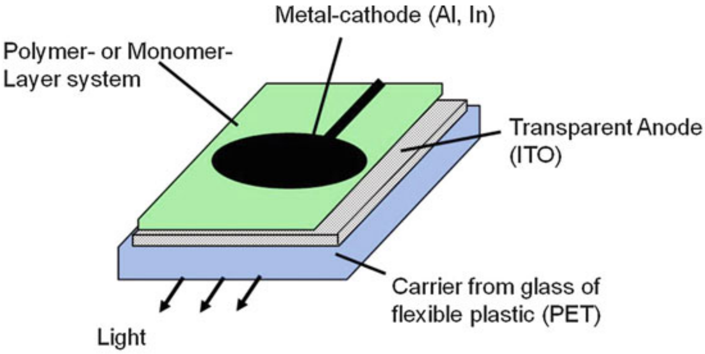


Fig. 12 Setup of an OLED

- Temperature range: −50 °C to +100 °C
- Lifetime: 10^8–10^{10} h

3.3.3 Organic Light-Emitting Diode (OLED)

Recently, also displays on the base of segmented organic light-emitting diodes can be found in vehicles as a substitute of small graphic LCDs or VFDs.

The organic layer system is embedded between two electrodes. Figure 12 shows the general setup of an OLED picture element (pixel).

Indium tin oxide (ITO), sputtered on a transparent plastic substrate, is used as electrode material for the anode. The cathode consists of metal, e.g., aluminum or indium.

The generation of luminescence is caused by injected charge carriers inside the layer system. When a voltage is applied, electrons are injected by the metallic electrode and, vice versa, holes by the anode. The charge carriers are propagated through the polymer layer and recombine under emission of light when hitting each other.

In principle, OLEDs are distinguished by their luminescent materials, by the conjugated polymers (poly-OLEDs) and small organic molecule compounds (monomers). The small molecules are evaporated in vacuum, while the conjugated polymers are spun on the electrode on a rotating table in dissolved form.

The mostly used material for OLEDs is tris(8-hydroxyquinoline)aluminum, Alq3. The spectral emission of the materials is broad banded. In principle, all colors can be realized with this technology.

OLEDs in polymer technology can be produced at low price, but only small displays with a hermetically sealed glass front plane have the ambient stability required for automotive application. OLEDs in monomer technology have a good homogeneity of the layer system with homogeneous and good optical properties, but their production is expensive due to the required vacuum processes.

In case of the intrusion of humidity at the display’s edges, a damage of the organic layer may occur. Temperatures above 70–80 °C are critical as well for the lifetime.

Fig. 13 OLED display in the Mercedes E-Class (2005) for climate control

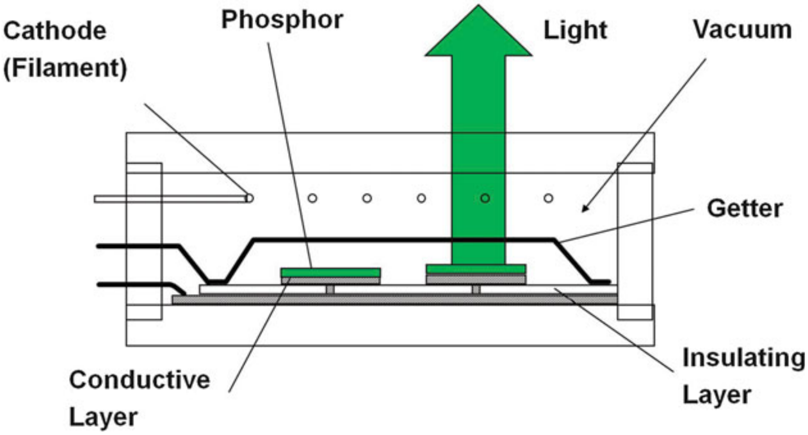


Fig. 14 Setup of a planar vacuum fluorescent display

These drawbacks have prevented the automotive application of colored graphic modules in OLED technology, as they are widely used in consumer applications (e.g., cellular phones). Due to the severe requirements to ambient parameters, OLEDs have been used until now only for small monochrome alphanumeric and simple graphic displays.

Figure 13 shows the display module of a climate control in OLED technology in a vehicle.

3.3.4 Vacuum Fluorescent Display (VFD)

The vacuum fluorescent display operates on the principle of a vacuum tube (triode). The anode is covered with a phosphor. As soon as the electrons, emitted by the cathode, are propagated with sufficient energy by the anode voltage, the phosphor emits monochrome light. Figure 14 shows the cut through a planar vacuum fluorescent display.

The directly heated cathode consists of tungsten wires covered with a layer with low work function. The getter electrodes are produced by etching thin foils of stainless steel and are structured like honeycombs. The structure of the anode (segment) is produced in thick-film layer technology. Standard VFDs use a ZnO phosphor whose spectral response covers a range from 400 to 700 nm with a maximum in the bluish-green spectrum at 505 nm. From this emission spectrum all colors can be produced from blue to red with appropriate filters, with a violet filter also white.

VFDs for consumer applications are driven in multiplex mode, while in the automobile they are directly driven (segment-wise) due to the high-brightness requirements in the automobile.

Typical technical data of VFD:

- Power dissipation: 15–125 mW/digit with a seven-segment display
- Lifetime: 50,000 h
- Temperature range: −40 °C to + 85 °C
- Luminance: 300 cd/m^2 with direct addressing

3.4 Passive Displays

Passive displays emit no light; they modulate incoming light and consume, therefore, very little energy. Some technologies (not for automotive application) also store the information. The most important representative of passive display technologies is the liquid crystal display (LCD).

3.4.1 Liquid Crystal Displays (LCDs), TN-LCD

Liquid crystal materials are liquid in a defined range of temperature, while they possess the anisotropic behavior of crystals. The dielectric anisotropy allows influencing the orientation of the cigar-shaped molecules, while the optical anisotropy allows making this effect visible in polarized light. For the application in displays, LCD segments are used as light valves.

From the wide variety of possible electro-optical effects in liquid crystals, the “twisted nematic display” (TN-LCD) has the highest importance. TN-LCD technology is the most widespread technology in automobiles. Figure 15 shows the principle.

In a TN cell the molecules are oriented parallel to the glass surface. In the off-state the main directions of the molecules and the limiting glass plates are twisted in an angle of 90° to each other (left picture). The polarization direction of linear polarized light entering the layer is twisted by 90°, corresponding to the twist of the molecules.

With crossed polarizers the liquid crystal element is transparent (positive contrast cell). When applying a voltage at the ITO electrodes (right picture), the molecules within the layer are turned into a position parallel to the electric field and perpendicular to the glass plates. The polarization direction of the light cannot follow the abrupt transition; the cell is opaque to the light.

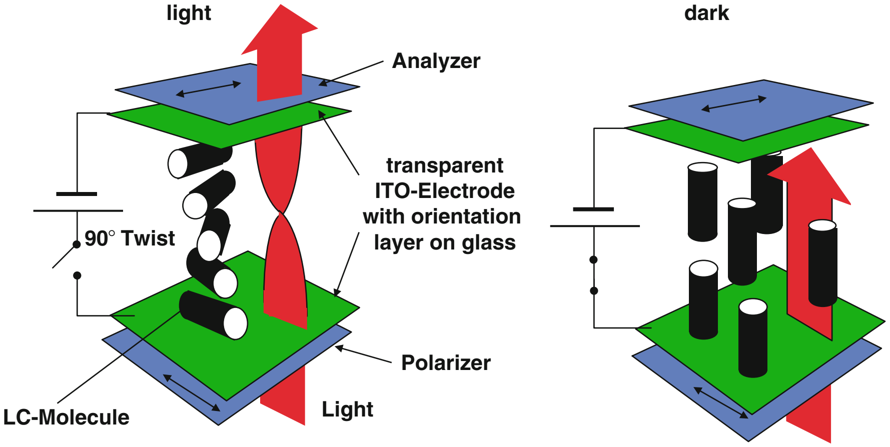


Fig. 15 Optical behavior of a TN cell between crossed polarizers: off-state (*left*) and on-state (*right*)

Typical technical data of LCDs:

- Power dissipation: μW/digit
- Lifetime: >100,000 h
- Temperature range: −40 °C to + 110 °C

LCDs with negative contrast are made by parallel arrangement of the polarizers on the liquid crystal cell.

TN technology is in no way suitable only for small display modules in the instrument cluster. Implementing even large display areas and designing modular structure or even whole-area LCD instrument clusters pose no problems. These segmented displays do not allow a realistic simulation of mechanical pointers. Consequently, speed and engine speed are displayed digitally or as bar graphs.

3.4.2 STN and DSTN Display for Displaying Maps

Super-twisted nematic (STN) and double layer STN (DSTN) technology can be used for modules with moderate resolution. In STN displays the liquid crystal layer is twisted by 180–270° giving the cell a steeper electro-optical response compared to a standard LCD and can thus be driven at a higher multiplex ratio allowing to realize smaller graphic displays. The major drawback is the color of the STN display: It can only switch between yellow and blue and is therefore not well suited for automotive application.

Adding a second LC layer twisted at the same twisting angle in the reverse direction than the first layer, a color-neutral appearance results. The DSTN display can be implemented as a monochrome or a multicolor version. One example of the use of this display technology is the instrument of the Mercedes compact-class vehicles as shown in Fig. 4. The red appearance is produced by backlighting with red LEDs.

Multicolor capability can be implemented by attaching red-, green-, and blue-colored filters using thin-film technology to the inner side of one of the two glass substrates. Gray levels are not possible under in-vehicle conditions. The color range is therefore restricted to the primary and secondary colors, to black and white.

The performance of a DSTN display is adequate for simple map displays. However, the color DSTN does have an inadequate color performance when viewed at a large viewing angle. It can therefore be used only to a limited extent.

3.4.3 Graphic Display Modules for Complex Information

High-resolution, video-capable AMLC displays are required for visually demanding presentation of complex information in the area of the instrument cluster, the center console, for TV applications and for screens for the mobile office. This requirement is met today only by the actively addressed liquid crystal display. The active addressing of the individual pixels is realized with thin-film transistors (TFT).

TFT-LCDs consist of an “active” glass substrate and the counter-plate with the color filter structures. The active substrate accommodates the pixel electrodes made from ITO, the metallic line and row electrodes and the semiconductor structures. At each intersection of line and row electrodes, a thin-film transistor is located being etched from a previously deposited layer structure in a sequence of four mask steps. The TFT is a polycrystalline variant of the MOS field-effect transistor whose production has been optimized in such a way that the structures can be produced by self-adjusting photolithographic processes (Glueck et al. 1994; Glueck and Lueder 1993; Glueck et al. 1993a, b).

The opposite substrate accommodates the color filters and a black matrix structure. The latter leads to an improvement of the display’s contrast mainly in high-ambient-light situations because it avoids reflections from the metallic electrodes. The color filters may be applied either in the form of continuous stripes or as mosaic filters. Stripes provide good rendition of graphic information, while a mosaic array is particularly suitable for video images. The filters today are mostly produced in lithographic processes. An alternative ink-jet printing process has been introduced in production. On top of these layers, one counter-electrode from ITO for all pixels is located.

Figure 16 shows an opened-up TFT matrix.

For vehicle application TFT displays are available with diagonals from 3.5″ to 8″ with extended temperature range from −25 °C up to +85 °C. This allows a reconfigurable display with formats exceeding 10″.

3.4.4 Display Technologies for Head-Up Displays

Both active and passive displays can be used for the pattern generation of HUDs.

For simple applications like digital speed, indicator lamps, or fuel gauge bar graphs, monochrome (bluish-green or white) VFDs are in common use mainly in

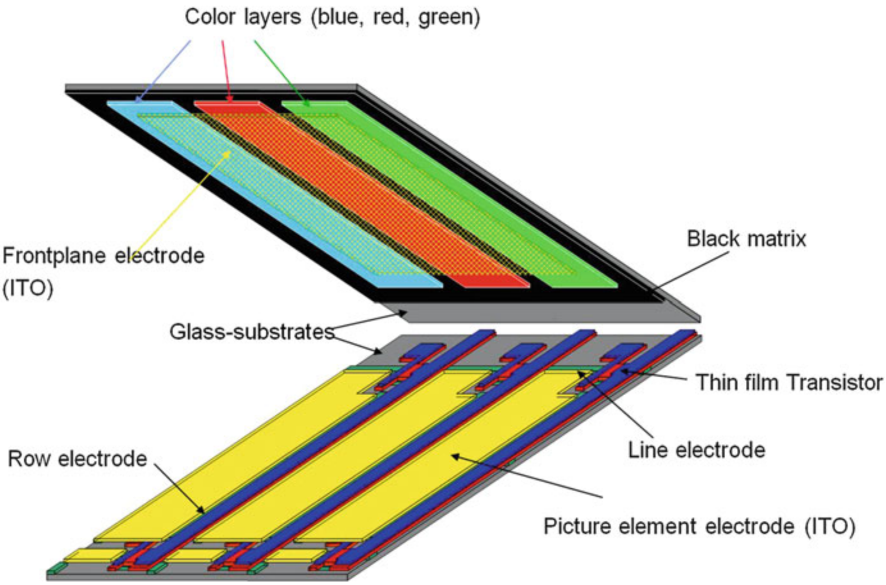


Fig. 16 TFT matrix display

Japanese cars. For displays with higher information content monochrome, highly transmissive LCDs are used as light valves in vehicles. These may be DSTN displays or (actively addressed) polymer-dispersed liquid crystal (PDLC) displays. DSTNs have the disadvantage that they absorb a substantial share of the light passing through, i.e., around 70 %, owing to the required polarizers. This necessitates very bright illumination leading to a high-power dissipation of the system (Roscoe 1981).

The HUD shown in Fig. 8 makes use of a TFT-LCD. To achieve the required luminance, an LED matrix with high brightness is used as backlight.

PDLCs are under development and may be used in the near future for HUDs. The PDLC consists of a polymer in which microscopic liquid crystal droplets are embedded. When the cell is de-energized, incident light is scattered. This means that the projected image appears dark. With applied voltage the pixels become transparent leading to a bright image on the windscreen. One very attractive feature of this technology is that it does not need light-absorbing polarizers (Lueder 1993).

3.5 Future Automotive Display Concepts

Besides the mentioned display concepts new developments mainly in the field of HUDs are under development.

3.5.1 Contact-Analog Head-Up Display

At a first glance a video picture from a night vision system projected on the windscreen seems to be a good solution. The first systems on the market from US and Japanese suppliers in the early 2000s showed the picture above the instrument cluster or above the center console. Experience with the product showed however that the distraction caused by moving pictures in the primary field of view or very close to it leads to high distraction and some suppliers took the systems from the market.

Today's systems show the picture in the instrument cluster or in the center console as described above. New picture processing algorithms allow detecting relevant obstacles in front of the car like pedestrians, bikers, or animals. The second generation of night vision systems warns the driver from these objects acoustically and marks, in parallel, the detected object in the graphic screen in the dashboard with a symbol; see Fig. 9. This leads to less distraction because the driver looks only on the display when he is warned.

An even better solution would be to refrain from showing the video picture on the graphic display but using instead only warning symbols on the windshield with an HUD. An optimum would be to show the warning symbol in the area of the windscreen where the object later appears for the driver. So he knows in advance in which direction he has to point his attention. Also arrows indicating the recommendation from the navigation system could be placed right into the intersection which is of significant use in complex intersection geometries.

An HUD of this kind is called "contact analog." It is available since 2014 in upper-class vehicles. This technology is ideal with respect to recognition and interpretation of objects with a minimum of distraction leading to a higher driving safety. The technical effort however is high with respect to the projection system because a very large projection area on the windscreen must be covered. Another drawback might be that the head position of the driver must be known by the system in order to project the information into the right area on the windscreen (Herzog 2006).

3.5.2 Laser Projection

Today's HUDs need a very bright luminance at bright ambient light to be readable. With direct sunlight the brightness of the HUD is often too low. The problem can be solved with a laser projection unit. Such a system projects the information generated by lasers on an intermediate screen, and from there it is projected on the windscreen by projection optics. A video controller drives the color lasers, emitting in the three primary colors red, green, and blue. With dichroic mirrors the three rays are united to one beam which is projected via a micromechanical two-axle scanning unit on the intermediate screen. Due to the high light output of the lasers, enough picture luminance is available even at the highest ambient brightness with about a factor of 10 compared to the standard HUD. Series introduction is intended mid of the decade depending on the availability of the blue laser (Herzog 2006).

References

Bouis D et al (1983) Ergonomische Optimierung von LCD-Anzeigen im Kraftfahrzeug (Ergonomic optimization of automotive LC-Displays). Proc. ISATA

European Commission (2007) Commission recommendation on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human machine interface. EU publication from February 6th, 2007. http://europa.eu.int/eur-lex/lex/JOHtml.do?uri=OJ:L:2007:032:SOM:DE:html

European Commission (2008) Code of practice for the design and evaluation of ADAS, 07.05.2008. www.prevent-ip.org/download/deliverables/IP_Level/PR-04000-IPD-080222-v16_PReVENT_Final_Report_080507.doc

Glueck J, Lueder E. (1993) Verfahren zur Herstellung einer Matrix aus a-Si:H Dünnschichttransistoren (A method for producing a matrix of Si:H thin film transistors) German Patent Specification P 4 310 640.4 C1

Glueck J et al (1993a) A three-mask process for high mobility a-Si:H TFTs demonstrated by a 14“AM-Display”. In: Proceedings of the EURODISPLAY Strasbourg, France

Glueck J et al (1993b) Process for manufacturing an active matrix display. US Patent 3388-60

Glueck J et al (1994) A 14-in. Diagonal A-Si TFT LCD for PAL-Television. SID Digest of Technical Papers, San José, USA

Heintz F, Knoll, PM (1982) LCD-Instrumentation in Automobiles, Proc. Soc. of Automotive Engineers (SAE) Conference, Detroit

Herzog B (2006) Laser-Projektion – die Technologie für brillante Head-up Displays (Laser projection – the technology for brilliant Head-up displays) VDI-Fachtagung „Optische Technologien in der Fahrzeugtechnik“(VDE conference „Optical technologies in automotive technology“)

Knoll PM (1986) Displays, Einführung in die Technik aktiver und passiver Anzeigen (Displays, introduction to the technology of active and passive displays). Hüthig, Heidelberg

Knoll PM (1997) Use of displays in automotive applications. J SID 5(3):165–172

Knoll PM (2007) An integrated HMI concept for driver information and driver assistance systems. SID Digest of Technical Papers, Los Angeles

Knoll PM, Elke D (1990) A conception of an integrated drivers information system. SID Digest of Technical Papers, Las Vegas, USA

Knoll PM, Reppich A (2006) A novel approach for a night vision system. In: Proceedings of the FISITA Int’l Conference, Yokohama

Knoll PM, Vollmer R (1994) Car information, communication and entertainment system – advances and new developments. In: Proceedings of the SAE international symposium, Detroit

Knoll PM et al (1993) Head-up display for automotive use. SID Digest of Technical Papers, Seattle, USA

Lueder E (1993) Ways towards brighter active addressed LC-projectors. SID Digest of Technical Papers, Seattle, USA

Mahlke S et al (2007) Evaluation of six night vision enhancement systems: qualitative and quantitative support for intelligent image processing. Hum Factors 49(3):518–531

Roscoe S (1981) Aviation psychology. University of Iowa Press, Iowa

Driver Warning Elements

36

Norbert Fecher and Jens Hoffmann

Contents

Abstract

This chapter first describes a model of human information processing and introduces the interfaces between human and machine. Requirements for warning elements are presented and so are examples of warning elements for forward and sideways guidance. Finally, a method for pregrouping the warning elements and criteria for assessing them during testing are presented.

N. Fecher (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: fecher@fzd.tu-darmstadt.de

J. Hoffmann
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: jens.hoffmann@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_37

1 Introduction

Both humans and machines may be responsible for errors especially in difficult driving conditions. Human fallibilities are due among other things to a limited ability to pay attention. If the attention is focused more on operating the navigation system than on driving the vehicle, the decisions made in emergency situations may be wrong, too late, or may not be made at all.

If a technical system is capable of identifying such emergency situations, the question is in what way the driver can be warned or in what way an intervention can take place. There are various aspects to consider when answering this question. It is important to question a warning's effectiveness when a danger exists and to assess a warning's nuisance level if triggered erroneously.

2 Human Information Processing

Very distinctive ideas have been elaborated in general psychology for the principles of human information processing. Permissible simplifications, according to (Jürgensohn and Timpe 2001), are restriction to the areas of information pickup, information processing in the narrower sense, and a system for executing action.

According to the model by Wickens (1984), human information transfer processes can be illustrated as summarized by Johannsen (1993) in Fig. 1. The model describes how stimuli are received by the sensory organs and how output variables are generated by body movements.

Input variables, which also include warning by a warning element, represent a stimulus for the human sensory organs. The intensity of the stimulus must be above a stimulus threshold specific to the sensory organs and below the pain threshold. The stimuli are received in the sensory short-term memory whose main purpose is to keep the received stimuli ready for the perception processes. Unlike when receiving stimuli, perception involves higher areas of the brain. Recognizing patterns and forming features are characteristics of perception. After perception, the human decides between possible alternative courses of action and selects a response in answer to the stimuli; at the same time, a constant exchange of information takes place between the working memory and the long-term memory. The working memory is also referred to as the short-term memory where not only the information itself but also its interpretation is stored. It takes considerably longer to save to and access the long-term memory than the short-term memory. A memory with appropriate stored body movements is also available for selection of the response. The process ends with execution of the response.

Humans have attention resources available for perceiving, deciding, and selecting responses, working memory, and response execution. These resources can be allocated at will. A detailed breakdown of human actions and human errors

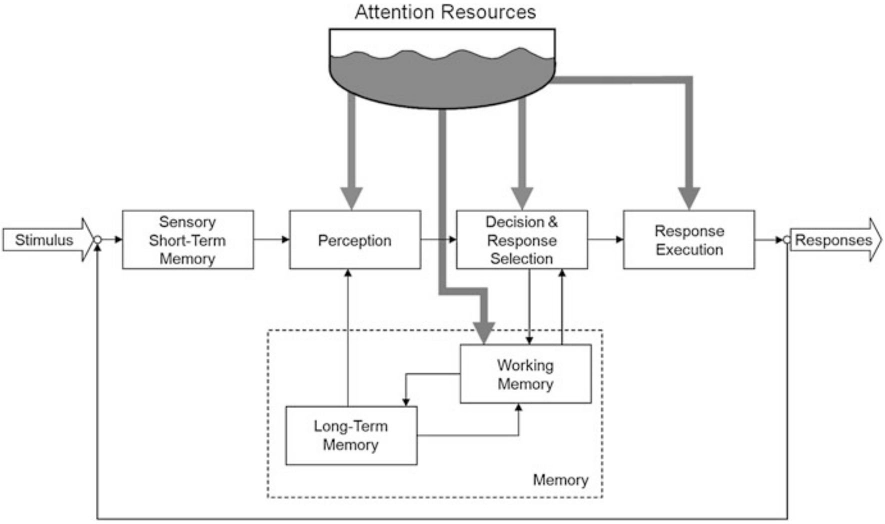


Fig. 1 Model of human information processing and distribution of attention (Jürgenson and Timpe 2001) according to Wickens (1984)

is the domain of ergonomics and is listed, for example, by Jürgensohn and Timpe (2001). According to Johannsen (1993) and Sheridan (1976), for example, humans have the ability to predict and preview when driving vehicles. In this case, current information is used to anticipate development of the traffic situation for the immediate future. If drivers classify the future situation as nonhazardous, they may tend to shift the allocation of this attention from driving the vehicle to other stimuli. If a sudden and unexpected event occurs during a moment when attention is unfavorably allocated, the result may be human errors and possibly accidents (Johannsen 1993).

Zomotor and Kiesewetter have shown that drivers do not apply the necessary brake activation speeds and forces in emergency braking situations (Zomotor 1987; Kiesewetter et al. 1997). These findings led to the development of braking assistants (BA). Rath and Knechtges described the physiological processes which take place in humans in emergency braking situations: The cerebrum is switched off due to the release of adrenaline, and the cerebellum takes over control and reacts with learned modes of action or instinctively with “fight, flight, or freeze” (Rath and Knechtges 1995). Generally, two objectives can be inferred for driver assistance systems with a warning function:

1. Assisting the human to allocate attention resources in order to prevent the collision by rapidly perceiving the traffic situation
2. Assisting the human to make the decision and select the response due to the type of warning

3 Human-Machine Interfaces

Humans receive information or stimuli via their sensory organs which have different sensitivities and operating ranges. Human sensory channels can be divided into five classes (see, e.g., the response checklist (Schmidt et al. 2000)):

1. Visual sensory channel
2. Auditory sensory channel
3. Haptic sensory channel
   (a) Tactile sensory channel
   (b) Kinesthetic-vestibular sensory channel
4. Olfactory sensory channel
5. Gustatory sensory channel

The first three sensory channels are mainly used by DAS in the vehicle. The visual and the auditory sensory channel are used in a variety of ways to transmit warnings; examples of this are shown in Sect. 5. Some of the more recent DAS specifically use the haptic sensory channel to transmit warnings. The haptic sensory channel can be divided into tactile and kinesthetic perception. Vestibular perception through the equilibrium organ in the inner ear and cerebellum is part of kinesthetic perception through the proprioceptors. The specific properties of the sensory channels are described below in relation to their use for warning elements. One of the important properties of the sensory channels for transmitting warnings is solving the question of what information should be transmitted and with what level of complexity. The transmittable information rate is a measure for this property. On the other hand, the time taken between outputting the warning of a technical system up to the beginning of perception in the human is a variable that determines the response time. This variable is referred to in the following as the perception delay time. Table 1 assigns selected criteria to the sensory channels. The sensory channels possess these properties under optimum conditions.

In addition to the properties described in the table, each sensory channel has further specific features which are explained in greater detail in Hoffmann (2008) and Schmidt et al. (2000).

4 Requirements for Warning Elements

Numerous sources exist for defining the requirements for warning elements. They are mainly familiar from the field of ergonomics. Some of the most familiar standardizations for designing driver assistance systems will be referred to in the following. No claim is made to completeness; rather, an attempt is made to provide an overview of the aspects that are to be taken into account. Guidelines for developing DAS are described in greater detail in ▶ Chap. 32, “Guidelines for

Table 1 Qualitative assessment of selected properties of the sensory channels based on Hoffmann (2008), Response-Checkliste (2006), Schmidt et al. (2000)

	Properties		
Sensory channel	Alternative name	Information rate	Perception delay time
Visual channel	Sense of sight	Very high	Fast
Auditory channel	Sense of hearing	Average	Average
Tactile channel	Sense of touch	Low	Very fast
Kinesthetic-vestibular channel	Sense of position and movement	Low	Very fast

User-Centered Development of DAS". Generally speaking, requirements for warning elements come from three areas:

1. Standards
2. Guidelines
3. Product development process

Re 1.: Standards place minimum requirements on the product. ISO 15623, among other standards, is relevant for warning elements; it is specifically aimed at collision warning systems. It explicitly defines requirements for visual and audible warnings. Other standards for auditory and visual applications are ISO CD 15006 and DIS 15008.

Re 2.: Guidelines contain requirements and specify the use of methods. Developers are advised to take existing guidelines into account when developing DAS. One such guideline, for example, is the RESPONSE checklist created as part of PReVENT (Response-Checklist 2006) in which advice is given on designing human-machine interfaces for DAS. The checklist essentially considers auditory, visual, and haptic human-machine interfaces.

Re 3.: According to VDI 2222 and VDI 2225, requirements in the product development process are generated by different methods. Unlike standards and guidelines, the product development process can be aimed specifically at use of the warning DAS. Some important general requirements for warning elements are shown by way of example in Table 2. These requirements are independent of the implementation of the warning element and the sensory channel used, and they are therefore referred to as nonfunctional requirements.

5 Examples for Warning Elements

In Sect. 2, a model for the human information process was introduced, human-machine interfaces were explained, and general requirements were defined. Several examples of warning elements are shown below, divided into warning elements for forward guidance and warning elements for sideways guidance.

Table 2 Selected requirements

Requirement	Description
Adverse effects, modalities	The requirement applicable to every warning element is that of excluding adverse effects on a human's health due to impacts. More consideration should be given to specific modalities (information rates, speed of perception, intensities, etc.)
Type and adaptation of the warning	To differentiate multiple warnings as to their type and urgency, the type of warning must be adapted to the existing danger. A collision warning is given differently to a lane departure warning. Adaptivity is required regarding the urgency of a warning so that greater urgency is achieved for greater danger
Exposure	One request in the development process may be to exclude the effect of a warning on other occupants so that the warning system is prevented from "exposing" the driver to other occupants. This exposure effect can occur in the case of both justified and unjustified warnings. For this reason, a lane departure warning is executed with an audible warning in trucks but as a seat vibration in buses (DaimlerChrysler 2000) and (Dörner 2006)

5.1 Warning Elements for Forward Guidance

Examples of warning elements for forward guidance are described for several systems currently available on the European market that warn against frontal collision.

Audi currently (2014) offers a four-stage system under the name "Audi pre sense plus" which emits audible and visual warnings in the first phase. If the driver fails to respond, the second stage alerts the driver by briefly jerking the brakes. In the third phase, the system automatically initiates partial braking to approximately 1/3 of the relevant maximum level. Pretensioning of the seat belts takes place. If there is still no response, the braking force is increased further in the fourth stage and ends in fully developed deceleration if the collision can no longer be avoided (Audi 2012).

Honda has been using the Collision Mitigation Brake System (CMBS) for the European Legend since 2006. It is available in a large number of other vehicle models in 2014 and works on a three-stage strategy. It alerts the driver audibly and visually in an early phase. If the driver brakes, the brake assistant immediately assists in adjusting the necessary braking pressure. If there is no response, a further warning is given by tensioning the seat belts; approximately 60 % deceleration takes place in the last phase (Honda 2014).

In the Lexus A-PCS (Advanced Pre-Crash Safety), the driver is initially alerted to an impending hazard by audible and visual signals. If the hazard continues to exist, the seat belt tensioners are activated, the trigger threshold for the brake assistant is adjusted, the damper control for the chassis is switched to hard, and partial deceleration of the vehicle takes place. Control of the superimposed steering is also adjusted to the situation. The driver is observed using a camera mounted on the steering column. If the processing unit identifies an inattentive driver in critical situations, the warning levels are activated at an earlier point (Toyota Deutschland 2006b).

In 2013 Mercedes introduced an enhancement of the familiar Pre-Safe Brake by adding the options "Plus" (rear-end collision protection) and "Impulse" (preconditioning of occupant protection) (Schopper et al. 2013). With the "Collision Prevention Assist Plus" assistance package, the driver is first alerted visually and audibly in several stages in the speed range between 7 and 250 km/h. Where the hazard level is higher, partial deceleration is initiated – below a speed of 105 km/h (Mercedes-Benz 2014).

In Volvo vehicles, the driver is alerted to an impending collision by a red flashing light that the instrument panel projects onto the windscreen and an audible signal (Volvo 2007).

For the automotive aftermarket, the Mobileye company offers a collision warning system (2014, model Mobileye-560) that can be retrofitted. This detects objects, such as vehicles, pedestrians, cyclists, and the lane markings, using a monocular camera. It alerts the driver visually via a separate display and audibly through additional speakers (Gat et al. 2005; Mobileye 2014).

Other driver warning elements are known from research. They include the Accelerator Force Feedback Pedal by Continental Teves: In this case, counterpressure or vibration is generated by an electric motor in addition to the passive accelerator pedal spring characteristics. When the driver depresses the accelerator pedal, increased counterpressure indicates that the distance to the vehicle in front is too short and a collision warning is transmitted by vibration (Conti 2008). Based on this principle, drivers only become aware of the warning when they operate the pedal.

Table 3 shows an overview of the sensory channels used for warnings of selected available DAS for forward guidance.

5.2 Warning Elements for Sideways Guidance

The description of warning elements for sideways guidance is aimed at the DAS currently available for warning against inadvertently wandering from the lane (referred to as Lane Departure Warning) and for lane-changing assistance (referred to as Lane Changing Decision Aid System); see ▶ Chap. 48, "Lateral Guidance Assistance". Only those systems that alert the driver to sideways guidance by means of the type of warning are described in greater detail.

One variation of tactile warning can be implemented through the steering wheel. Vibrations of the steering wheel are used, by BMW (2007) and Mercedes, for example, to warn drivers before they cross the lane markings. Although this type of warning clearly indicates the sideways guidance, it is nonspecific with regard to the direction of an impending departure from the lane. With continued deviation in relation to the center of the lane, it is also possible with the Mercedes system to brake individual wheels to guide the vehicle back into the lane. In the broadest sense, this represents a warning through the kinesthetic channel (Mercedes-Benz 2014).

Table 3 Overview of the sensory channels used for warnings of some examples of selected DAS for forward guidance in Europe (yes: present as a warning, –: not present as a warning)

	System						
Manufacturer	Audi	Honda	Lexus	Mercedes	Volvo	Mobileye	Continental Teves
System description	Pre Sense	CMBS	A-PCS	Collision prevention assist	Collision warning	560	Accelerator Force Feedback Pedal
Sensory channel							
Visual	Yes	Yes	Yes	Yes	Yes	Yes	–
Auditory	Yes	Yes	Yes	Yes	Yes	Yes	–
Haptic tactile	Yes	Yes	Yes	–	–	–	Yes
Kinesthetic	Yes	Yes	Yes	Yes	–	–	–

Fig. 2 Implementation of the visual warning of the Audi Side Assist

On introducing the current A6 in 2011, Audi offered the "Audi Active Lane Assist" system which warns drivers before they leave the lane by means of steering interventions or assists drivers to stay in lane by means of steering interventions (Freyer et al. 2011). The application of a steering torque to the steering wheel enables a warning with an indication of the direction such as Lexus has also implemented, for example (Toyota Deutschland 2006a). This form of warning, however, represents a borderline case between a warning and lane-keeping assistant (see ▶ Chap. 48, "Lateral Guidance Assistance").

With the lane change assistant, "Audi Side Assist," the traffic behind the vehicle is monitored by two RADAR sensors and a visual warning is emitted on the inside housing of the side mirrors at speeds above 30 km/h if they detect a vehicle travelling in the blind spot. However, if the driver activates the direction indicator to change lanes, the indicator becomes brighter and flashes multiple times (see Fig. 2). Similar warning elements are also used in BMW and Mercedes (BMW 2007; Mercedes-Benz 2014).

In the Mercedes trucks of the Actros series, the driver is warned by an audible rumble strip sound before an unintended lane departure. This type of warning may be described as a "representational noise" or "auditory icon." A noise is emitted from the left- or right-hand speaker depending on whether the vehicle is threatening to cross the left- or right-hand lane marking. This represents a specific directional warning which indicates sideways guidance on the basis of the formation of the representational noise (DaimlerChrysler 2000).

Citroën and Peugeot (PSA) offer the "AFIL" system in various models which warns the driver against unintended departure from the lane by means of vibrations under the side bolsters of the seat surface (Jungmann 2004). This warning element simulates tactile perception of the rumble strip sound, as a result of which the passengers in the vehicle barely notice the warning.

The Mobileye system 560 also assists in sideways guidance and emits audible and direction-oriented visual warnings (Mobileye 2014; Conti 2008).

Table 4 shows an overview of the types of warning of selected driver assistance systems for sideways guidance.

Table 4 Types of warning of DAS for sideways guidance in Europe selected as examples (D: warning element direction specific, U: warning element unspecific with regard to direction, –: not available)

	LDW						Lane changing	
Manufacturer	Audi	BMW	Mercedes	Mercedes truck	PSA	Mobileye	Audi	Mercedes
System description	Lane assist	Driving assistant	Active lane keeping assist	Telligent lane assistant	AFIL	560	Side assist	Active blind spot assist
Sensory channel								
Visual	–	–	–	–	–	D	D	D
Auditory	–	–	U	D	–	U	–	–
Haptic tactile	D	U	U	–	D	–	–	–
Kinesthetic	–	–	D	–	–	–	–	D

Studies on the effectiveness of certain warning elements in LDW systems are mainly restricted to audible and tactile warnings (Buld et al. 2005). The visual indicators of these systems highlight the system status. While it is obvious in the case of lane change assistance that a uniform warning concept has become accepted, this is still not identifiable in the LDW systems.

6 Pregrouping of Warning Elements

In the product development process (PD process) for developing warning elements, after the definition of requirements, the next step is to search for a solution while applying different methods. Typically several potential solutions then result. The aim in the further PD process is to reduce the number of variants, and to do this, suitable criteria must be available. Information content, coverage rate, and nuisance level appear to be suitable criteria for reducing the variants of driver warning elements for warning DAS. The information content of a message is a variable that specifies how much information has been transmitted in it. The coverage rate criterion is a measure for the availability of a sensory channel from the warning element to the driver as a result of which a driver is given the opportunity to respond to the warning. The nuisance level assesses the excusability of a false warning. Table 5 lists the ordinal criteria established.

One challenge when developing warning elements is specifying the timing point of the warning element before an impending collision while taking the so-called warning dilemma into account. This means that warning the driver is more effective the earlier the warning is emitted before a collision. In current systems, however, the earlier the warning is emitted, the greater the risk of a false alarm, because an environment sensing system cannot interpret the situation so accurately. On the other hand, it is expected that the fewer false alarms the system produces, the higher the chance of a warning system being accepted.

The requirement for warning DAS often runs counter to this: warning as late and as effectively as possible with a maximum coverage rate and a low nuisance level.

To allocate warning elements in the development process according to their suitability, the warning element is classified based on the criteria summarized in Table 5 and is assigned to a timing point that is early, average, or late before a collision. Each of the solution variants for a warning element is assessed by the developers involved in the PD process using the established criteria. Links between the criteria are defined.

- The lower the coverage rate of a warning element, the earlier it must be used in order to create time for other warnings or similar alerts.
- The better a warning element indicates the danger, the later it can be used because the response time is shorter.
- The lower the nuisance level of a warning element, the earlier it can be used because a false warning is less disruptive.

Table 5 Ordinal criteria for pregrouping warning elements

Information content	Nuisance level	Coverage rate
Attracts attention	Low nuisance level	High
Indicates the situation	Average nuisance level	Average
Indicates action	High nuisance level	Low

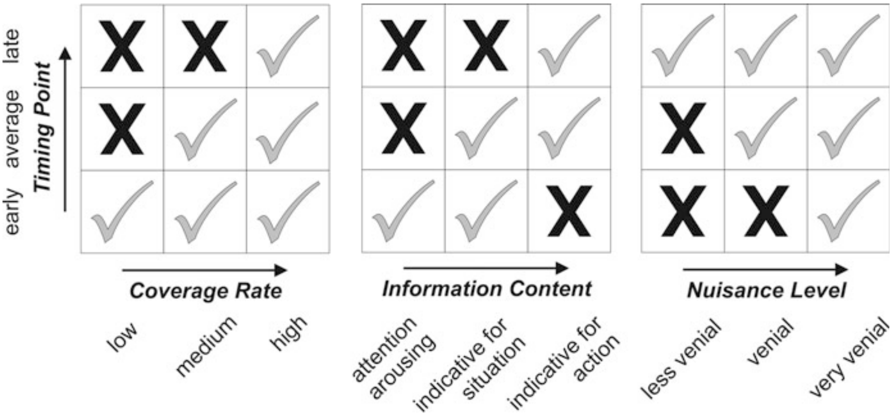


Fig. 3 Portfolio diagram (*tick*: suitable; *cross*: unsuitable)

These links are recorded in three two-dimensional matrices (portfolio diagrams); see Fig. 3; a black/white criterion (cross and tick) is available for the assessment. The timing point early, average, or late is assigned. A cross means that a driver warning element would not be appropriate at this point; the tick highlights an area that can be filled. Warning elements that indicate action cannot be used at an early point due to their high nuisance level. The three two-dimensional matrices are transferred, forming intersections with their assignment to the timing point, to a three-dimensional matrix which forms the basic shape of the compatibility matrix in Fig. 4.

Every warning element can be classified in the compatibility matrix with reference to the assessment that has taken place beforehand: It becomes obvious from the classification at which point in a collision warning the warning element can be used. A potential analysis is also carried out so that the weak points of the warning element become clear and it can be optimized in the desired direction.

Example: An auditory icon is a representational noise, such as the “screeching of tires” on full braking. The information content is indicative of a situation because drivers expect a fully braking vehicle with stationary tires in their vicinity. With a corresponding noise volume and appropriate placement of the speaker in the cockpit area, the coverage rate is high. The nuisance level is assessed as “average” because drivers may initially be irritated by a false warning; however, according to Hoffmann (2008), in over 70 % of cases, they will not respond excessively, such as by braking fully. According to the compatibility matrix, the warning element is

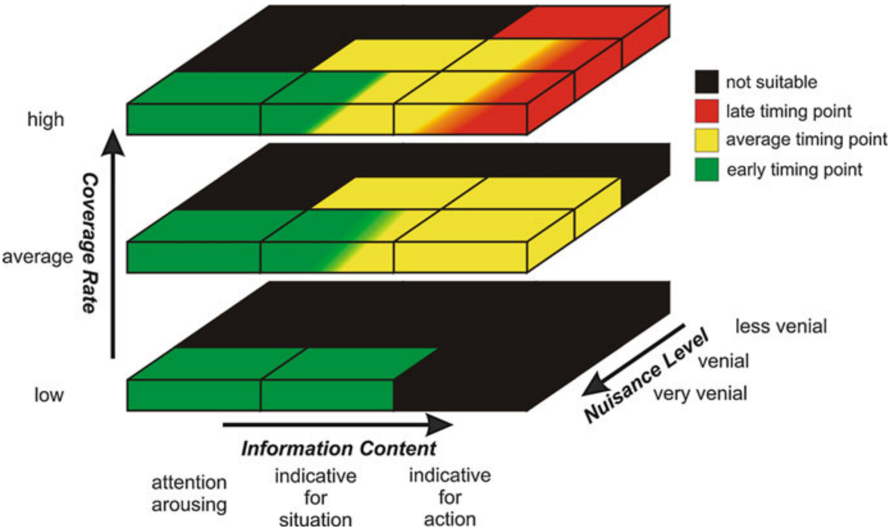


Fig. 4 Compatibility matrix for assessing the timing point of warning elements

suitable for an early to middle timing point; more detailed descriptions of auditory icons can be found in Fricke (2006) and Graham (1999).

To sum up, the compatibility matrix represents a manual tool for filtering the flood of different variations of driver warning elements, for specifying the timing point, and for determining the optimization direction.

Acknowledgment The authors would like to thank Dr. Ing. Jens Gayko for his work on the ► Chap. 36, "Driver Warning Elements". He was substantially involved in preparing and updating the chapter for the first two editions and is retiring with revising the third edition. Many thanks.

References

Audi Press Release (2012). www.audi-mediaservices.com. Accessed 08 Mar 2012

BMW Press Release (2007) Der neue BMW 5er (The new BMW 5 series). Jan 2007

Buld S, Tietze H, Krüger HP (2005) Auswirkungen von Teilautomation auf das Fahren (Effects of partial automation on driving). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung. Springer, Berlin

Conti Press Release (2008) Mit ContiGuard zum unfall- und verletzungsvermeidendem Auto (With ContiGuard to the accident an injury avoiding car). Jan 2008

DaimlerChrysler AG (2000) Hightechreport 2000: Rettendes Rattern (Rescuing judder)

Dörner K (2006) Assistenzsysteme für Nutzfahrzeuge und deren Unfallvermeidungspotential (Assistance systems for commercial vehicles and their accident avoidance potential), IAA-Symposium Entwicklungen im Gefahrgutrecht und Sicherheit von Gefahrgutfahrzeugen, 26.09.2006

Freyer J, Winkler L, Held R, Schuberth S, Khlifi R, Popken M (2011) Assistenzsysteme für die Längs- und Querführung (Assistance systems for longitudinal and lateral guidance). In: ATZ-Extra: Der neue Audi A6

Fricke N (2006) Zur Gestaltung der Semantik von Warnmeldungen (Design of the semantics of warning messages). VDI Ber 1960:133–136

FVWS ISO Norm 15623 (2013)144.19 Road vehicles – Forward vehicle collision warning system – Performance requirements and tests procedures

Gat I, Benady M, Shashua A (2005) A monocular vision advance warning system for the automotive aftermarket; SAE-2005-01-1470

Graham R (1999) Use of auditory Icons as emergency warnings. Ergonomics 42(9):1233–1236

Hoffmann J (2008) Das Darmstädter Verfahren (EVITA) zum Testen und Bewerten von Frontalkollisionsgegenmaßnahmen (The Darmstadt method (EVITA) for testing and evaluation of anti-collision systems). Fortschritt-Berichte VDI Reihe 12 Nr. 693, Düsseldorf, ISBN 978-3-18-369312-2

Honda (2014) Technical specification. http://www.honda.de/innovation/sicherheit_technik/sicherheit_kollisions-warnsystem.php. Accessed Apr 2014

ISO/CD15006-1: Auditory information presentation

ISO/DIS15008-1: Visual presentation of information

Johannsen G (1993) Mensch-Maschine-Systeme (Human-machine systems). Springer, Berlin

Jürgensohn T, Timpe KP (2001) Kraftfahrzeugführung. Springer, Berlin

Jungmann T (2004) Citroën C4, all4engineers news. 22 Nov 2004

Kiesewetter W, Klinker W, Steiner M (1997) Der neue Brake Assist von Mercedes-Benz – aktive Fahrerunterstützung in Notsituationen (The new Brake Assist by Mercedes-Benz – active driver assistance in emergency situations), ATZ

Mercedes-Benz Homepage (2014) Sicherheit – Collision prevention assist plus. http://www.mercedes-benz.de. Accessed Apr 2014

Mobileye technical specification (2014). http://de.mobileye.com. Accessed Apr 2014

Rath H, Knechtges J (1995) Effective active safety to reduce road accidents. SAE technical paper series. pp 35–42

Response-Checklist (2006) Code of practice for the design and evaluation of ADAS, PReVENT

Schmidt RF, Thews G, Lang F (2000) Physiologie des Menschen. Springer, Berlin

Schopper M, Henle L, Wohland T (2013) Intelligent drive, Vernetzte Intelligenz für mehr Sicherheit (Networked intelligence for more safety). In: ATZ-Extra: Die neue S-Klasse von Mercedes-Benz, p 107. July 2013

Sheridan TB (1976) Toward a general model of supervisory control. In: Monitoring behavior and supervisory control. Plenum, New York, pp 271–281

Toyota Deutschland GmbH (2006a) Press release lexus: Der neue Lexus LS 460. Jan 2006

Toyota Deutschland GmbH (2006b) Press release lexus: lexus Pre-Crash-Safety (PCS) in LS 460. Accessed Nov 2006

VDI 2222 (1996) und VDI 2225 (1977). www.vdi.de

Volvo (2007) Press release, Volvo mit aktivem Geschwindigkeits- und Abstandregelsystem inklusive Bremsassistent Pro (Volvo with active speed and distance control system incl. brake assist). 22 Feb 2007

Wickens CD (1984) Engineering psychology and human performance. Merrill, Columbus

Zomotor A (1987) Fahrwerktechnik – Fahrverhalten. Vogel, Würzburg

Driver Condition Detection

37

Ingmar Langer, Bettina Abendroth, and Ralph Bruder

Contents

I. Langer (✉) • B. Abendroth
Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: i.langer@iad.tu-darmstadt.de; abendroth@iad.tu-darmstadt.de

R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: bruder@iad.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_38

Abstract
This following chapter deals with driver condition detection. After delineating the factors relevant to detecting a driver's condition and discussing the reasons for addressing the subject in terms of accident risk and the corresponding potentials and challenges (Sect. 1), three potential uses of driver condition detection are examined: detection of inattentiveness (Sect. 2), detection of drowsiness (Sect. 3), and detection of medical emergencies (e.g., a heart attack; Sect. 4). The respective driver conditions are defined, relevant measuring variables and their corresponding measuring procedures present, and selected applications expanded upon. Section 5 addresses driver condition monitoring systems currently available on the market and names the measuring variables and procedures used by those systems, before giving a short overview of the problem of potential false alarms in Sect. 6.

1 Introduction and Motivation

1.1 Definition of the Term "Driver Condition"

Driver condition encompasses all driver characteristics which change over the course of time and which can be relevant to the driving task. Since a driver's condition is subject to intraindividual fluctuations, it is possible to differentiate, depending on the time period over which a change occurs, between factors influencing driver condition which change over a short period (within minutes and seconds) and over a medium period (within hours or days) (based on Kopf 2005).

For example,

- Factors which change over a medium period (days, hours)
  - Fatigue
  - Current state of health or illness
  - Daily rhythm
  - Influence of alcohol/drugs
- Factors which change over a short period (minutes, seconds)
  - Attention (e.g., selective, divided; visual, auditory)
  - Sustained attention (vigilance, wakefulness)
  - Stressors
  - Acute health problems and medical emergencies (e.g., heart attack)
  - Situational awareness
  - Emotions

Furthermore, factors which do not change or which only change on a long-term scale also affect driver condition (e.g., constitution or personality). However, these will not be further examined (for more on this, see ▶ Chap. 1, "Capabilities of Humans for Vehicle Guidance"). The following sections cover fatigue, attention, and medical emergencies in more detail.

1.2 Influence of Critical Driver Condition on Risk of Accident

A driver's condition has a strong influence on the risk of accident. Analyses of accident causes demonstrate that inattentiveness, i.e., neglect of information intake, is the chief cause of accidents. According to the analysis of Vollrath et al. (2006), 455 out of 695 accidents during turning/crossing (about 65 %) can be attributed to disregard for other road users due to inattentiveness. In the so-called 100-car study of Klauer et al. (2006), a clear connection between accidents or near-accidents due to inattentiveness and the performance of secondary tasks was identified. It has been shown that using mobile devices (e.g., mobile phones) is the most common form of secondary task and that eye diversions of more than 2 s significantly increase the risk of accident.

Other factors resulting in an increased risk of accident include fatigue (according to Platho et al. (2013), 10–20 % of traffic accidents can be attributed to fatigue behind the wheel), alcohol consumption, or driving under the influence of drugs. Klauer et al. (2006) discovered that existing fatigue increases the danger of accidents or near-accidents by a factor of four to six and leads to accidents with the most serious consequences (cf. Hargutt 2003). This is because tired drivers have difficulty accomplishing the tasks necessary for avoiding collision (braking or steering) (Jan et al. 2005). Approximately 3 % of all traffic deaths can be attributed to a medically determined incapacity on the part of the driver (Mirwaldt et al. 2013).

1.3 Potentials and Challenges in the Driver Condition Detection

Considering traits which describe driver condition allows (novel) advanced driver assistance systems (ADAS) to further build upon their already very high potential for avoiding accidents. It is conceivable, for example, that relevant system information could be transmitted in such a way that the driver could effectively perceive them independent of driver condition, e.g., in the case of inattentiveness. Similarly, warning and system intervention strategies can be adapted to driver condition, thereby increasing both the effectiveness and acceptance of driver assistance systems. It seems directly useful, for example, to warn an inattentive driver earlier or more clearly – although an early or very conspicuous warning runs the risk of the "warning dilemma" (for more on this, see ▶ Chaps. 36, "Driver Warning Elements" and ▶ 46, "Fundamentals of Collision Protection Systems").

In order to realize these potentials, it must be possible to determine driver condition. Currently, many research projects are engaged with the question of how to reliably ascertain driver condition and interpret the values determined.

The following various requirements for systems which recognize driver condition are mentioned in the literature (Knipling and Wierwille 1994; Schleicher et al. 2008, Karrer-Gauß 2011 among others):

– Unobtrusiveness of sensors through contact-free measurement
– Low rate of false alarms (see Sect. 6)

- Adequate warning and intervention strategies which, for example, motivate the driver to rest when fatigued or bring the vehicle into a minimal-risk state during a medical emergency (this would mean, at the least, stopping by the side of the road)
- Consideration of undesired behavior adaptation (cf. risk homoeostasis)

One complication is that the borders between various states are difficult to define due to large fluctuations between individuals (cf. Rauch et al. 2007). Furthermore, most sensors for monitoring driver condition require a high degree of robustness against artifacts (including movement, forces, and environmental light).

A further challenge with recognizing inattentiveness is that such a state can only be confidently identified if the attention resources necessary for a particular driving situation and the resources supplied by the driver for the task (or the control processes underlying them) are known. Since this is not possible with technical measurements, attention can only be evaluated with the help of other criteria (Blaschke 2011): for example, eye and head movements can demonstrate the driver's line of vision and thus identify potential visual inattentiveness. In order to examine the demands on attention arising from a given driving situation, it is necessary to obtain confident environment recognition and classification as well as knowledge of what level of attention is sufficient in each situation. A study by Rauch et al. (2007) also shows that the impact of disruptive factors relevant to attention is strongly situation dependent and that various indicators for recognizing the state of attention are suitable depending on the type of inattentiveness which occurs. Long-term impairments to vigilance (see also Sect. 2.1) can be recognized by continuous indicators which describe horizontal or longitudinal adjustment. Short-term distractions, on the other hand, are more readily identified by readiness to react to specific events – for example, brake reaction time to a sudden decelerating vehicle in front of the own vehicle.

Fatigue is not directly measurable but can only be quantified based on measuring aftereffects. However, aftereffects can fluctuate from person to person. In order to evaluate them, it is necessary to know the values at which a reduction of a driver's performance capacity has an impact on driving safety.

It must be noted that not all measurement categories for evaluating driver condition which will be delineated in the following sections fulfill the aforementioned demands. Although the following chapters primarily occupy themselves with methods which can be implemented using currently available sensors, they represent starting points for further research thanks to their potential for continued development.

2 Detection of Inattentiveness

2.1 Definition of Attention

Attention is commonly subdivided into the following three categories of Posner and Rafal (1987): selective attention, divided attention, and sustained attention.

Under selective attention, relevant information is selected from the environment, and irrelevant information is filtered out. Closer examination of selective attention is pertinent to a driving context, since drivers must allocate attention to all potentially relevant sources in order to process information necessary to a driving situation (Blaschke 2011). If the driver receives an influx of too much information at one time (reaching a capacity limit), there is a risk that relevant information will be perceived at a time delay or not at all.

Under divided attention, information is simultaneously received and processed, allowing the simultaneous accomplishment of various tasks (with sufficient performance among the various tasks). This requires coordination of attention distribution: for example, divided attention is required of a driver in order to visually monitor distance to a vehicle in front and follow the acoustic instructions of the navigation system. Depending on which sensory channels are simultaneously addressed, attention distribution will be more or less successful (cf. Wickens 2002).

Sustained attention – also called vigilance – describes the ability to extract relevant information from the environment over a longer period of time and react to it (cf. Posner and Rafal 1987).

These components of attention show that processing resources are limited not only with respect to scope (selection and division) but also with respect to being maintained over a long period of time (sustained attention). During vehicle guidance, most information is gathered through the visual sensory channel. All previously mentioned factors play a large role here, since the driver needs to select important information, detect relevant changes in the driving environment or in the vehicle itself (system information), while carrying out the primary driving task (distribution of attention) and remaining as attentive as possible in order to react to changes, even in time-critical situations.

Attention is often discussed in conjunction with distraction: distraction during driving is when the driver's attention is focused on an object, task, or direction not belonging to the primary task of driving. When information perception is not disrupted through distraction by other information, the term "focused attention" is also used (Schlick et al. 2010).

Inattentiveness refers to insufficient or nonexistent attention to activities which are crucial for safe driving (Regan et al. 2011; also cf. Lee et al. 2008).

2.2 Measuring Variables and Procedures for the Detection of Inattentiveness

There are various possibilities for determining a driver's level of attentiveness (Blaschke 2011):

- Detecting eye movement or head orientation via camera
- Detecting secondary activities/operational actions via vehicle sensors or cameras
- Detecting vehicle operation behavior (e.g., steering and braking patterns) via vehicle sensors

Head orientation is of limited use, since glances at an infotainment display are also possible without large movements of the head; however, detection of eye movement has strong potential to detect a driver's level of distraction.

According to Rauch et al. (2007), it is useful to distinguish between long-term (continuous) and short-term driving indicators. Long-term indicators allow decreases in vigilance to be recognized, while the current state of attentiveness can be described by means of short-term indicators.

According to Rauch et al. (2007), suitable long-term indicators include

- Tracking, above all of standard deviation in lateral position (SDLP) within the lane
- Variations in steering behavior (increase in fast, large steering movements; decrease in small corrective movements)
- Variations in distance and speed
- Length of time before speed is adjusted to external conditions

However, situational dependence must always be taken into account.

In order to detect the current state of attentiveness or short-term decreases in attention, Rauch et al. (2007) assert that it is possible to use indicators which are typically implemented as criteria in warning systems. These include, for example, TTC (time to collision), how hard the brakes are applied, or reaction time while braking. What becomes problematic is that these indicators are only active when the situation has already become critical.

In certain circumstances, the performance of a secondary activity and associated driver inattentiveness can be inferred from changes in steering movements, and it is possible to detect secondary activities – such as controlling the infotainment system – directly (Blaschke 2011).

According to Rauch et al. (2007), the repeat occurrence of longer phases without steering intervention followed by large, rapid steering motions is a sure sign of an inattentive driver (cf. also Sect. 3.2).

Sonnleitner et al. (2014) assert that alpha spindle rates from an electroencephalogram (EEG, see also Sect. 3.2) make it possible to evaluate driver distraction and to differentiate between driving with or without secondary tasks in real traffic.

2.3 Applications of Inattentiveness Detection

Inattentiveness detection can, for example, influence adaptive warning strategies – in which a warning is issued or suppressed depending on the state of attentiveness – and adjustment of warning times, depending on whether a driver is inattentive or not.

In order to monitor attention orientation, a research vehicle from Continental AG ("driver focus vehicle") was fitted with a camera on the steering column. By using an infrared camera, the driver's line of view can be detected at the highest degree of independence from environmental light conditions. In order to direct the driver's

Fig. 1 LED light strip to guide the driver's attention (Continental 2013)

attention toward a dangerous situation, Pfromm et al. (2013) describe an approach using an LED light strip (Fig. 1). Directing attention is particularly relevant if it can be determined beforehand that the driver's attention is not currently focused on the critical area.

3 Detection of Fatigue

3.1 Definition of Fatigue and Tiredness

Fatigue is generally understood to be the reversible reduction in an organ or organism's functional capacity as a consequence of activity. Fatigue can be fully reversed through recovery.

Following the modificated stress–strain concept (Rohmert 1984), fatigue can appear as a result of stressors and lead to an adjustment of human resources and capacities.

Hacker (1989) defines fatigue as a state of temporary impairment to performance conditions because of sustained demands on activity where the potential for continual restoration of performance conditions is exceeded.

Using various characteristics, the concept of fatigue can be unpacked into stress consequences which differ systematically (see Fig. 2): in the literature, the terms "fatigue," "tiredness," and "drowsiness" are rarely differentiated clearly. In this chapter, the terms will be used synonymously, since detection of drowsiness is often discussed in a driving context.

In his successive destabilization theory, Luczak (1983) defines four degrees of fatigue which enable a description of the fatiguing process. While at the first level the first scarcely noticeable disruptions in psychophysiological functions appear, disruptions at the second degree of fatigue can be observed directly by the fatigued

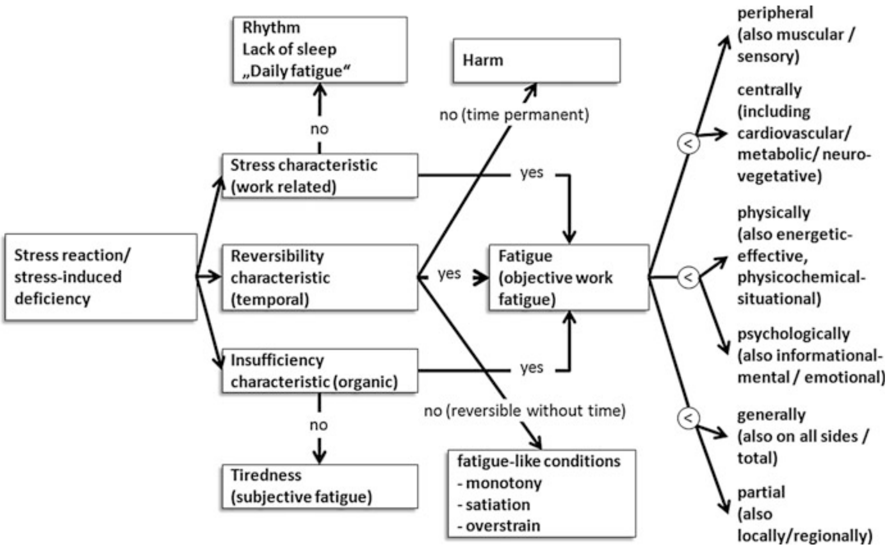


Fig. 2 Delimitations of the concept of fatigue according to Luczak (1983) (translated by the authors)

individual. Their average on the performance curve remains the same even if a high degree of performance variation occurs and the incidence of erroneous actions (e.g., driving errors) increases. At level 3 fatigue, however, performance diminishes. A further intensification to the fourth degree leads to conditions similar to exhaustion, which generally end in a refusal to perform.

This shows that fatigue is a slowly incipient process and that detection of fatigue is pertinent even at early stages of the process in order to undertake initial measures (e.g., warning the driver) during phases of fatigue which have not yet led to critical reductions in performance.

A study of truck drivers identified weakening attention and delayed reaction times to critical events as some of the consequences of fatigue (Wylie et al. 1996).

3.2 Measuring Variables and Procedures for Fatigue Detection

Identifying driver performance (via steering behavior and lane-keeping), blinking behavior (e.g., via special eye-tracking systems), EEG, and the pupillography sleepiness test are considered among the most valid possibilities for determining fatigue (cf. Platho et al. 2013). An electrocardiogram (ECG, used to measure heart rate among other factors) or subjective inquiry regarding fatigue can also be implemented. More reliable detection of fatigue is generally achieved through the combination of two or more measuring procedures.

Table 1 Descriptions of a range of potential eye-oriented measuring variables

Measuring variable	Explanation of measuring variable	Literature
Pupil diameter	Measured by pupillometry (camera-based, infrared light) fatigue can be determined through changes in diameter (frequency of pupil oscillations decreases) high susceptibility to environmental factors (brightness above all)	(Karrer-Gauß 2011), cf. (Schwalm 2009)
Eye openness	Camera-based measurement is possible distance between upper and lower eyelid (smaller with emerging fatigue)	(Jan et al. 2005), (Karrer-Gauß 2011)
Blink duration	Camera-based measurement is possible longer if fatigue is present	(Jan et al. 2005), (Schleicher et al. 2008), (Karrer-Gauß 2011)
Time delay before reopening of eyelid	Camera-based measurement is possible longer if fatigue is present	(Schleicher et al. 2008)
Blink frequency	Camera-based measurement is possible increases if fatigue is present	(Schleicher et al. 2008), (Karrer-Gauß 2011)
Blink speed	Camera-based measurement is possible becomes slower with increasing fatigue	(Platho et al. 2013), (Jan et al. 2005), (Schleicher et al. 2008), (Karrer-Gauß 2011)
PERCLOS (PERcentage of eye CLOSure)	Percentage of time for which the eyes are closed 80 % or more with regards to the space between the eyelids camera-based measurement is possible increases with fatigue, but only begins to respond at advanced stages of fatigue	(Wierwille et al. 1994), (Trutschel et al. 2011)

Indicators which make fatigue recognizable can be fundamentally subdivided into human-oriented and vehicle-oriented indicators. The following section illustrates some of the potential indicators. An overview of potential procedures for measuring fatigue and existing systems for measuring fatigue is presented in Platho et al. (2013).

Human-Oriented Measuring Variables Detection of eye activities is a valid, widely used procedure for detecting fatigue during vehicle guidance (see Table 1). Detection of blinking behavior is principally accomplished through camera-based systems or detection of glance behavior through eye-tracking systems.

Using duration of eye opening and blink duration as indicators, Hargutt (2003) identified four stages of fatigue in road trials (partially comparable with Luczak (1983), Fig. 3).

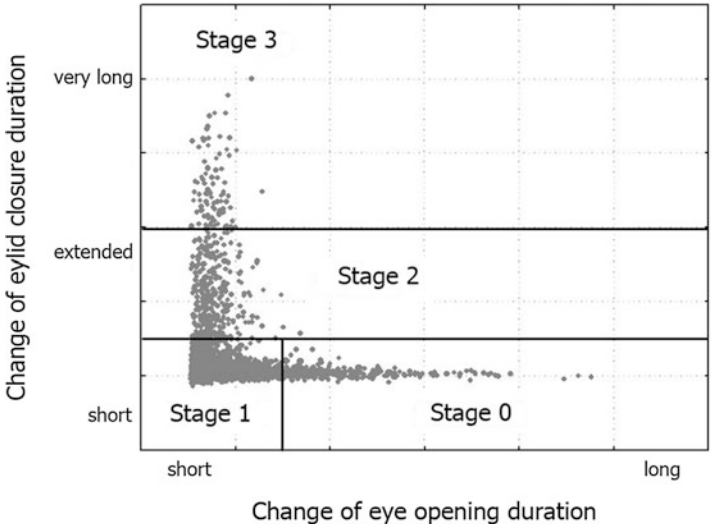


Fig. 3 Classifying stages of fatigue (Hargutt 2003) (translated by the authors)

While at Stage 1 (“decreased vigilance”) attention performance lessens, driving performance still remains unchanged. Changes are identified in secondary tasks, while the driver potentially still considers themselves absolutely wakeful. At Stage 2 (“tired”), the driver’s impaired condition affects driving performance. If fatigue continues, the driver reaches Stage 3 (“drowsy”), at which all resources are used up, making gross driving errors much more probable. At this point, at the very latest, the journey should be interrupted. Combining duration of eye opening with blink duration was determined to be superior to comparison using PERCLOS, since this affords higher sensitivity for earlier stages of fatigue (before, only ca. 40 % of earlier stages of fatigue could be resolved) and since phases on the brink of falling asleep can be identified more reliably (Hargutt 2003).

Using electrodes on the scalp, an EEG can determine changes in the frequency bands of brainwave activity, during which the incidence, duration, and amplitude of so-called alpha spindles give indications of the existing degree of fatigue (e.g., Karrer-Gauß (2011)). At the moment, however, EEG measurement does not fulfill the requirement of contact-free measurement.

Finally, heart rate, heartbeat variability, and skin conductivity can be used as measuring variables to register signs of fatigue.

Vehicle-Oriented Measuring Variables As fatigue rises, driving errors occur more frequently (see Platho et al. (2013); cf. Stage 2 fatigue (Luczak 1983)). There are many approaches to evaluating data regarding driving behavior (e.g., steering motions, speed and braking behavior, deviations from the ideal course or parameters such as TTC; Knipling and Wierwille 1994) in order to determine a driver’s level of fatigue. The advantages of detecting fatigue from driving behavior data lie in the contact-free and cost-effective recording of data. However, one disadvantage is that fatigue detection using data for horizontal vehicle guidance is difficult in city traffic due to the susceptibility to disruption based on route characteristics (cf. Platho et al. 2013).

In trials conducted by Schramm et al. (2009), during which fatigue was induced over the course of a 3-h test drive in the monotonous environment of a test facility, significant correlation was discovered between steering wheel reversal rate (SWRR, according to McLean and Hoffman (1975) the frequency at which the direction of steering was adjusted beyond a minimum angle ("gap")) and a self-evaluation. With increasing fatigue, the frequency of large steering wheel movements grows, while the total number of steering wheel movements decreases. High standard deviations appear in the results of Schramm et al. (2009), which demonstrate the presence of strong differences between individuals.

Various applications of fatigue recognition are described in Sect. 5.

4 Detection of Medical Emergencies

The performance capacity of an individual is influenced by their current health condition, among other factors. This becomes particularly relevant in the case of changes to health condition which arise suddenly during a journey – e.g., the occurrence of medical emergencies, including heart attack or stroke.

Due to changing demographics, the number of older participants in road traffic is going to increase. A rise in medically determined loss of control at the wheel can be expected. In particular, increased occurrence of cardiovascular conditions (heart attacks, for example) is to be expected (Mirwaldt et al. 2013), which can cause sudden incapacitation of the driver, often resulting in serious accidents. Monitoring health condition in order to be able to intervene during problems and, e.g., bring the vehicle to a complete stop will therefore gain in significance.

4.1 Measuring Variables and Procedures for Recognizing Medical Emergencies

Mirwaldt et al. (2012) summarize that data from ECG, plethysmography (a procedure which measures volume fluctuations of an organ or part of the body), and monitoring blood pressure can be used to identify cardiac emergencies (heart attacks and cardiac arrhythmia) and syncopes (circulatory collapse) (see also Table 2). Additionally, indicators for epilepsy and strokes are also useful, with EEG data making both of these emergencies more readily detectable. While blood sugar concentration can help to detect sugar shock, monitoring breathing can help to detect epilepsy and syncopes – at the moment, however, neither of these indicators are measurable during a journey with the current sensors.

Table 2 Measuring variables for detecting medical emergencies in vehicle (Excerpt from (Mirwaldt et al. 2012); translated by the authors): + readily detectable, (+) detectable, o less readily detectable

	Cardiac Emergencies	Epilepsy	Syncopes	Sugar shock	Stroke
Electrocardiogram (EKG)	+	(+)	+	o	(+)
Plethysmography	+	(+)	+	o	(+)
Blood pressure	+	(+)	+	o	(+)
Electroencephalogram (EEG)		+			+
Blood sugar concentration				+	
Breathing	o	(+)	(+)	o	o

Similarly, information about oxygen saturation levels in the blood, body temperature, along with the driver's position and movements are, in principle, suited to the identification of driver medical emergencies (Nguyen-Dobinsky et al. 2010).

Recent research results demonstrate which sensors are capable of detecting health indicators during a journey – here, emphasis is placed on detecting heart rate. Assessment procedures using ECG, skin conductivity, and oxygen saturation are also under discussion. Some research studies have evaluated the suitability of camera-based procedures, since they fulfill the requirement of contact-free measurement and can be combined with further applications in the vehicle (e.g., detection of fatigue or inattentiveness). Camera-based procedures can detect heart rate through changes in blood volume to blood vessels in the face (plethysmography), since this presents no limitations due to clothing. However, potential artifacts arise due to the lighting environment.

Mirwaldt et al. (2012) have determined that a color camera built into the combination instrument ensures good detectability of heart rate frequency at medium rates of artifact exposure, while other sensors (e.g., capacitive or magnetic inductive) are more susceptible to artifacts.

Even simple webcams are capable of detecting abnormalities in the heart's circulatory system by means of changes in degree of light reflection (Poh et al. 2010). The ascertained values correlate up to $r = 0.98$ with the reference value measurement via finger sensor. Even though the best results were ascertained from subjects sitting peacefully, equally good results were achieved during small movements. Problems arise in the case of large head movements and poor lighting conditions (Poh et al. 2010). In addition to heart rate, which was investigated in the study, other indicators such as heartbeat variation can be measured with this method.

In the passenger car seat developed by Eilebrecht et al. (2011) with a multichannel ECG system in the backrest, capacitive electrodes measure heart activity imperceptibly and without contact, since this can be detected from potentials on the surface of the body even through clothing. Signal quality depends on pressure applied on the seat and is thus also dependent on body weight, height, and stature. With a suitable configuration of electrodes, statistical tests can determine values for approximately 90 % of subjects. Factors which further influence signal quality include movement artifacts, which could arise during very dynamic driving behavior, and the driver's clothing.

Using a sensor unit in the steering wheel, it is possible to measure heart rate, oxygen saturation, and skin resistance (D'Angelo and Lüth 2011). With this sensor installed on the edge of the steering wheel, values were ascertained more than 81 % of the time in realistic driving tests. Over 90 % of test subjects wanted an emergency braking system which can identify a medical emergency and subsequently bring the vehicle to a safe halt.

4.2 Applications of an "Emergency Stopping Assistant"

The requirements for an automatic emergency stopping system on highways, according to Mirwaldt et al. (2012), are automatic continuation of driving and lane change execution until a risk-minimal stopping position is reached, no automatic increase in vehicle speed, strategies for warning or informing other road users, maintaining certain minimum speeds, integrating map data to determine suitable stopping possibilities, and choosing a suitable control concept in order to avoid involuntary oversteering or braking (e.g., due to unconsciousness).

Waldmann et al. (2010) describe an emergency stopping assistant which would make it possible to avoid accidents caused by health-related loss of control or to

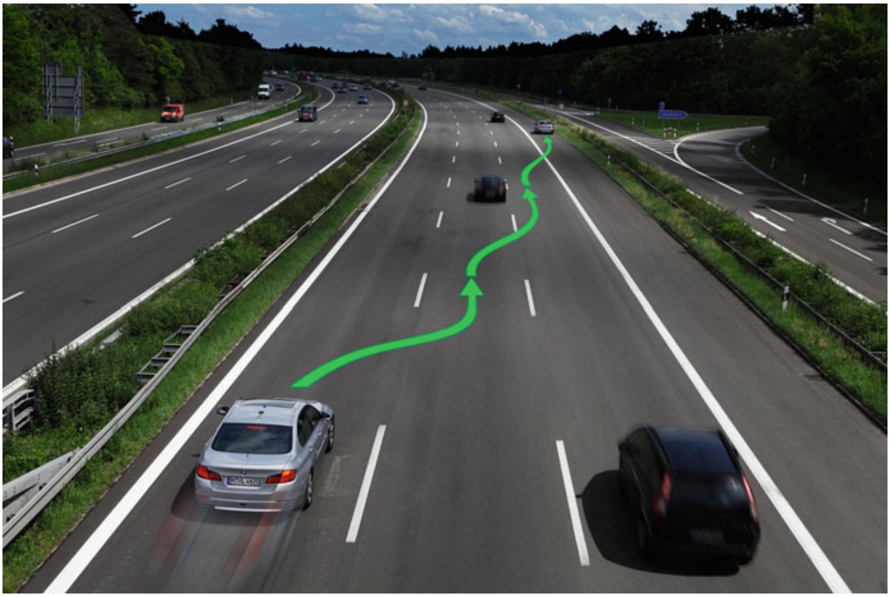


Fig. 4 Schematic execution of safe lane change (Source: BMW Group Forschung und Technik)

reduce the severity of such accidents. To this end, the assistant would bring the vehicle into a safe position in which a secured emergency stopping maneuver would be carried out, in an ideal case allowing the vehicle to come to a halt in the emergency lane of a highway (see Fig. 4). After coming to a stop, further steps such as first aid or an emergency call could be initiated (eCall, see Sect. 5). Waldmann et al. (2010) see particular challenges in the safe execution of a lane change – particularly in the case of heavy oncoming traffic.

5 Driver Condition Monitoring Systems Available on the Market

This section describes systems for monitoring inattentiveness and fatigue which are currently available on the market in current vehicle models.

In addition to these, there are further systems which react, for example, to inattentiveness but which are not assigned to monitor a driver's condition. One such system is Daimler's traffic sign assistant, which according to Missel et al. (2013) was further developed such that warnings are also issued to drivers entering the highway incorrectly ("ghost drivers"). Another example are systems which warn when drivers unintentionally leave their lane (e.g., Audi's "lane assist", see also ▶ Chap. 48, "Lateral Guidance Assistance"). Distance and collision warning systems also become active when a driver does not react because of their present condition. So-called eCall systems (e.g., BMW's "advanced/enhanced emergency call"), which become active during an accident after restraint systems (airbag, belt tensioner), are triggered – and automatically transmit data such as accident location and some data on the severity of the accident to a service center – enabling an initial evaluation of the occupant's condition.

It can be expected that in the future, more systems will refer to driver condition as a direct input value.

Attention Assist (Mercedes-Benz) This system monitors driver condition with respect to fatigue and resulting inattentiveness. According to Missel et al. (2013), drivers can always remain informed of the so-called attention level (state of attentiveness in five levels) detected by the system and begin planning a break at an early stage. If the driver is recommended to take a break, the system displays the "coffee cup symbol" (see Fig. 5) already seen in the first generation. After issuing a warning message, the navigation situation offers a rest stop search. The system is active between speeds of 60 and 200 km/h. The driver has the option of setting the system to "sensitive" mode (alternative mode: "standard"), in which the algorithm reacts more sensitively and the driver receives earlier warning.

At the start of a journey, the system creates an individual driver profile which is then continuously compared with the driver's current behavior (Schopper et al. 2011). The following indicators are consulted for recognition of increasing fatigue or inattentiveness: steering behavior, driving conditions (speed, current time, and length of journey), external influences such as crosswinds or road

Fig. 5 Various stages of attention levels depending on travel time since last break ("Pause vor") from low (*left*, "niedrig") to high (*right*, "hoch") (Missel et al. 2013)

unevenness, and behavior regarding controls (e.g., the question of whether the turn signal is activated when changing lanes).

Driver Alert (Ford) Using a front camera installed behind the inside rear-view mirror, Ford's system detects lane markings on both sides (Ford 2010). By comparing the ideal lateral position and the vehicle's current position, it is possible to infer whether the driver is fatigued, since a tired driver tends to swerve from side to side. As soon as a significant deviation is ascertained – and as long as this cannot be attributed to a lane change – a two-stage warning process is initiated. First, a warning signal is displayed on the combination instrument for 10 s, along with an auditory warning; if the driver subsequently shows further signs of fatigue, a more intrusive warning follows, which the driver must confirm by pressing a button.

Driver Alert Control (Volvo) This system analyzes how a driver proceeds between lane markings using a front camera, using a warning tone and display on the combination instrument to warn if the driver is fatigued or distracted (Lindman et al. 2012). The system compares steering behavior with previously observed patterns and recognizes fluctuations in horizontal distance from lane markings.

Driver Monitoring Camera (Toyota) and Driver Attention Monitor (Lexus) In this system, a camera installed on the steering column observes whether the driver is looking straight ahead and issues a warning if there is a threat of collision with an obstacle. Additionally, the system can provide braking support (Kurodo et al. 2009).

Fatigue Detection (Volkswagen) Volkswagen's system (in the VW Passat, for example) warns the driver by means of a display on the combination instrument and an auditory signal when fatigue is detected and a break is recommended. According to Nessenius (2010), steering angle is the most important signal for detection. Other signals such as pressing the accelerator, lateral acceleration, and the driver's activity with system controls are taken into account. The signals are compared with characteristic behavior from the beginning of the journey.

6 False Alarms and Failure to Alarm During Detection of Driver Condition

The fewer the false alarms ("false positives," i.e., the driver is not tired, but the system detects fatigue anyway) which occur in a warning system, the higher the system's degree of acceptance (cf. ▶ Chap. 36, "Driver Warning Elements"). During system design, the target conflict between false alarms and failure to alarm ("false negatives," i.e., the system detects no fatigue although the driver is tired) must be taken into account by adjusting limit values and system algorithms accordingly.

This results in the issue that although there are possibilities for evaluating a driver's condition through various measuring procedures, research studies have not reached agreement as to limit values beyond which relevant effects of driver condition on driving safety are to be expected.

Since driver condition cannot be measured directly, but can only be inferred based on indicators of condition, it is recommended that several different indicators be used to measure and evaluate in parallel, even though the necessity of different sensors results in cost disadvantages.

Trust in the system can be lost, particularly when the driver receives false feedback about potentially recognized driver conditions such as fatigue, since drivers can generally evaluate their own condition and identify respective system errors. The driver may then potentially ignore further warnings from the system.

Current measuring procedures only seldom cite convincing proof of validation, and also statements about the number of false or failed alarms are rare (Platho et al. 2013). Since driver condition recognition systems represent a clear improvement in safety, further research and development is needed here.

References

Blaschke C (2011) Fahrerzustandserkennung zur Optimierung von Spurhalteassistenzsystemen (Driver condition detection for optimization of lane keeping assistance systems). Dissertation, Universität der Bundeswehr München

Continental (2013) Continental Counts on LEDs as Co-pilot, http://www.continental-corporation.com/www/pressportal_com_en/themes/press_releases/3_automotive_group/interior/press_releases/pr_2013_02_07_driver_focus_en.html

D'Angelo LT, Lüth T (2011) Integrierte Systeme zur ablenkungsfreien Vitalparametermessung in Fahrzeugen (Integrated Systems for distraction-free measurement of vital signs in vehicles). Automobiltechnische Zeitschrift (Automotive technical review) (ATZ) 113(11):890–894

Eilebrecht B, Wartzek T, Lem J, Vogt R, Leonhardt S (2011) Kapazitives Elektrokardiogrammmesssystem im Autositz (Capacitive electrocardiogram measurement system in car seat), In: Automobiltechnische Zeitschrift (Automotive technical review) (ATZ) 113(03):232–237

Ford (2010) Ford Technology Newsbrief 08–2010: Driver Alert

Hacker W (1989) Ermüdung (Fatigue). In: Greif S, Holling H, Nicholson N (eds) Arbeits- und Organisationspsychologie: Internationales Handbuch in Schlüsselbegriffen (Industrial and Organizational Psychology: International Handbook in keywords). Psychologie Verlags Union, München

Hargutt V (2003) Das Lidschlussverhalten als Indikator für Aufmerksamkeits- und Müdigkeitsprozesse bei Arbeitshandlungen (The eyelid closure as an indicator of attention and fatigue processes at work acts), In: VDI Fortschritt-Bericht Nr. 17 (VDI progress report no. 17) (223), VDI Verlag, Düsseldorf

Jan von T, Karnahl T, Seifert K, Hilgenstock J, Zobel R (2005) Don't sleep and drive – VW's fatigue detection technology, In: Proceedings – 19th international technical conference on the enhanced safety of vehicles, National Highway Traffic Safety Administration, Washington, DC

Karrer-Gauß K (2011) Prospektive Bewertung von Systemen zur Müdigkeitserkennung: Ableitung von Gestaltungsempfehlungen zur Vermeidung von Risikokompensation aus empirischen Untersuchungen (Prospective evaluation of systems for drowsiness detection: derivation of design recommendations for the prevention of risk compensation from empirical studies), Dissertation, TU Berlin

Klauer SG, Dingus TA, Neale VL, Sudweeks JD, Ramsey DJ (2006) The impact of driver inattention on near-crash/crash risk: an analysis using the 100-car naturalistic driving study data: report no. DOT HS 810 594, National Highway Traffic Safety Administration, Washington, DC

Knipling RR, Wierwille WW (1994) Vehicle-based drowsy driver detection: current status and future prospects. In: IVHS America 4th annual meeting, Atlanta

Kopf M (2005) Was nützt es dem Fahrer, wenn Fahrerinformations- und -assistenzsysteme etwas über ihn wissen? (How helpful is it for the driver when driver information systems and driver assistance systems know anything about him?). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung (Driver assistance systems with machine perception). Springer, Berlin/Heidelberg

Kuroda K, Izumikawa I, Kouketsu O (2009) Logical mediation structures for Toyota's driver support systems. In: Proceedings der international technical conference on the enhanced safety of vehicles (ESV), Stuttgart, Germany

Lee JD, Young KL, Regan MA (2008) Defining driver distraction. In: Regan MA, Lee JD, Young KL (eds) Driver distraction – theory, effects and mitigation. CRC Press, Boca Raton

Lindman M, Kovaceva J, Levin D, Svanberg B, Jakobsson L, Wiberg H (2012) A first glance at driver alert control in FOT-data, In: Proceedings der IRCOBI Konferenz, international research council on the biomechanics of injury, Zurich

Luczak H (1983) Fatigue. In: Rohmert W, Rutenfranz J (eds) Praktische Arbeitsphysiologie (Practical Physiology). Georg Thieme Verlag, Stuttgart/New York

McLean JR, Hoffman ER (1975) Steering reversals as a measure of driver performance and steering task difficulty. Hum Factors 17(3):248–256

Mirwaldt P, Bartels A, To T-B, Braer M, Malberg H, Zaunseder S, Lemmer K (2013) Evaluation von Sensoren zur kontaktlosen Messung der Herzrate im Fahrzeug (Evaluation of sensors for contactless measurement of the heart rate in the vehicle), In: VDI-Berichte 2205: Der Fahrer im 21. Jahrhundert (VDI Reports 2205: The driver in the 21st century), VDI Verlag, Düsseldorf

Mirwaldt P, Bartels A, To T-B, Pascheka P (2012) Gestaltung eines Notfallassistenzsystems bei medizinisch bedingter Fahrunfähigkeit (Design of an emergency assistance system in medically related driving disability), In: 5. Tagung Fahrerassistenz (5th Meeting of driver assistance), München

Missel J, Mehren D, Reichmann M, Lallinger M, Bernzen W, Weikert G (2013) Intelligent Drive – Entspannter und sicherer fahren (Intelligent Drive - Driving more relaxed and more safely), In: Automobiltechnische Zeitschrift (Automotive technical review) (ATZ) extra. No. 07/2013, pp. 96–104

Nessenius D (2010) Der neue VW Passat (The new VW Passat). Automobiltechnische Zeitschrift (Automotive technical review) (ATZ) 112(12):916–925

Nguyen-Dobinsky T-N, Jacob C, Dobinsky M (2010) Mobile Notfallassistenz – Herausforderungen (Mobile Emergency Assistance - challenges), In: Proceedings zum 3. Deutschen Ambient Assisted Living-Kongress (Proceedings of the 3rd German Ambient Assisted Living Congress), VDE Verlag GmbH, Berlin, Offenbach

Pfromm M, Cieler S, Bruder R (2013) Driver assistance via optical information with spatial reference. In: Proceedings of the 16th international IEEE annual conference on intelligent transportation systems (ITSC 2013), The Hague

Platho C, Pietrek A, Kolrep H (2013) Erfassung der Fahrermüdigkeit: Berichte der Bundesanstalt für Straßenwesen – Fahrzeugtechnik Heft F 89 (Detection of driver fatigue: Reports of the Federal Highway Research Institute - Automotive booklet F 89), Fachverlag NW in der Carl Schünemann Verlag GmbH, Bremen

Poh M-Z, McDuff DJ, Picard RW (2010) Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt Express 18(10):10762–10774

Posner MI, Rafal RD (1987) Cognitive theories of attention and the rehabilitation of attentional deficits. In: Meier MJ, Benton AL, Diller L (eds) Neuropsychological rehabilitation. Churchill Livingstone, Edinburgh

Rauch N, Schoch S, Krüger H-P (2007) Ermittlung von Fahreraufmerksamkeit aus Fahrverhalten (Determination of drivers' attention handling), BMWi Project AKTIV-AS, Project FSA

Regan MA, Hallett C, Gordon CP (2011) Driver distraction and driver inattention: definition, relationship and taxonomy. Acc Anal Prevent 43(5):1771–1781

Rohmert W (1984) Das Belastungs-Beanspruchungs-Konzept (The stress–strain concept), In: Zeitschrift für Arbeitswissenschaft (Journal of Ergonomics), vol. 1984, No. 4

Schleicher R, Galley N, Briest S, Galley L (2008) Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired? Ergonomics 51(7):982–1010

Schlick CM, Bruder R, Luczak H (2010) Arbeitswissenschaft (Ergonomics), Springer, Berlin [u. a.]

Schramm T, Fuchs K, Wagner N, Bruder R (2009) Driver behaviour in a monotonous environment: a test track study, In: 16th world congress and exhibition on intelligent transport systems and services (ITS), Stockholm, Schweden

Schopper M, Mehren D, Baumann M, Köhnlein J (2011) Der beste Unfall ist der, der nicht passiert (The best accident is the one who does not happen), In: Automobiltechnische Zeitschrift (Automotive technical review) (ATZ) extra. No. 12/2011, pp 100–109

Schwalm M (2009) Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automotiven Kontext (Pupillometry as a method for detecting mental stress in the automotive context), Dissertation, Universität des Saarlandes

Sonnleitner A, Treder M, Simon M, Willmann S, Ewald A, Buchner A, Schrauf M (2014) Analysis and single-trial classification of EEG alpha spindles on prolonged brake reaction times during auditory distraction in a real road driving study. Accid Anal Prev 62(Jan):110–118

Trutschel U, Sirois B, Sommer D, Golz M, Edwards D (2011) An alertness measure of the past. In: Proceedings of the 6th international driving symposium on human factors in driver assessment, training and vehicle design, California, USA

Vollrath M, Briest S, Schießl C, Drewes J, Becker U (2006) Ableitung von Anforderungen an Fahrerassistenzsysteme aus Sicht der Verkehrssicherheit: Berichte der Bundesanstalt für Straßenwesen – Fahrzeugtechnik Heft F 60 (Derivation of requirements for driver assistance systems from the perspective of road safety: Reports of the Federal Highway Research Institute - Automotive booklet F 60), Wirtschaftsverlag NW Verlag für neue Wissenschaft GmbH, Bremerhaven

Waldmann P, Kaempchen N, Ardelt M, Homm F (2010) Der Nothalteassistent – abgesichertes Anhalten bei plötzlicher Fahrunfähigkeit des Fahrzeugführers (The Emergency Stopping assistant system – Safe stopping for sudden driving inability of the driver), In: Proceedings zum 3. Deutschen Ambient Assisted Living-Kongress (Proceedings of the 3rd German Ambient Assisted Living Congress), VDE Verlag GmbH, Berlin, Offenbach

Wickens CD (2002) Multiple resources and performance prediction. Theor Issues Ergon Sci 3 (2):159–177

Wierwille WW, Wreggit SS, Kirn CL, Ellsworth LA, Fairbanks RJ (1994) Research on vehicle-based driver status/performance monitoring, development, validation, and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration, Washington, DC

Wylie CD, Shultz T, Miller JC, Mitler MM, Mackie RR (1996) Commercial motor vehicle driver fatigue and alertness study: technical summary. Federal Highway Administration, Washington, DC

Driver Intent Inference and Risk Assessment

38

Martin Liebner and Felix Klanner

Contents

M. Liebner (✉)
Forschung und Technik, BMW Group, München, Germany
e-mail: martin.liebner@bmw.de

F. Klanner
BMW Forschung und Technik, München, Germany
e-mail: felix.klanner@bmw.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_39

Abstract

Both for driver assistance systems and highly automated driving, the in-depth understanding of traffic situations becomes more and more important. From the viewpoint of a warning driver assistance system, the authors analyze the requirements and challenges of risk assessment and driver intent inference in complex urban scenarios and provide a systematic overview of existing approaches. Furthermore, the ability of each approach to deal with more than two alternative maneuvers, partially observable feature sets, and potential interaction between traffic participants is evaluated. It is found that generative approaches and Bayesian networks in particular show great potential for driver intent inference, but it is also argued that more effort should be put into modeling the driver's situation awareness. Based on four concrete examples, the benefits of awareness-based situation analysis are demonstrated with respect to the avoidance of unnecessary warnings, the detection of occluded traffic participants, further improvement of the driver intent inference itself, as well as the prediction of the future trajectories of relevant traffic participants.

In 2013, the Federal Bureau of Statistics registered 3,340 fatalities on German roads (Statistisches Bundesamt 2014a). Even though this number is an all-time low, it is still an avoidable waste of human lives that needs to be reduced in the future. Corresponding goals are set both by the European Union (European Transport Safety Council 2013) and the German Government (Bundesministerium für Verkehr, Bau und Stadtentwicklung 2011). Next to advancements in road infrastructure and occupant protection, driver assistance systems in particular seem to be well suited to contribute to this end. While early systems such as ABS and ESC were limited to improve the controllability of the vehicle, the new generation of driver assistance systems is able to point out potential dangers in advance and thereby help to avoid critical driving situations at an early stage (Deutscher Verkehrssicherheitsrat e.V. 2009).

The huge potential of driver assistance systems is underlined by the fact that 69 % of all traffic accidents resulting in injuries occur in town and that 61 % of all people that are killed in those accidents are either pedestrians or cyclists (Statistisches Bundesamt 2014b). In contrast to occupants inside a car, there is little that can be done to protect either of them in case of a collision. Hence, the top priority must be to avoid such collisions in the first place or at least to minimize the vehicle's speed before contact.

Driver assistance systems can achieve this goal in two different ways: Either by issuing a warning to the driver or by initiating an automated braking or evasive steering maneuver. The key advantage of the latter is that the driver's reaction time does not have to be taken into account, which leaves more time to observe the situation and thereby reduces the uncertainty regarding the future trajectories of all relevant traffic participants. This, however, comes at the price of more expensive sensors as false alarms in particular might pose a serious liability problem. Therefore, most of the methods introduced in this chapter focus on long-term risk assessment that can be used to decide at an early stage whether or not the driver should be warned.

1 Guiding Research Question

In contrast to pure driving accidents such as veering off the street due to bad weather conditions or speeding, most accidents that happen in town result from the driver either not seeing or misjudging a relevant traffic participant. The situation shown in Fig. 1 is a typical example: there is a good chance that the driver of the red vehicle has overlooked the cyclist in his vehicle's blind spot.

In order to resolve potential information deficits such as the one above, driver assistance systems need to be able to detect relevant traffic participants in the vicinity of the vehicle. While current driver assistance systems are still very limited in their range of addressable situations, it is to be expected that this will change due to ongoing improvements in sensor technology as well as the introduction of various communication technologies between vehicles, infrastructure units, and central back-end servers. With enhanced environment perception capabilities, driver assistance systems will be able to anticipate an increasing number of potential threats – the question is whether or not this can be allowed to lead to an increase in the number of warnings presented to the driver.

Published results of well-known scientists in the field show that an excessive amount of available information is likely to lead to a cognitive overload of the driver and may even distract him from his actual driving task (Endsley 1995). In addition, it seems likely that the driver will feel disturbed and patronized when presented with unnecessary warnings. Therefore, driver assistance systems should be able to assess the actual collision risk and issue warnings only when necessary. In addition to the current position and velocity of all relevant traffic participants, this may require knowledge of the driver's intent and his current situation awareness.

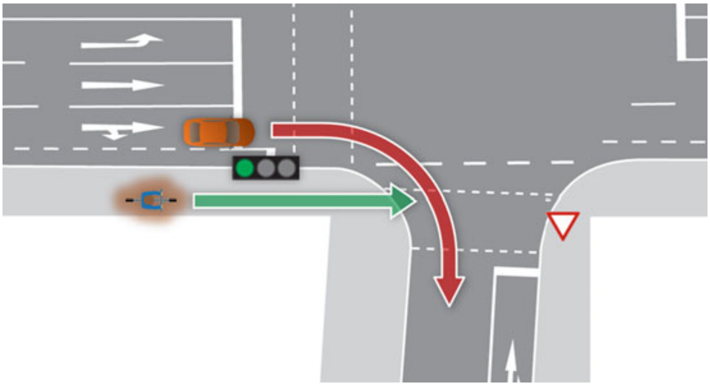


Fig. 1 Right-turn situation in an urban setting

1.1 Driver Intent Inference

From extensive driving simulator studies, it is known that warnings should be given at least 2–3 s before the actual conflict in complex inner-city situations so that the driver has time enough to process the information and initiate an appropriate action (Naujoks et al. 2012). In order to assess the collision risk prior to issuing the warning, the driver assistance system has to predict the future trajectory of all relevant traffic participants. In many situations, this is not possible based on the vehicle's dynamic state alone, as the driver's steering behavior does have a major impact on the vehicle's trajectory within this time frame. The right-turn maneuver visualized in Fig. 1 is a typical example: Even though neither the vehicle nor the cyclist shows any sign of a future conflict in their current direction of travel, the conflict is still there given the driver's intent to turn right.

Sometimes the risk of a situation does not only depend on the vehicle's own driver's intent but also on that of other traffic participants. In the situation shown on the left side of Fig. 2, for instance, the driver of the red vehicle should only be warned of the gray vehicle if the latter is not about to turn right.

Unfortunately, driver intent inference turns out to be more difficult for other traffic participants than for the own vehicle's driver as – in general – there is not

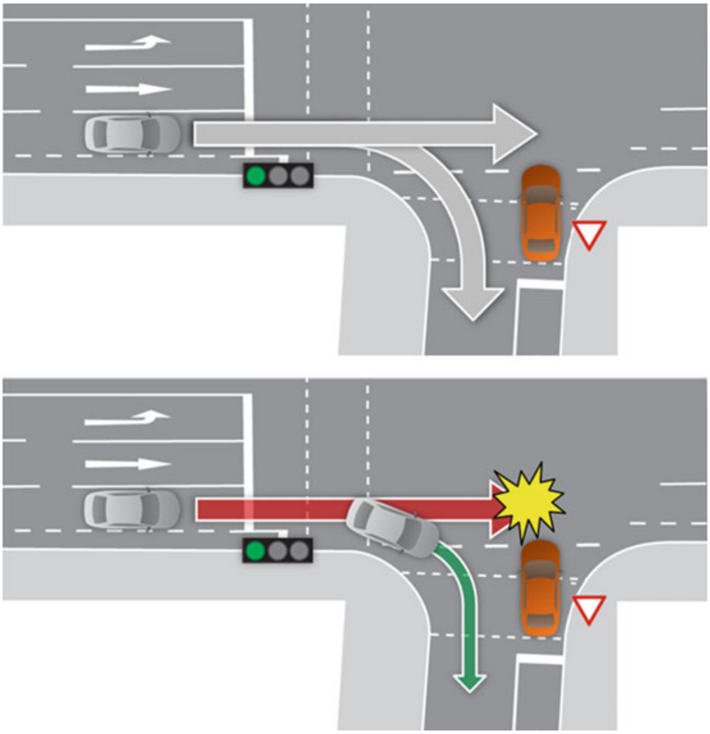


Fig. 2 Challenges for driver intent inference

nearly as much information available. In particular, it might be that the indicator status of the gray vehicle cannot be detected by the red vehicle's onboard sensors, so that driver intent inference can only rely on the gray vehicle's velocity profile. If the indicator status is known, however, the system should still be able to make use of the additional information to improve its prediction accuracy. The same holds true for information about the gray vehicle's steering angle or even its driver's gaze direction if both of them are shared via car-to-car communication. Therefore, driver intent inference algorithms should be able to deal with a variable set of available features.

In addition, the situation itself is characterized by a high degree of variability: The gray vehicle might approach the intersection at an arbitrary speed and if the driver wants to turn, the braking point will be in accordance to his typical driving behavior. The amount of deceleration needed depends both on the intersection geometry and whether or not he is required to stop at a pedestrian crossing. Also, the observed driving behavior might be caused by interaction with other traffic participants: In the situation shown on the right side of Fig. 2, the velocity profile of the gray vehicle at the rear resembles that of a right-turn maneuver even though the driver actually wants to go straight.

Even in this comparably simple situation, there are numerous factors that influence both the actual driving behavior and its implications with respect to the most probable intent. This and the fact that in most cases only a part of the driving behavior and the factors that influence it can be observed make driver intent inference an exciting but also challenging field of research.

1.2 Situation Awareness

Making predictions about the driver's intent helps to ensure that the driver is only presented with warnings of potential threats that are relevant to the path that he is about to take. Still, even these warnings can be annoying if the driver has already perceived the potential threat by himself. In principle, the situation is the same as with human codrivers: While constant advices and warnings are bothersome and – in the worst case – may even lead to critical situations caused by driver distraction, a good human codriver observes both the traffic situation and the driver and gives warnings only if the driver seems to have missed a critical aspect of the current situation that this might lead to a dangerous situation (Liebner et al. 2012a).

Besides reducing the amount of disturbance to the driver, the reliance on communication-based environment perception could be another reason to minimize the number of warnings issued by the system: As full market penetration of car-to-car communication will take at least another 20 years, there can be no guarantee that such a system will detect relevant traffic participants and issue appropriate warnings at all times. Therefore, habituation effects such as those known from parking assistance systems must be avoided at all cost. Again, this is the same as with human codrivers that may help to prevent accidents but cannot be held responsible if they fail to provide a necessary warning.

In addition to its direct application for suppressing unnecessary warnings, the driver's situation awareness is also the foundation of his interaction with other traffic participants. Examples on how this can be used to obtain a more complete representation of the vehicle's environment, to reason about the observed driver behavior, or to predict the outcome of the current traffic situation will be discussed in Sect. 6.

2 Classification of Existing Work

Being such an important and challenging field of research, numerous approaches to risk assessment of a traffic situation can already be found in literature. Therefore, we will now provide a systematic overview of the most important categories that are then discussed in detail in the following sections.

As shown in Fig. 3, the first distinction to be made is between methods that involve a simple prediction based on kinematic, dynamic, or map-based models, methods that include the driver's intent into their evaluations, as well as methods that directly obtain some kind of risk measure through knowledge-based systems. From all of those methods, risk assessment based on the driver's intent is by far the most important. As for the task of inferring the driver's intent, it can be further distinguished between discriminative and generative methods. The third level of classification introduces the categories "Hardly Any Interaction," "Limited Interaction," and "Full Interaction" that evaluate the ability of the corresponding methods to account for the interaction between traffic participants as described in Sect. 1. Besides being able to allow for an initially unknown number of influences to the driver behavior, this does also include the possibility of modeling the driver's situation awareness with respect to each of the traffic participants individually.

Finally, the lowest level of this taxonomy lists the actual methods for driving intent inference that are assigned to these categories.

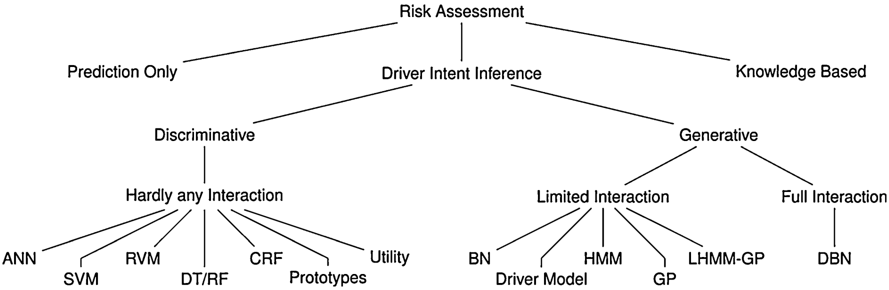


Fig. 3 Classification of existing methods for risk assessment and driver intent inference

3 Prediction-Only Methods

Methods that belong to the prediction-only category are characterized by the fact that they do not require nor obtain knowledge of the driver's intent. They can be distinguished by the kind of motion model that is employed for prediction, by their method of detecting potential collisions as well as by how they deal with uncertainties.

3.1 Motion Models

Most prediction-only methods make use of either dynamic (Brännström et al. 2010, 2013; Ammoun and Nashashibi 2009) or kinematic (Hillenbrand et al. 2006; Lytrivis et al. 2001; Tamke et al. 2011) motion models for predicting the future trajectories of all relevant traffic participants. As the driver's intent is not taken into account, the resulting collision probabilities are valid only for short prediction horizons of up to 1 or 2 s depending on the situation. Therefore, prediction-only methods are mainly used for last-second warnings as well as for systems that initiate automated braking or evasive steering maneuvers. A typical application of a dynamic motion model is visualized in Fig. 4.

In order to improve the accuracy for prediction horizons of more than 1 s, it can be beneficial to incorporate the road layout into the motion model (Eidehall and Petersson 2008; Althoff et al. 2007; Petrich et al. 2013). The additional information can be obtained either from a high-resolution digital map or from the car's onboard sensors for lane detection. Also, even though the driver's intent is not modeled explicitly, his behavior may be incorporated as a random input to the system in order to detect potential dangers in case that he or another traffic participant changes direction.

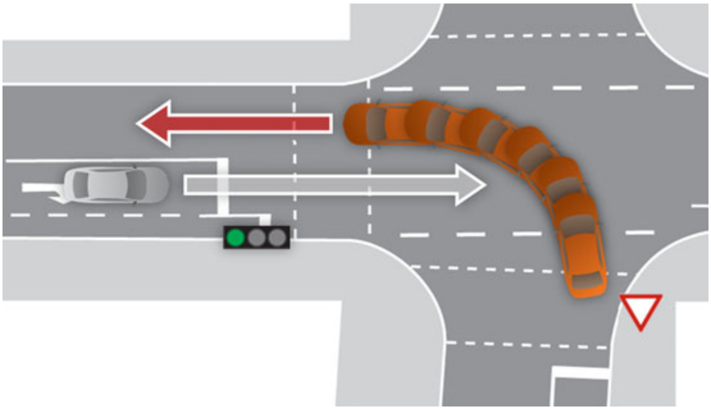


Fig. 4 Prediction of the turning vehicle based on a dynamic motion model

3.2 Collision Detection

An important part of any risk assessment algorithm is its ability to detect potential collisions between traffic participants. In most cases, a closed-form analytical solution for potential collision points such as in Brännström et al. (2010) and Hillenbrand et al. (2006) is only available for kinematic or dynamic motion models. A frequently used alternative is simulation-based solutions that predict the dynamic state of relevant traffic participants using discrete time steps (Ammoun and Nashashibi 2009; Lytrivis et al. 2001; Tamke et al. 2011). For each time step, the collision risk is then calculated based on either geometric considerations (Berthelot et al. 2011; Brännström et al. 2010; Hillenbrand et al. 2006) that include, for instance, estimating the overlap of rectangular vehicle bounding boxes or so-called conflict areas (Weidl et al. 2013) that are only allowed to be entered by one vehicle at a time.

3.3 Consideration of Uncertainties

Another key difference between existing approaches for trajectory prediction is the way in which they deal with uncertainties regarding the current dynamic state as well as the motion model itself. Some approaches neglect both kinds of uncertainties and predict the most probable trajectory only (Tamke et al. 2011; Brännström et al. 2010), while others calculate a probability distribution for the vehicle's location over time. The latter is most commonly realized using one of the following methods:

– *Assumption of normal distributions* (Ammoun and Nashashibi 2009; Berthelot et al. 2011; Brännström et al. 2013; Lytrivis et al. 2001): The main advantage of normal distributed uncertainties is that their propagation to future time steps is comparably simple. Also, the initial dynamic state of the traffic participants is likely to be available as a normal distribution already as most of the current environment perception modules rely on Kalman filters for object tracking. Nonlinearities of the dynamic model can be accounted for by methods known from the extended or unscented Kalman filter, whereas collision detection may rely on the shape of the one-sigma ellipse of the individual traffic participants (Ammoun and Nashashibi 2009). Alternatively, it is also possible to assume the minimal distance between traffic participants to be normally distributed and estimate the corresponding parameters using the unscented transformation (Berthelot et al. 2011).
– *State space discretization* (Althoff et al. 2009): By dividing the state space into small discrete cells, it is possible to propagate the probability for each cell individually and thereby account for existing nonlinearities. In particular, this allows the application of complex motion models that may even include the interaction between traffic participants as described in Sect. 1. The collision probability is estimated by aggregating the probability of each individual cell to

be occupied by more than one vehicle at once. The drawback of this method is its comparably high computational complexity or, conversely, the resulting discretization error of up to 2.5 m if real-time computation is required (Althoff and Mergel 2011).

– *Particle-based approaches* (Althoff and Mergel 2011; Broadhurst et al. 2005; Eidehall and Petersson 2008): As an alternative to a fixed discretization, the state space probability distribution can also be approximated by a number of particles that each represent one particular assignment to the dynamic state of all relevant traffic participants. After random initialization based on the original state space distribution, each particle is propagated individually according to both the motion model and its uncertainties until either a collision occurs or the maximum prediction horizon is reached. The overall risk is again calculated by the aggregation of the individual collision probabilities.

If the driver assistance system's decision depends not on the actual collision risk but rather its mere possibility, the prediction of future trajectories can be solved very efficiently either by calculating the envelope of the reachable state space (Althoff and Althoff 2009; Greene et al. 2011) or by creating an instance of a rapidly exploring random tree (Aoude et al. 2010). A typical application of such approaches is the validation of automated driving maneuvers.

4 Knowledge-Based Methods

In contrast to those of the previous section, the aim of knowledge-based methods is to estimate the collision risk based on the current situation itself. For some situations, this can be achieved by a set of deterministic rules. A typical example is that of currently available lane change assistants: As long as there is another vehicle within the driver's blind spot, the system registers a medium risk and hence triggers a warning triangle within the outer mirror. If the driver then activates the indicator to that side, the risk level is increased and additional steps like steering wheel vibrations and flashing of the triangle are initiated. Other examples include warnings of pedestrians on highways or the advice to take a coffee break if the driver seems to be tired. In general, rule-based risk assessment works best in situations where the relationship between risk and relevant context is simple.

For more complex situations, logic-based approaches as in Schwering and Lakemeyer (2013) try to infer the individual plans for relevant traffic participants in order to identify potential threats. Alternatively, a database of previously observed situations can be used to retrieve similar situations along with either the best possible system behavior or the most likely outcome of the situation (Boury-Brisset and Tourigny 2000; Vacek et al. 2007; Graf et al. 2013). A brief insight into the workings of case-based reasoning is given in Fig. 5.

A major issue with case-based reasoning is the difficulty of obtaining situations that have actually led to accidents or near misses and do therefore require an intervention of the driving assistance system. In addition, only comparably simple

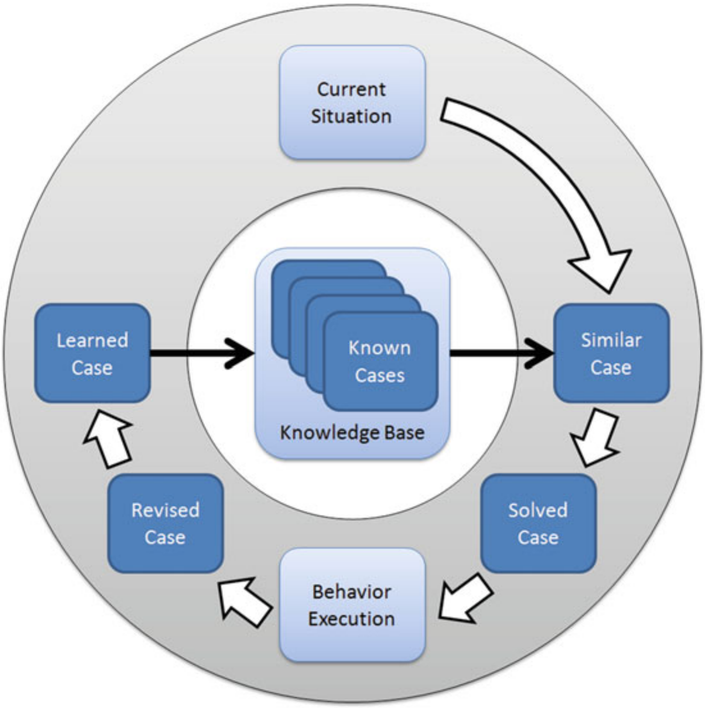


Fig. 5 Case-based reasoning (Vacek 2008). The current traffic situation is first translated into an abstract case description. Afterwards, a similar case is retrieved from the database along with information about the optimal system behavior. After execution, the behavior is evaluated, revised, and stored together with a new case if necessary

situation can be addressed because of the exponential growth of the state space. The more complex a situation, the more likely it is that the database does not contain an appropriate example. The same problem applies to the approaches proposed in Chinea and Parent (2007) and Salim et al. (2007) that try to capture typical accident scenarios by machine learning.

5 Risk Assessment Based on the Driver's Intent

Except for a couple of works that model the driver's steering input as a continuous random variable (Broadhurst et al. 2005; Eidehall and Petersson 2008), nearly all the approaches in this category aim at predicting discrete maneuvers such as lane changes (Morris et al. 2011; Gindele et al. 2010; Kasper et al. 2011; Ortiz et al. 2012), overtaking (Kretschmer et al. 2006; Firl 2011), turning (Tran and Firl 2012; Berndt and Dietmayer 2009; Klanner 2008; Lidström and Larsson 2008; Cheng and Trivedi 2006), stopping (Aoude and Desaraju 2012; Armand et al. 2013), or – as a special case of going straight – car following (McCall and

Trivedi 2007; Schneider et al. 2008). In recent years, there has also been a trend of evaluating the driver's behavior based on high-resolution digital maps (Lefèvre and Laugier 2011; Schendzielorz et al. 2013; Gindele et al. 2013; Petrich et al. 2013; Zhang and Roessler 2009; Hermann and Schroven 2012). Besides the ability to extract the current set of possible maneuvers directly from the map, this enables corresponding approaches to account for the intersection geometry's influence on the driver behavior (Liebner et al. 2013a) and to distinguish between different maneuvers of the same kind (e.g., multiple possibilities to turn right) or even combinations of subsequent driving maneuvers (Liebner et al. 2013b).

Although there are many potential features for driver intent inference, only a few can be observed from outside and therefore used to infer the intent of other traffic participants. Most commonly used are the vehicle's position, velocity, acceleration, and turn rate, as well as context information such as the distance and velocity of a preceding vehicle. In addition, it is to be expected that camera systems will soon be able to provide the indicator signal of other traffic participants. For driver intent inference within the driver's own vehicle, additional signals such as the steering wheel angle or degree of pedal activation may be used. In the past few years, a lot of effort has also been put into direct observation of the driver's head orientation, gaze direction, and foot posture that seem to have great potential for very early maneuver recognition (Doshi and Trivedi 2008; McCall et al. 2007; Liebner et al. 2012a, 2013b; Cheng and Trivedi 2006).

An overview of existing works on driver intent inference can be found in Doshi and Trivedi (2011). However, the distinction made there is mainly that of the recognized maneuvers and the features used by the individual approaches. In contrast, we will now provide a systematic overview over existing methods of driver intent inference.

5.1 Driver Intent Inference Based on Discriminative Methods

Discriminative methods are used to distinguish a fixed set of classes based on a number of observed features. In contrast to generative methods, most discriminative methods neither require nor allow prior independency assumptions as correlations between the individual features are discovered based on the training data itself. For driver intent inference, the following methods are most commonly used:

- *Artificial neural networks* (*ANNs*) (Ortiz et al. 2011): ANNs are one of the oldest methods of artificial intelligence, with the multilayer perception being the most commonly known representative. It consists of several layers of binary decision nodes that get activated if the weighted sum of all their inputs reaches a certain threshold. By cascading several layers of such nodes, even complex nonlinear decision boundaries can be represented. The weights of the individual nodes are learned (i.e., optimized with respect to the error rate) based on labeled training data. To avoid overfitting, it is important to choose the model complexity in proportion to the amount of training data.

- *Support vector machines* (*SVMs*) (Aoude and Desaraju 2012; Kumar et al. 2013; Mabuchi and Yamada 2011): Since the late 1990s, SVMs have been one of the most important methods in machine learning. Their name refers to the idea of using a distinctive subset of the training data as support vectors to represent the – in general – multidimensional and nonlinear decision boundary between two different classes. During training, which usually requires a cross-validation in order to optimize the internal parameters of the SVM, the set of support vectors is chosen by an optimization-based procedure. SVMs are quite robust against overfitting, but same as ANNs, they only allow to predict the most probable class and not the class probability itself. In order to still be able to make a probabilistic prediction of the driver's intent, a Bayesian filter may be used to estimate the class probability based on the sensitivity and specificity of the SVM as well as its results of the last N time steps (Aoude and Desaraju 2012; Kumar et al. 2013). While this procedure offers some measure of probability, it remains ignorant of the information of how close each individual result has been to the decision boundary.
- *Relevance vector machines* (*RVMs*) (McCall and Trivedi 2007; Morris et al. 2011; Cheng and Trivedi 2006): Published by Tipping (2001), the RVM is quite similar to SVMs except for the fact that it is able to provide the class probability in addition to the most probable class. The extension for distinguishing more than two classes is still an active field of research (Psorakis et al. 2010).
- *Decision trees/random forests* (*DT/RF*) (Reichel et al. 2010): Decision trees estimate the most probable class by a series of individual decisions. Starting at the root, each branch of the tree represents a decision with respect to the value of one particular feature, whereas leaf nodes represent possible classification results. The order in which the decisions have to be made is usually determined by information-theoretical measures. While decision trees are easy to read and interpret, their classification performance cannot compare to that of other methods when continuously valued features are used. This disadvantage is addressed by the so-called random forests that consist of many simple decision trees each accounting for a randomly chosen subset of the available features. The final class decision is made by a majority vote of all trees.
- *Conditional random fields* (*CRFs*) (Tran and Firl 2012): CRFs are undirected graphical models that account for correlations between individual features. In contrast to other discriminative methods, it is possible to improve the ability to generalize or – conversely – to reduce the amount of necessary training data by incorporating independency assumptions directly into the structure of the CRF. For evaluating a time series such as subsequent measurements of the vehicle's velocity, this can be a big advantage.
- *Prototype-based methods* (*prototypes*) (Hermes et al. 2009; Käfer et al. 2010): Instead of predicting abstract maneuvers, it is also possible to infer the most probable future trajectory directly. Corresponding approaches compare the current trajectory with a large number of so-called prototypes. The prototype that fits best is then used as prediction result. Similar to the knowledge-based

methods mentioned above, a major drawback of this method is that the context of the current situation – such as the intersection geometry or interaction with other traffic participants – is difficult to account for as the number of required prototypes increases exponentially with the number of influences to the driver behavior.

– *Utility-based methods* (*utility*) (Eichhorn et al. 2013): A simple yet intuitive approach to driver intent inference is to examine the plausibility of the driver's behavior with respect to his potential destinations. Based on a cost function, a planning algorithm, and a digital map, the cost of reaching each of those destinations is repeatedly evaluated in order to determine how it is changed by the current driver behavior. It is assumed that the probability of each destination is related to the corresponding cost-to-go gradient. Given that fact that a real driver is not necessarily optimal, finding an appropriate cost function seems to be a major difficulty of this approach.

As a conclusion, it seems evident that discriminative methods, given enough training data and an appropriate degree of model complexity, are well suited to predict the driver's intent in simple situations even at an early stage. A major drawback, however, is that the class probability often needs to be estimated by subsequent Bayesian filtering, which is an impediment both to risk assessment and to the combination of multiple classificators in order to distinguish between more than two different maneuvers. Also, due to the fact that discriminative methods require a fixed set of features as an input, most methods discussed above have problems if some of the features are temporarily unavailable – such as the indicator status of other vehicles or the driver's gaze direction. Another implication is that all possible influences to the driving behavior must be accounted for during training already. As the amount of required training data rises exponentially with the number of features, their capability of dealing with interaction between traffic participants appears to be very limited.

5.2 Driver Intent Inference Based on Generative Methods

Generative methods are characterized by the fact that they represent the joint probability distribution over all random variables. If a part of those random variables is observed, the posterior distribution of the other variables can be calculated using Bayes' rule. The probability of individual variables can be obtained by marginalization of the joint distribution, i.e., summing up the probabilities of all other variables.

With respect to driver intent inference, generative methods thus have the advantage of providing probabilities for all possible driver intents. Also, it does not matter if some of the features cannot be observed at all times as it is possible to calculate both prior and posterior probabilities with the same model. Finally, as each generative model represents a complete probability distribution, there are no limitations in combining them as needed. This is why they are nearly always preferred to discriminative methods if more than two different maneuvers are to be distinguished (Doshi and Trivedi 2011).

In the following, we will try to provide an overview over the most common generative methods for driver intent inference:

- *Bayesian networks* (*BNs*) (Klanner 2008; Schneider et al. 2008; Lefèvre et al. 2011; Kasper et al. 2011; Herrmann and Schroven 2012): BNs (Perl 1988) are directed acyclic graphs used to represent the structure of probabilistic models. The independency assumptions implied by that structure reduce the complexity of the model and thereby the amount of training data needed for parametrization. Although the structure itself may be determined by training data, more often it is designed according to expert knowledge so as to reflect the natural causal relationship between the individual random variables. The parametrization of the individual nodes can then be learned from training data or again be based on expert knowledge. Within the BN, each node represents self-contained knowledge of the statistical relationship between the corresponding random variable and its neighbors. In contrast to discriminative methods, it is therefore possible to reuse individual nodes for "on-the-fly" creation of new BNs as the traffic situation evolves. This fact is crucial when it comes to accounting for the interaction with an a priori unknown number of traffic participants.
- *Parametric models* (*driver model*) (Lidström and Larsson 2008; Liebner et al. 2013a, b): Typical human driving behavior has long been investigated by numerous studies in the field of human factors. The results have been used as a basis for several parametric models that describe typical driver behavior with respect to a single feature (Gipps 1981; Treiber and Helbing 2002; Land and Lee 1994), a particular type of maneuver (Rahman et al. 2013), or even as an integrated model for a wider range of situations (Hochstädter et al. 2000; Salvucci 2006). Instead of using training data, the nodes of a BN can be parametrized based on such models. In many cases, the disadvantage of simplifying the statistical relationship of the corresponding variables is by far outweighed by the benefits of using a simple, transparent, and computationally efficient model that is able to account even for complex environment conditions.
- *Hidden Markov models* (*HMMs*) (Berndt and Dietmayer 2009; Firl 2011; Meyer-Delius 2009): HMMs (Rabiner 1989) are a special case of dynamic Bayesian networks that consist of a single hidden state and an observable random variable that depends on that state. In contrast to BNs, HMMs take the temporal dependencies of subsequent time steps into account and can therefore be used for time series analysis. For driver intent inference, the first step is to obtain the prior transition and emission probabilities that best explain the training data. Based on standard optimization techniques, this step is repeated for all possible driver intents. For a new set of input data, the resulting model can then be used to determine the likelihood of each possible maneuver to have produced that time series. In order to obtain the maneuver probability, each of those likelihoods can then be combined with a prior probability for the

corresponding maneuver and – in analogy to the procedure in Liebner et al. (2013b) – even additional features from outside the HMM.

A commonly used argument against HMMs is the so-called Markov assumption that requires conditional independence between current and past observations given the system's current state. Therefore, it may be necessary to add additional complexity to the state variable of the HMM in order to approximate the persistent component of the signals that are to be observed. This disadvantage, however, is partly compensated by the fact that the calculation of the observation likelihood is in general much more efficient for HMMs than for CRFs. Regarding the interaction with other traffic participants, a much greater disadvantage of HMMs is that, similarly to discriminative methods, the model is learned as a whole. Therefore, it is not possible to adjust the model to the current situation as discussed in the context of BNs.

- *Gaussian processes* (*GPs*) (Armand et al. 2013; Tran and Firl 2013): As a generalization of the multidimensional normal distribution to an infinite number of dimensions, GPs can be used to represent temporal dependencies between subsequent observations of a multidimensional normally distributed random variable based on a mean value function m(t) and a covariance function k(t, t'). For driver intent inference, each type of maneuver is represented by its own set of parameters. After training, these can be used to determine the observation likelihood for each possible maneuver and – similar to HMMs – determine the corresponding probabilities by combining them with their priors. Same as HMMs, GPs can be used as part of a greater BN. While being well suited to capture the dynamics of continuous signals, their main disadvantage is their high computational complexity that rises cubical with the amount of training data and their inability to generalize well to situations that are not contained in it.
- *Layered HMMs based on GPs* (*LHMM-GPs*) (Laugier et al. 2011): Layered or hierarchical HMMs make use of the combinability of HMMs and GPs in order to describe the driving behavior at different levels of abstraction. In Laugier et al. (2011), the topmost layer captures the transition probabilities between individual maneuvers like going straight, turning, or overtaking. Depending on the type of maneuver, the second layer is then used to represent constitutive parts of that maneuver such as braking, driving around the corner, and accelerating. Finally, the concrete instantiation of each of those parts is taken care of by a GP on the lowest level of the model. As with BNs, the advantage of hierarchical models is that by considering smaller parts instead of holistic maneuvers, a higher degree of reusability can be achieved. Besides enabling more efficient training and a better ability to generalize to new situations, this again is a major precondition for being able to deal with the interaction between traffic participants.
- *Dynamic Bayesian networks* (*DBNs*) (Gindele et al. 2010, 2013; Dagli et al. 2003; Oliver and Pentland 2000; Lefèvre et al. 2012): By combining the reusability of BNs with the ability of HMMs to capture temporal dependencies,

DBNs can be used to infer the driver's intent, to model his situation awareness (cf. Sect. 6), and to make predictions of the future trajectory even if interaction with other traffic participants is involved. Unfortunately, such predictions are computationally very expensive and have therefore been investigated by only a small number of works (Brechtel et al. 2001). Most of them make use of particle-based methods in order to deal with the computational complexity (Dagli and Reichardt 2002). If only the current state of the driver's intent and situation are to be inferred, the driving behavior of all relevant traffic participants is likely to be observed. In this case, the resulting conditional independencies reduce the complexity of the model and may even allow methods of exact inference to be used (Lefèvre et al. 2012).

From the examples shown above, it can be seen that generative methods have great potential for modeling complex traffic situations. As this is mainly due to the reusability of generative submodels, it makes sense to further improve this feature by making use of an object-oriented design (Koller 1997; Kasper et al. 2011). A major drawback is their – compared to discriminative methods – high computational complexity that can hinder a pure data-driven design up to the point of infeasibility. For driver intent inference, however, this disadvantage is well compensated by the possibility of incorporating expert knowledge into the model.

5.3 Risk Assessment Based on the Driver's Intent

After estimating the probability distribution of the driver's intent, the collision risk and thus the need for the systems intervention can be determined either by a simple set of rules or by the predicting of the future trajectory of the vehicle. The latter usually requires a high-resolution digital map in order to determine the future path for each possible driver intent (Petrich et al. 2013; Lefèvre and Laugier 2011; Doshi and Trivedi 2008), although in simple-structured environments such motorways or highways, it may also be sufficient to rely on the vehicle's onboard sensors. Based on the road layout, the prediction itself is carried out by one of the methods introduced in Sect. 3 or – in the case of HMMs, DBNs, or parametric models – by means of the model that has been used for driver intent inference.

6 Situation Awareness and Its Applications

Most of the works available today address the situation of overlooking a relevant traffic participant. The necessity of a warning therefore depends only on the path that the driver is about to follow and whether or not a collision would occur if he did. At the same time, the driver assistance system is assumed to have a perfect knowledge of the situation – in Fig. 6 this corresponds to the lower right quadrant.

In reality, it is much more often the case that the driver has already spotted the potential threat by himself and is acting accordingly (upper half-plane) or that the

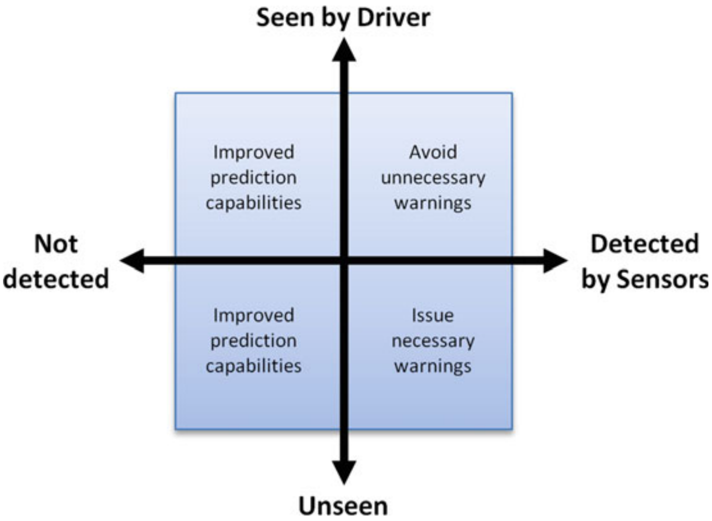


Fig. 6 Possible combinations of the system’s and the driver’s situation awareness

driver assistance system has insufficient knowledge of the vehicle’s surroundings (left half-plane). In order to obtain a better understanding of the traffic situation as well as to provide target-oriented support to the driver, it is therefore necessary to take both kinds of situation awareness into account. While a more thorough discussion of how this can be achieved is published in Liebner (2015), this section will focus on demonstrating its potential use based on four concrete examples.

6.1 Avoidance of Unnecessary Warnings

Based on the discussion in Sect. 1.2, it seems reasonable to account for the so-called inhibitors to the driver’s intent when calculating the collision risk (Schroven and Giebel 2009). In Morris et al. (2011), this is carried into effect by using the distance of following vehicles as a feature for predicting lane changes. Unfortunately, there is usually not enough data of real accidents or near misses for the parametrization of such models, so the resulting system will in effect not be able to predict critical situations if a pure data-driven approach is followed. As an alternative, the situation awareness of the driver can be captured by a separate model, parametrized using uncritical driving situations, and then used to decide whether or not the driver will carry out the intent detected by the driver intent inference method (Liebner et al. 2012a). Although this approach works well both for discriminative and generative driver intent inference methods, the former are somewhat limited when it comes to predicting the consequences of not carrying out a maneuver due to the driver’s situation awareness. For instance, it would be rather difficult for a

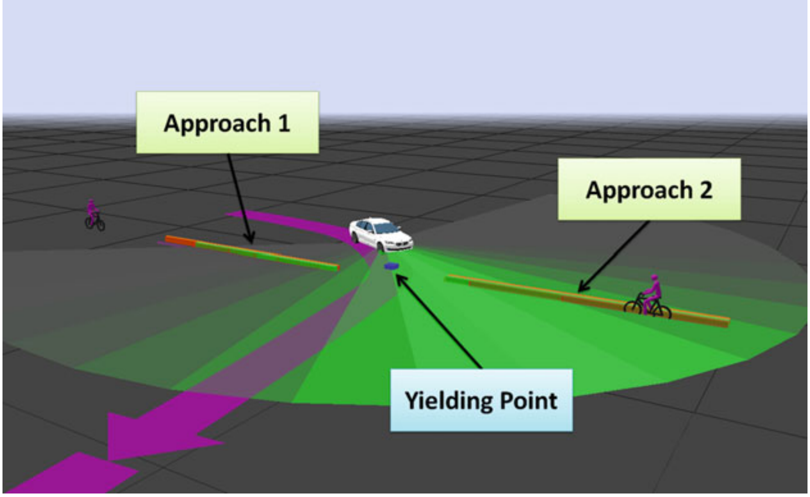


Fig. 7 Yielding point, approaches, and the driver's field of view in a right-turn situation

discriminative method to predict a lane change that has been postponed in order to let another vehicle pass.

Generative methods, on the other hand, preserve all relevant correlations when combined from submodels and are therefore much better suited to take the situation awareness into account. This is particularly true for BNs and DBNs as their high degree of modularity enables them to account for the interaction with other traffic participants. By evaluating the driver behavior, it is thus possible to infer whether or not the driver is aware of a particular traffic participant. In Liebner et al. (2012b), for instance, two different submodels for the vehicle's expected acceleration have been used in order to distinguish between whether or not the driver is aware of a cyclist moving parallel to the street. The approach could be further improved by taking the temporal dependencies into account, as the driver will most likely remember traffic participants that he has already been aware of in a previous time step. This is particularly important if the driver's gaze direction as shown in Fig. 7 is used as an input to the model.

6.2 Detection of Occluded Traffic Participants

While most existing work on situation analysis and risk assessment focuses exclusively on traffic participants that have already been detected, it is possible to reason about the existence of non-observed traffic participants by means of evaluating the behavior of those that can be observed. A typical example is shown in Fig. 8: Based on the fact that the gray vehicle is waiting in front of the cycleway, it is likely that either a pedestrian or a cyclist is approaching. Taking into account that the environment perception of the red vehicle should be able to detect either of them

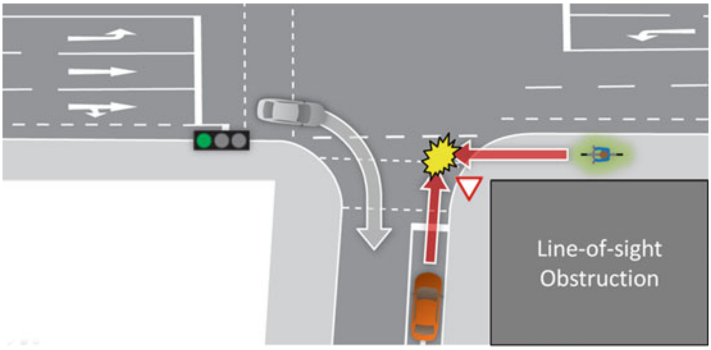


Fig. 8 Detection of an occluded cyclist based on the behavior of the gray vehicle

if approaching from the left and that it is quite unlikely that the gray vehicle is waiting for a pedestrian that is still behind the occlusion, it could even be concluded that – most likely – a cyclist is approaching from the right. Either way, a more or less specific warning could be issued to the driver of the orange vehicle when approaching the intersection.

As the cyclist is not observed and thus not available as a physical instance within the model, some other means of describing the interaction between him and the gray vehicle's driver needs to be found. One possibility is to use context information from a digital map – such as the approaches and the yielding point shown in Fig. 7. Each approach is represented by a random variable that can be either free or occupied, while the yielding point can only be passed if both of them are free.

6.3 Enhancement of Driver Intent Inference

Independently of all other features, the model of the previous section can also be used to infer the gray vehicle's future path: If it was to go straight, there would be no reason for waiting in front of the cycleway. For the orange vehicle, the situation is a bit more complicated as the deceleration of the car neither indicates a turn maneuver nor the existence of a vehicle that has the right of way. Instead, it is simply due to the fact that the driver does not know whether or not he will be required to stop at the corresponding yielding point. Therefore, besides modeling the objective and subjective occupation of each approach, it is also necessary to represent the driver's confidence in his state of knowledge.

Even though there is no immediate benefit in terms of issuing a warning, being able to explain the driver's braking by his lack of knowledge helps to avoid misinterpretations such as the conclusion that his braking is due to his wish to turn right, as might occur if the situation awareness was not to be taken into account.

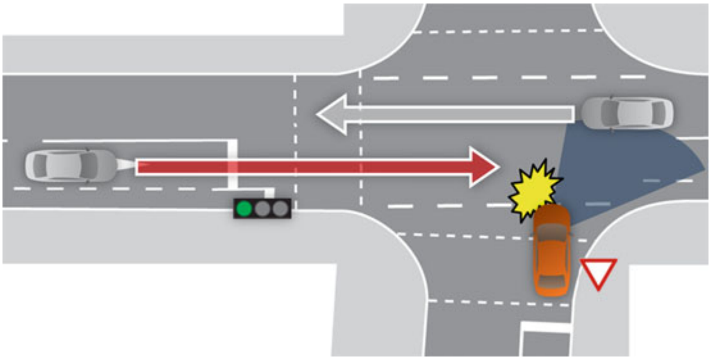


Fig. 9 Time of departure prediction based on the situation awareness of the orange vehicle's driver

6.4 Prediction of Future Trajectories

As discussed in Sect. 3, the main problem for prediction is that the number of possible future trajectories is growing exponentially over time. This is particularly true if the driver's situation awareness is to be taken into account, as his interaction with other traffic participants has a major influence on the future trajectory of the vehicle. It is thus necessary to capture not only the vehicle's dynamic state and the driver's intent but also his situation awareness at each point in time. In order to reduce the computational complexity, it may be assumed that the driver keeps his current situation awareness throughout the whole prediction horizon. This makes sense as by the time a positive risk assessment would trigger a warning, the driver should either already have the correct situation awareness or obtain it at the very latest with the help of the corresponding warning.

By incorporating the driver's situation awareness into the model, his interaction with other traffic participants can be predicted even before it is observable in his actions. A simple example is that of predicting the time of departure as in Fig. 9: While the driver of the orange vehicle is waiting for the gray vehicle approaching from the right, it is not desirable to warn him of each new vehicle approaching from the left even though – given his observed gaze direction – he is not likely to have seen it. The problem is that due to the driver's reaction time, the warning needs to be issued well before he activates the gas pedal. Thus, the only possibility to issue both an effective and target-oriented warning is to estimate the most likely time of departure based on the trajectory of the vehicle approaching from the right which – according to the driver's situation awareness – is the reason of why he is waiting.

7 Conclusion and Future Directions

With respect to the guiding research question discussed in Sect. 1, the analysis of the current state of the art leads to the following conclusions:

- In most cases, a meaningful risk assessment for driver assistance systems that aim to warn the driver of potential threats is only possible based on the driver's intent. This is due to the fact that for typical prediction horizons of 2–3 s, his steering input has a major influence on the trajectory of the vehicle. Moreover, direct knowledge-based methods seem to be unable to account for the great number of factors that influence the collision risk in urban scenarios as this would require an unrealistically large set of training data that, in addition, would need to include a currently unavailable amount of accident recordings.
- Although discriminative methods show good results for the prediction of individual driving maneuvers, they cannot be combined as well as generative models and thus have difficulties to account for situations with more than two possible maneuvers or if the interaction between traffic participants is to be taken into account. As discriminative methods require fixed feature sets, partially observable input data may pose an additional problem.
- While generative models are well suited to address even complex situations in general, limitations exist if they are learned as a whole (e.g., HMMs, GPs). Both BNs and DBNs offer great flexibility in combining submodels in accordance to the current situation which is especially important for modeling the interaction between traffic participants. Corresponding temporal dependencies are usually modeled by DBNs, although only a small number of works have attempted it so far.
- By the use of parametric driver models for on-the-fly creation of BN or DBN submodels (Liebner et al. 2013b), the context of a traffic situation can be taken into account even if it shows a high degree of variability. This includes variable initial dynamic states, intersection geometries, and behaviors of a preceding vehicle as has been identified as a problem in Fig. 2 but also enables the prediction of maneuver combinations and the distinction between two maneuvers of the same kind, e.g., if there are several possibilities to turn right.

In contrast to the driver's intent, his situation awareness has so far not been part of most investigations. This is unfortunate, as it has been shown in Sect. 6 that explicitly modeling the driver's situation awareness has great potential for the avoidance of unnecessary warnings, the detection of occluded traffic participants, the improvement of driver intent inference, and the prediction of the future trajectories of all relevant traffic participants. Especially in light of the rapid development of automated driving, the ability of such systems to understand the behavior of human drivers should be advanced by further research.

References

Althoff M, Althoff D (2009) Safety verification of autonomous vehicles for coordinated evasive maneuvers. In: Proceedings of the intelligent vehicles symposium (IV)

Althoff M, Mergel A (2011) Comparison of Markov chain abstraction and Monte Carlo simulation for the safety assessment of autonomous cars. Transact Intell Transport Syst 12(4):1237–1247

Althoff M, Stursberg O, Buss M (2007) Safety assessment of autonomous cars using verification techniques. In: Proceedings of the American control conference, pp 4154–4159

Althoff M, Stursberg O, Buss M (2009) Model-based probabilistic collision detection in autonomous driving. Transact Intell Transport Syst 10(2):299–310

Ammoun S, Nashashibi F (2009) Real time trajectory prediction for collision risk estimation between vehicles. In: Proceedings of the 5th conference on intelligent computer communication and processing (ICCP), pp 417–422

Aoude GS, Luders BD, Lee KKH, Levine DS, How JP (2010) Threat assessment design for driver assistance system at intersections. In: Proceedings of the 13th international conference on intelligent transportation systems (ITSC), pp 1855–1862

Aoude G, Desaraju V (2012) Driver behavior classification at intersections and validation on large naturalistic data set. Transact Intell Transport Syst 13(2):724–736

Armand A, Filliat D, Ibanez-Guzmán J (2013) Modelling stop intersection approaches using Gaussian processes. In: Proceedings of the 16th international conference on intelligent transportation systems (ITSC), pp 1650–1655

Berndt H, Dietmayer K (2009) Driver intention inference with vehicle onboard sensors. In: Proceedings of the international conference on vehicular electronics and safety (ICVES), pp 102–107

Berthelot A, Tamke A, Dang T, Breuel G (2011) Handling uncertainties in criticality assessment. In: Proceedings of the intelligent vehicles symposium (IV), pp 571–576

Boury-Brisset AC, Tourigny N (2000) Knowledge capitalisation through case bases and knowledge engineering for road safety analysis. Knowledge-Based Syst 13(5):297–305

Brännström M, Coelingh E, Sjoberg J (2010) Model-based threat assessment for avoiding arbitrary vehicle collisions. Transact Intell Transport Syst 11(3):658–669

Brännström M, Sandblom F, Hammarstrand L (2013) A probabilistic framework for decision-making in collision avoidance systems. Transact Intell Transport Syst 14(2):637–648

Brechtel S, Gindele T, Dillmann R (2001) Probabilistic MDP behavior planning for cars. In: Proceedings of the 14th international conference on intelligent transportation systems (ITSC), pp 1537–1542

Broadhurst A, Baker S, Kanade T (2005) Monte Carlo road safety reasoning. In: Proceedings of the intelligent vehicles symposium (IV), pp 319–324

Bundesministerium für Verkehr, Bau und Stadtentwicklung (2011) Verkehrssicherheitsprogramm 2011 vom 28.10.2011

Cheng S, Trivedi M (2006) Turn-intent analysis using body pose for intelligent driver assistance. Pervasive Comput 5(4):28–37

Chinea A, Parent M (2007) Risk assessment algorithms based on recursive neural networks. In: Proceedings of the international joint conference on neural networks, pp 1434–1440

Dagli I, Brost M, Breuel G (2003) Action recognition and prediction for driver assistance systems using dynamic belief networks. In: Agent technologies, infrastructures, tools, and applications for eservices. Springer, Berlin, pp 179–194

Dagli I, Reichardt D (2002) Motivation-based approach to behavior prediction. In: Proceedings of the intelligent vehicles symposium (IV), IEEE, pp 227–233

Deutscher Verkehrssicherheitsrat e. V. (2009) Was leisten Fahrerassistenzsysteme (how do driver assistance systems help)?

Doshi A, Trivedi M (2008) A comparative exploration of eye gaze and head motion cues for lane change intent prediction. In: Proceedings of the intelligent vehicles symposium (IV), pp 49–54

Doshi A, Trivedi M (2011) Tactical driver behavior prediction and intent inference: a review. In: Proceedings of the 14th international conference on intelligent transportation systems (ITSC), pp 1892–1897

Eichhorn A, Werling M, Zahn P, Schramm D (2013) Maneuver prediction at intersections using cost-to-go gradients. In: Proceedings of the international conference on intelligent transportation systems (ITSC), pp 112–117

Eidehall A, Petersson L (2008) Statistical threat assessment for general road scenes using Monte Carlo sampling. Transact Intell Transport Syst 9(1):137–147

Endsley MR (1995) Toward a theory of situation awareness in dynamic systems. Human Factors 37:32–64

European Transport Safety Council (2013) Road safety manifesto for the European parliament elections May 2014

Firl J (2011) Probabilistic maneuver prediction in traffic scenarios. In: Proceedings of the European conference on mobile robots (ECMR)

Gindele T, Brechtel S, Dillmann R (2010) A probabilistic model for estimating driver behaviors and vehicle trajectories in traffic environments. In: Proceedings of the 13th international conference on intelligent transportation systems (ITSC), pp 1625–1631

Gindele T, Brechtel S, Dillmann R (2013) Learning context sensitive behavior models from observations for predicting traffic situations. In: Proceedings of the 16th International conference on intelligent transportation systems (ITSC), pp 1764–1771

Gipps P (1981) A behavioural car-following model for computer simulation. Transportation Research Part B: Methodological, vol. I, no. 2, pp 105–111

Graf R, Deusch H, Fritzsche M, Dietmayer K (2013) A learning concept for behavior prediction in traffic situations. In: Proceedings of the intelligent vehicles symposium (IV), pp 672–677

Greene D, Liu J, Reich J, Hirokawa Y, Shinagawa A, Ito H, Mikami T (2011) An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists. Transact Intell Transport Syst 12(4):942–953

Hermes C, Wöhler C, Schenk K, Kummert F (2009) Long-term vehicle motion prediction. In: Proceedings of the intelligent vehicles symposium (IV), pp 652–657

Herrmann S, Schroven F (2012) Situation analysis for driver assistance systems at urban intersections. In: Proceedings of the international conference on vehicular electronics and safety (ICVES), pp 151–156

Hillenbrand J, Spieker AM, Kroschel K (2006) A multilevel collision mitigation approach – its situation assessment, decision making, and performance tradeoffs. Transact Intell Transport Syst 7(4):528–540

Hochstädter A, Zahn P, Breuer K (2000) Ein universelles Fahrermodell mit den Einsatzbeispielen Verkehrssimulation und Fahrsimulator (A universal driver model with the applications in traffic simulation and driving simulators). Fahrzeug- und Motorentechnik, 9. Aachener Kolloquim

Käfer E, Hermes C, Wöhler C, Ritter H, Kummert F (2010) Recognition of situation classes at road intersections. In: Proceedings of the international conference on robotics and automation (ICRA), pp 3960–3965

Kasper D, Weidl G, Dang T, Breuel G, Tamke A, Rosenstiel W (2011) Object-oriented bayesian networks for detection of lane change maneuvers. In: Proceedings of the intelligent vehicles symposium (IV), pp 673–678

Klanner F (2008) Entwicklung eines kommunikationsbasierten Querverkehrsassistenten im Fahrzeug (Development of a communication-based onboard intersection assistent). Dissertation, Technische Universität, Darmstadt

Koller D (1997) Object-oriented Bayesian networks. In: Proceedings of the 13th annual conference on uncertainty in artificial intelligence (UAI), pp 302–313

Kretschmer M, König L, Neubeck J, Wiedemann J (2006) Erkennung und Prädiktion des Fahrerverhaltens während eines Überholvorgangs (Recognition and prediction oft he driver behavior during overtaking). In: Proceedings of the 2. Tagung Aktive Sicherheit durch Fahrerassistenz, Garching

Kumar P, Perrollaz M, Laugier C (2013) Learning-based approach for online lane change intention prediction. In; Proceedings of the Intelligent Vehicles Symposium (IV)

Land M, Lee D (1994) Where do we look when we steer. Nature 369:742–744

Laugier C, Paromtchik IE, Perrollaz M, Yong M, Yoder JD, Tay C, Mekhnacha K, Nègre A (2011) Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety. ITSM, October 2011, pp 4–19

Lefèvre S, Ibañez-Guzmán J, Laugier C (2011) Context-based estimation of driver intent at road intersections. In: Proceedings of the symposium on computational intelligence in vehicles and transportation systems, pp 583–588

Lefèvre S, Laugier C (2011) Exploiting map information for driver intention estimation at road intersections. In: Proceedings of the intelligent vehicles symposium (IV), pp 583–588

Lefèvre S, Laugier C, Ibañez-Guzmàn J (2012) Risk assessment at road intersections: comparing intention and expectation. In: Proceedings of the intelligent vehicles symposium (IV), pp 165–171

Lidström K, Larsson T (2008) Model-based estimation of driver intentions using particle filtering. In: Proceedings of the 11th International Conference on intelligent transportation systems (ITSC), pp 1177–1182

Liebner M, Klanner F, Stiller C (2012a) Der Fahrer im Mittelpunkt – Eye Tracking als Schlüssel zum mitdenkenden Fahrzeug (Focussing the driver – eye tracking as a key to intelligent vehicles)? In: Proceedings of the 8th workshop Fahrerassistenzsysteme, UniDAS e.V., Walting, pp 87–96

Liebner M, Baumann M, Klanner F, Stiller C (2012b) Driver intent inference at urban intersections using the intelligent driver model. In: Proceedings of the intelligent vehicles symposium (IV), pp 1162–1167

Liebner M, Klanner F, Baumann M, Ruhhammer C, Stiller C (2013a) Velocity-based driver intent inference at urban intersections in the presence of preceding vehicles. ITSM 5(2):10–21

Liebner M, Ruhhammer C, Klanner F, Stiller C (2013b) Generic driver intent inference based on parametric models. In: Proceedings of the 16th international conference on intelligent transportation systems (ITSC), pp 268–275

Liebner M (2015), Fahrerabsichtungserkennung und Risikobewertung für warnende Fahrerassistenzsysteme. Dissertation, Karlsruher Institut für Technologie

Lytrivis P, Thomaidis G, Tsogas M, Amditis A (2001) An advanced cooperative path prediction algorithm for safety applications in vehicular networks. Trans Intell Trans Syst 12(3):669–679

Mabuchi R, Yamada K (2011) Study on driver-intent estimation at yellow traffic signal by using driving simulator. In: Proceedings of the intelligent vehicles symposium (IV)

McCall J, Trivedi M (2007) Driver behavior and situation aware brake assistance for intelligent vehicles. Proceedings of the IEEE 95(2):374–387

McCall J, Wipf D, Trivedi M, Rao B (2007) Lane change intent analysis using robust operators and sparse bayesian learning. Trans Intell Trans Syst 8(3):431–440

Meyer-Delius D (2009) Probabilistic situation recognition for vehicular traffic scenarios. In: Proceedings of the international conference on robotics and automation (ICRA), pp 459–464

Morris B, Doshi A, Trivedi M (2011) Lane change intent prediction for driver assistance: on-road design and evaluation. In: Proceedings of the intelligent vehicles symposium (IV), pp 895–901

Naujoks F, Grattenthaler H, Neukum A, (2012) Zeitliche Gestaltung effektiver Fahrerinformationen zur Kollisionsvermeidung auf der Basis kooperativer Perzeption (Timing of effective driver information for collision avoidance based on cooperative perception). Proceedings of the 8th Workshop Fahrerassistenzsysteme, UniDAS e.V., Walting, pp 107–117

Oliver N, Pentland A (2000) Graphical models for driver behavior recognition in a SmartCar. In: Proceedings of the intelligent vehicles symposium (IV), pp 7–12

Ortiz G, Fritsch J, Kummert F, Gepperth A (2011) Behavior prediction at multiple time-scales in inner-city scenarios. In: Proceedings of the intelligent vehicles symposium (IV), pp 1066–1071

Ortiz G, Kummert F, Schmudderich J (2012) Prediction of driver behavior on a limited sensory setting. In: Proceedings of the 15th international conference on intelligent transportation systems (ITSC), pp 638–643

Perl J (1988) Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann Publishers, San Francisco

Petrich D, Dang T, Kasper D, Breuel G, Stiller C (2013) Map-based long term motion prediction for vehicles in traffic environments. In: Proceedings of the 16th international conference on intelligent transportation systems (ITSC), pp. 2166–2172

Psorakis I, Damoulas T, Girolami MA (2010) Multiclass relevance vector machines: sparsity and accuracy. Trans Neural 21(10):1588–98

Rabiner L (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286

Rahman M, Chowdhury M, Xie Y, He Y (2013) Review of microscopic lane-changing models and future research opportunities. Trans Intell Trans Syst 14(4):1942–1956

Reichel M, Botsch M, Rauschecker R, Siedersberger KH, Maurer M (2010) Situation aspect modelling and classification using the scenario based random forest algorithm for convoy merging situations. In: Proceedings of the 13th international conference on intelligent transportation systems (ITSC), pp 360–366

Salim FD, Loke SW, Rakotonirainy A, Srinivasan B, Krishnaswamy S (2007) Collision pattern modeling and real-time collision detection at road intersections. In: Proceedings of the intelligent transportation systems conference, pp 161–166

Salvucci DD (2006) Modeling driver behavior in a cognitive architecture. Hum Factors 48(2):362–80

Schendzielorz T, Mathias P, Busch F (2013) Infrastructure-based vehicle maneuver estimation at urban intersections. In: Proceedings of the 16th international conference on intelligent transportation systems (ITSC), pp 1442–1447

Schneider J, Wilde A, Naab K (2008) Probabilistic approach for modeling and identifying driving situations. In: Proceedings of the intelligent vehicles symposium (IV), pp 343–348

Schroven F, Giebel T (2009) Fahrerintentionserkennung für Fahrerassistenzsysteme. VDI-FVT-Jahrbuch, pp 54–58

Schwering C, Lakemeyer G (2013) Spatio-temporal reasoning about traffic scenarios. In: Proceedings of the 11th international symposium on logical formalizations of commonsense reasoning

Statistisches Bundesamt (2014a) 7.2% weniger Todesopfer auf deutschen Straßen im Jahr 2013 (7.2% less fatalities on German roads in 2013)

Statistisches Bundesamt (2014b) Verkehrsunfälle Fachserie 8 Reihe 7 Jahr 2013 (road accidents edn. 8.7/2013).

Tamke A, Dang T, Breuel G (2011) A flexible method for criticality assessment in driver assistance systems. In: Proceedings of the intelligent vehicles symposium (IV), pp 697–702

Tipping M (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244

Tran Q, Firl J (2012) A probabilistic discriminative approach for situation recognition in traffic scenarios. In: Proceedings of the intelligent vehicles symposium (IV), pp 147–152

Tran Q, Firl J (2013) Modelling of traffic situations at urban intersections with probabilistic non-parametric regression. In: Proceedings of the intelligent vehicles symposium (IV), pp 334–339

Treiber M, Helbing D (2002) Realistische Mikrosimulation von Straßenverkehr mit einem einfachen Modell (Realistic micro-simulation of road traffic with a simple model). 16. Symposium Simulationstechnik ASIM, p 80

Vacek S (2008) Videogestützte Umfelderfassung zur Interpretation von Verkehrssituationen für kognitive Automobile (video supported environment perception for interpretation of traffic situations for cognitive automobiles). Dissertation, Universität Karlsruhe (TH)

Vacek S, Gindele T, Zöllner JM, Dillmann R (2007) Situation classification for cognitive automobiles using case-based reasoning. In: Proceedings of the intelligent vehicles symposium (IV), pp 704–709

Weidl G, Singhal V, Petrich D, Kasper D, Wedel A, Breuel G (2013) Collision detection and warning at road intersections using an object oriented bayesian network. In: Proceedings of the 9th international conference on intelligent computer communication and processing (ICCP)

Zhang J, Roessler B (2009) Situation analysis and adaptive risk assessment for intersection safety systems in advanced assisted driving. In: Dillmann R, Beyerer J, Stiller C, Zöllner JM, Gindele T (eds) Autonome mobile systeme. Springer, Berlin, pp 249–258

Part VIII

DAS on Stabilisation Level

Brake-Based Assistance Functions

39

Anton van Zanten and Friedrich Kost

Contents

Anton van Zanten is retired (formerly Robert Bosch GmbH)

A. van Zanten (✉)
Ditzingen, Germany
e-mail: anton.vanzanten@web.de

F. Kost
Robert Bosch GmbH, Abstatt, Germany
e-mail: frieder.kost@de.bosch.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_40

Abstract

Many driver assistance systems (DAS) use the electronic stability control (ESC) system for their control tasks, while ESC itself uses the antilock brake system (ABS) and traction control system (ASR, TCS) for the control of the lateral dynamics of the vehicle. This chapter starts with the description of the systems ABS and ASR as they are used by ESC. Then the description of the system ESC as is used by DAS follows. At the end of the chapter, the brake-based assistance functions as used by DAS are described.

1 Introduction

In daily traffic, the longitudinal and lateral accelerations of a vehicle are seldom larger than 0.3 g. Therefore, the absolute value of the tire slip is seldom larger than 2 %, while the absolute values of the slip angles of the tires and the vehicle are seldom larger than 2°. Within those values, the tires and the vehicle behave more or less in a linear manner. Experience of most drivers in handling is thus limited to the linear behavior of the vehicle. If a vehicle approaches the physical limit between the tires and the road, its behavior becomes highly nonlinear. In those situations, most drivers are not able to handle the vehicle in a safe manner. Moreover, if the wheels lock during braking or spin during traction, then the driver is not able to influence the vehicle motion anymore and control over the vehicle is lost. If, e.g., the rear wheels lock before the front wheels, then the vehicle may skid (Fig. 1). Control systems which control the wheel rotation help the driver to keep the vehicle under control.

These control systems are the antilock brake system (ABS) which keeps the wheels from locking, the traction slip control system (TCS or ASR) which keeps the wheels from spinning, and the electronic stability control (ESC) which keeps the vehicle from skidding and leaving the turn. Since these control systems help the tires and the vehicle to behave in a predictable manner, they may be seen as vehicle assistance systems rather than driver assistance systems. Driver assistance systems may help the driver with his tasks to steer, accelerate, and decelerate the vehicle and to coordinate these tasks.

2 Fundamentals of Vehicle Dynamics

2.1 Stationary and Transient Behavior of Tires and Vehicles

This section deals with the handling of the vehicle in the linear and nonlinear region however not in full detail but only as is required to understand the control systems. The vehicle motion is mainly determined by the forces between the tires and the road. Therefore, understanding the tire behavior is a prerequisite for understanding

Fig. 1 Skidding car on a dry asphalt road

the vehicle behavior. Since transients of the vehicle motion have a main influence on the control system, the transient behavior of the vehicle will also be discussed.

If a wheel is not steered and neither braked nor driven, then its rotational velocity, called the free rolling rotational velocity ω_{WhlFre}, can be computed from the vehicle velocity v_{v}, $\omega_{\mathrm{WhlFre}} = v_{\mathrm{v}}/r$ where r is the radius of the wheel. If the wheel is braked by a brake torque M_{BR}, then the rotational velocity ω_{Whl} will be smaller than the free rolling rotational velocity ω_{WhlFre}. In the description of the control systems, the wheel velocity is used instead of the rotational velocity. The wheel velocity is defined as the product of the rotational wheel velocity and the wheel radius r. The free rolling wheel velocity is then $v_{\mathrm{WhlFre}} = \omega_{\mathrm{WhlFre}} \cdot r$, while the wheel velocity is $v_{\mathrm{Whl}} = \omega_{\mathrm{Whl}} \cdot r$.

The difference between the free rolling and the braked wheel velocity is called the slip velocity. The slip velocity divided by the free rolling wheel velocity is called the wheel or tire slip λ. It is a dimensionless quantity whose value is one if the wheel is locked. Often, the slip is also expressed in percent, and the slip is 100 % if the wheel locks.

$$\lambda = \frac{v_{\mathrm{WhlFre}} - v_{\mathrm{Whl}}}{v_{\mathrm{WhlFre}}} \tag{1}$$

If a brake torque M_{B} acts on the wheel, then a brake force F_{B} between the tire and the road results. If the normal force on the tire is F_{N}, then the brake coefficient of friction between the tire and the road is $\mu_{\mathrm{B}} = F_{\mathrm{B}}/F_{\mathrm{N}}$. The road torque on the wheel M_{R} is defined as the product of the brake force F_{B} and the wheel radius r,

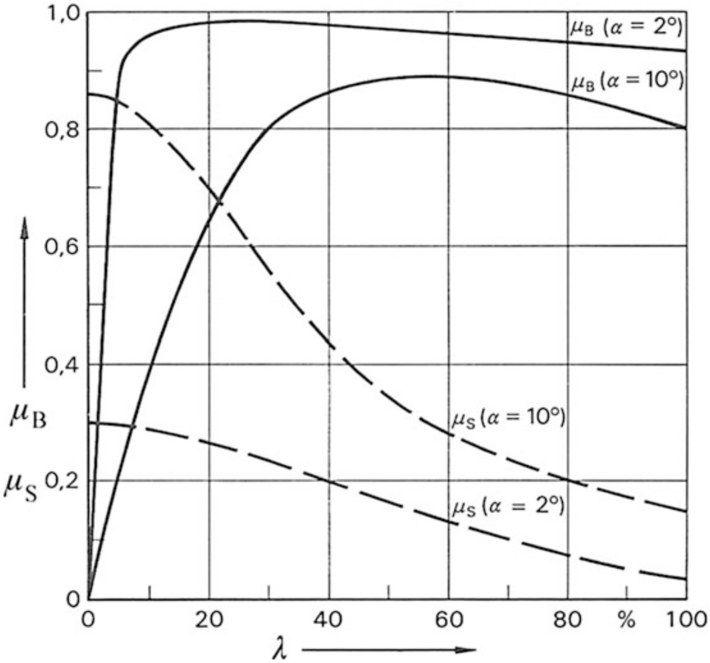
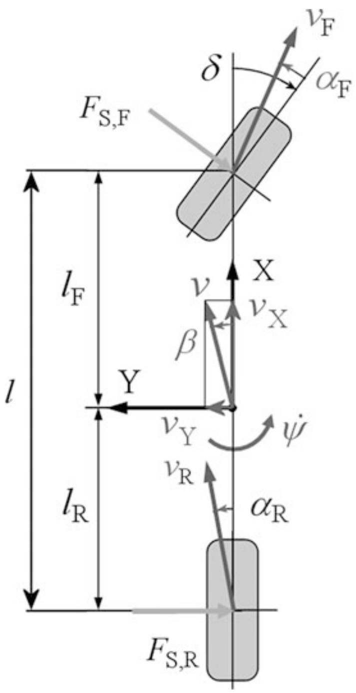


Fig. 2 μ-slip curves at some slip angle values and dependence of the lateral coefficient of friction on the slip

$M_R = \mu_B \cdot F_N \cdot r$. Between the brake slip and the brake coefficient of friction between the tire and the road, a nonlinear relationship exists, $\mu_B(\lambda)$, called the μ-slip curve. Figure 2 shows typical μ-slip curves. The curves usually exhibit a maximum, but for loose road surfaces like snow and gravel, the curve may not exhibit a maximum. The slip value at the maximum of the μ-slip curve is often called the target slip λ_T. For values of the tire slip larger than λ_T, the μ-slip curve is called unstable, since there is no stable equilibrium between the brake torque and the road torque: the tire slip will not be stable for a constant brake torque but usually increase until the wheel locks.

If a free rolling wheel whose longitudinal velocity is v_x is pushed sideways, then the wheel will also move in the lateral direction and the resultant velocity of the wheel center is v_V. The angle between the resultant wheel velocity and the wheel plane is called the slip angle α. Because the wheel is pushed sideways, the road pushes with a side force F_S in the opposite lateral direction on the wheel. The side force depends on the slip angle and on the normal force on the tire. Note that the maximum side force is not proportional to the normal force. The lateral coefficient of friction is the side force divided by the normal force on the wheel $\mu_S = F_S/F_N$. The nonlinear relation between the slip angle and the lateral coefficient of friction, $\mu_S(\alpha)$, is called the μ-slip angle curve which looks similar to that between the slip and the longitudinal coefficient of friction caused by braking.

Fig. 3 Bicycle model of the vehicle



The side force and the lateral coefficient of friction are also influenced by the tire slip and are reduced if the slip is increased. This is shown in Fig. 2. Similarly, if the slip angle is increased, then the brake force is reduced.

As mentioned above, for small steering wheel angles, vehicle handling is almost linear on a dry road. Handling is then described by the relationship between the steering angle of the front wheels and the vehicle yaw velocity using simple relations. First, the vehicle is simplified to a bicycle model that runs with constant velocity and where the transients have died out (Fig. 3). The lateral tire forces are supposed to increase linearly with the slip angles of the tires, where the slip angle is also influenced by compliance in the suspension and the steering system. This model is the basis for the vehicle dynamics control system ESC.

The steady-state yaw velocity can then be expressed by the following equation:

$$\dot{\psi} = \frac{v_X \cdot \delta}{(l_F + l_R) \cdot \left(1 + \frac{v_X^2}{v_{ch}^2}\right)} \tag{2}$$

The characteristic velocity v_{ch} determines the handling behavior of the vehicle. Its value depends on the effective lateral stiffness at the front axle, $c'_{\alpha F}$, and at the rear axle, $c'_{\alpha H}$, on the wheelbase $l = l_F + l_R$, on the vehicle mass m, and on the position of the front and rear axle w.r.t. the center of mass of the vehicle l_F and l_R, respectively.

$$v_{\text{ch}} = l \cdot \sqrt{\frac{1}{m} \cdot \left(\frac{c'_{\alpha\text{F}} \cdot c'_{\alpha\text{R}}}{l_{\text{R}} \cdot c'_{\alpha\text{R}} - l_{\text{F}} \cdot c'_{\alpha\text{F}}} \right)} \tag{3}$$

However, since $c'_{\alpha\text{F}}$ and $c'_{\alpha\text{H}}$ depend almost linearly with the vehicle mass and the position of the axles, v_{ch} is in a first approximation almost independent of changes in the vehicle mass and changes in the location of the axles. If v_{ch} is positive, then the vehicle behavior is called understeer. If v_{ch} is infinite, then the vehicle behavior is called neutral steer, and if v_{ch} is imaginary, then the vehicle behavior is called oversteer.

Since the lateral acceleration of the vehicle is limited by the maximum lateral coefficient of friction between the tires and the road, $\mu_{\text{S,max}}$, the steady-state yaw velocity is also limited by $\mu_{\text{S,max}}$.

$$|a_{\text{Y}}| = \left| \frac{v_{\text{X}}^2}{R} \right| = |\dot{\psi} \cdot v_{\text{X}}| \leq \mu_{\text{S,max}}, \Rightarrow |\dot{\psi}| \leq \left| \frac{\mu_{\text{S,max}}}{v_{\text{X}}} \right| \tag{4}$$

in which R is the radius of the turn.

The yaw velocity as a function of the vehicle velocity according to Eqs. 2 and 4 is shown in Fig. 4a for several steering wheel angle values. In this figure, curves of constant lateral acceleration of the vehicle, a_{Y}, are also shown for several values of a_{Y}. If the vehicle velocity increases and the value of a_{Y} reaches the value of $\mu_{\text{S,max}}$ (in Fig. 4a: $a_{\text{Y}} = 0.775$ g), then the yaw velocity is limited according to Eq. 4. For larger values of the vehicle velocity, the yaw velocity is determined by Eq. 4.

In Fig. 4b, the yaw gain is shown, which is defined as

$$\frac{\dot{\psi}}{\delta} = \frac{v_{\text{X}}}{(l_{\text{F}} + l_{\text{R}}) \cdot \left(1 + \frac{v_{\text{X}}^2}{v_{\text{ch}}^2} \right)} \tag{5}$$

Sudden rotation of the steering wheel can induce transients in the yaw velocity if the vehicle velocity is high enough (e.g., if $v_{\text{X}} > 60$ km/h). Figure 5 shows the yaw velocity after a sudden rotation of the steering wheel as measured in a front-wheel-drive middle-class car for two different car velocities. The measurement shows an oscillation of approximately 0.6 Hz in the yaw velocity. By comparison of the yaw velocities at the two car velocities, it follows that the decay of the oscillation is slower if the car velocity is higher. This can be explained by using the linear bicycle model for the evaluation of the yaw velocity and the lateral velocity of the car (Isermann 2006).

2.2 Rating Vehicle Dynamics

Handling properties are judged using vehicle maneuvers and include objective as well as subjective comparisons (Isermann 2006).

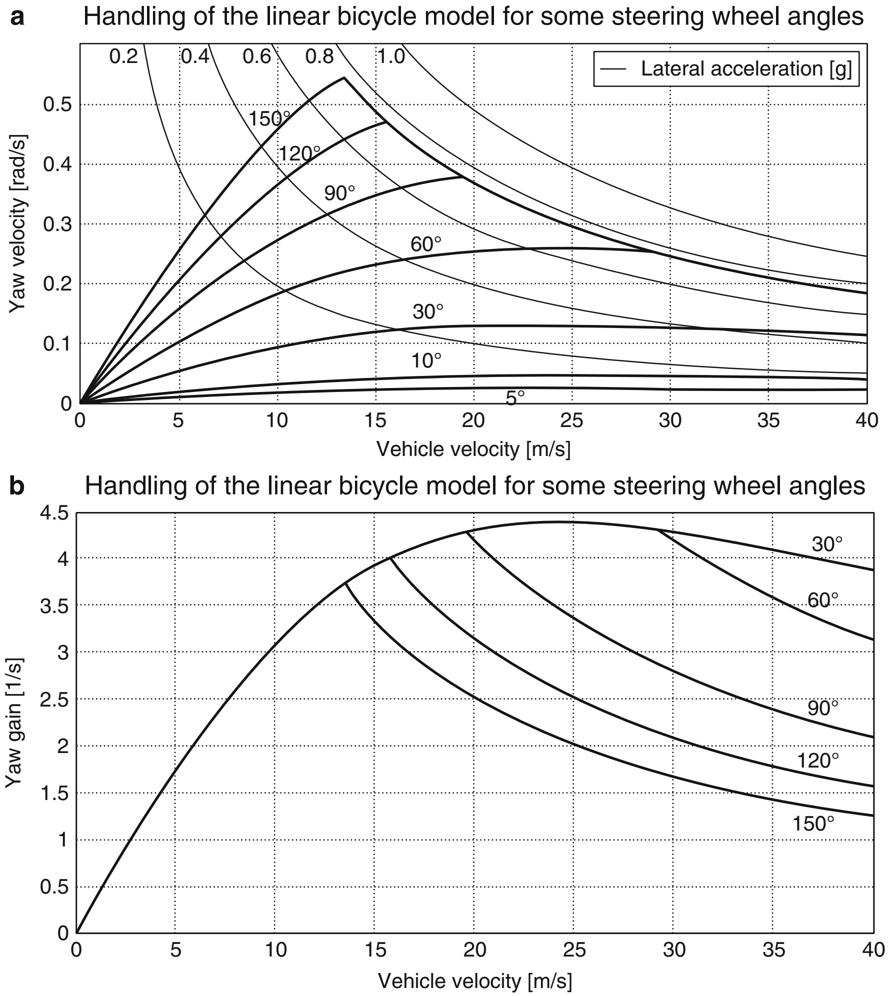


Fig. 4 Yaw velocity and yaw gain as a function of the vehicle velocity and the coefficient of friction of the road for various steering wheel angle values

For the objective rating of ABS, several ISO requirements exist, like "ISO 7975 (1996): Passenger cars – Braking in a turn – Open-loop test procedure." In Germany, the braking distance is usually measured with straight-line braking on a dry asphalt surface with initial velocity just above 100 km/h. The test procedure is described in DIN 70028: "Passenger cars – Measuring the stopping distance with ABS in straight-ahead stops." The test is carried out a couple of times and an average is computed for the braking distance. Usually the test is done for cold and hot brakes. For a straight-line braking maneuver on a μ-split road surface, the German magazine "Auto Motor und Sport" has defined a rating system in which both the braking distance and the vehicle stability are considered.

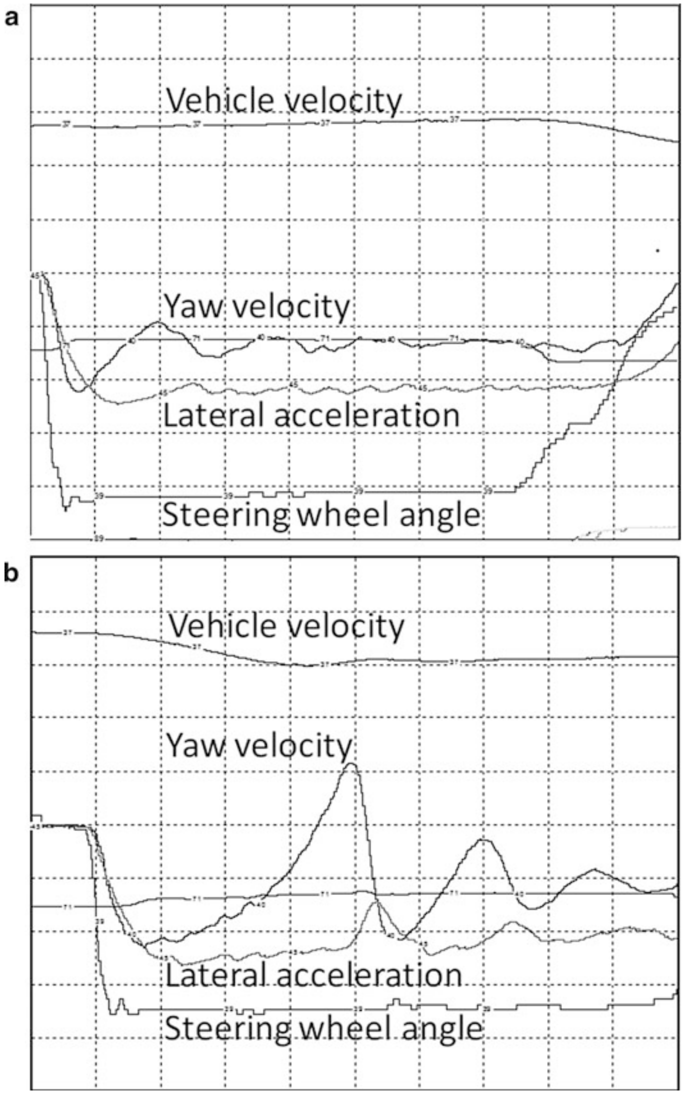


Fig. 5 Yaw velocity (**a**) after a step steering wheel angle input of 121° at a car velocity of 28 m/s and (**b**) after a step steering wheel angle input of 100° at a car velocity of 37 m/s. Plotting limits: time, 0–8 s; car velocity, −50 to +50 m/s; steering wheel angle, −145° to +145°; yaw velocity, −1 to +1 rad/s; lateral acceleration, −20 to +20 m/s^2

In the USA, where ESC is compulsory on all cars and light trucks (FMVSS 126) since 2011, some standard maneuvers have been defined for the rating of vehicles with ESC. In particular with the test "sine with dwell steering," the vehicle has to fulfill well-described performance limits (Fig. 6).

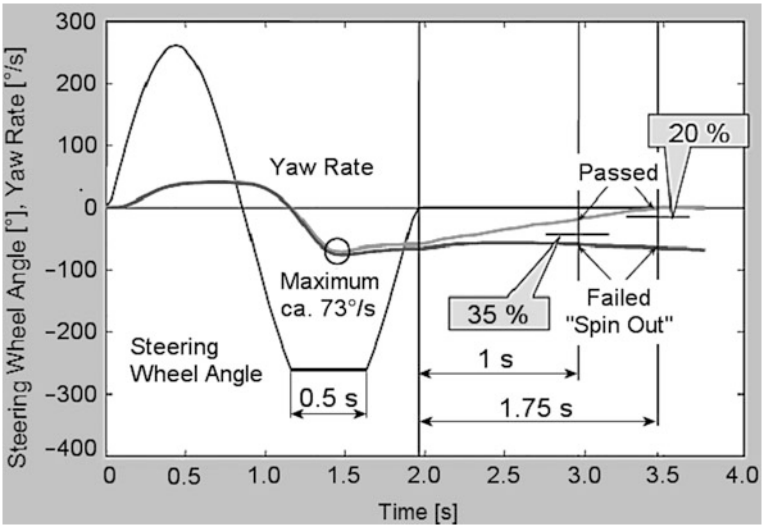


Fig. 6 Minimum requirements on ESC as defined by the NHTSA

The test maneuver is defined for a horizontal, smooth, dry, and solid road surface, with an initial velocity of 80 km/h, standard vehicle weight, tires with which the vehicle is sold, “sine with dwell steering” frequency 0.7 Hz, 500 ms dwell time, increasing the steering amplitude step by step by a factor of 0.5 of the initial steering angle up to the factor of 6.5 or up to the maximum of 270° steering wheel angle. The initial steering wheel angle is that angle at which the lateral acceleration of the vehicle is 0.3 g. The vehicle passes the test if 1 s after the sine with dwell steering the yaw velocity has dropped below 0.35 of its maximum value during the steering maneuver and below 0.2 of its maximum value during the steering maneuver after 1.75 s (“spinout” criteria of the NHTSA). For vehicles with gross vehicle weight up to 3500 kg, the lateral displacement must be at least 1.83 m after 1.07 s from the beginning of the steering maneuver. For heavier vehicles, this value is 1.52 m.

3 ABS, ASR, and MSR

3.1 Control Concept

In order to maintain some level of stability and steerability on all solid road surfaces, ABS must at least avoid the locking of the wheels during braking (Burkhardt 1993). For economical reasons, the vehicle velocity v_X is not measured, so that the free rolling wheel velocities and the slip levels during ABS control are unknown. For this reason, the control concept of ABS cannot be based on slip control. Instead, the control concept of ABS uses the wheel acceleration and

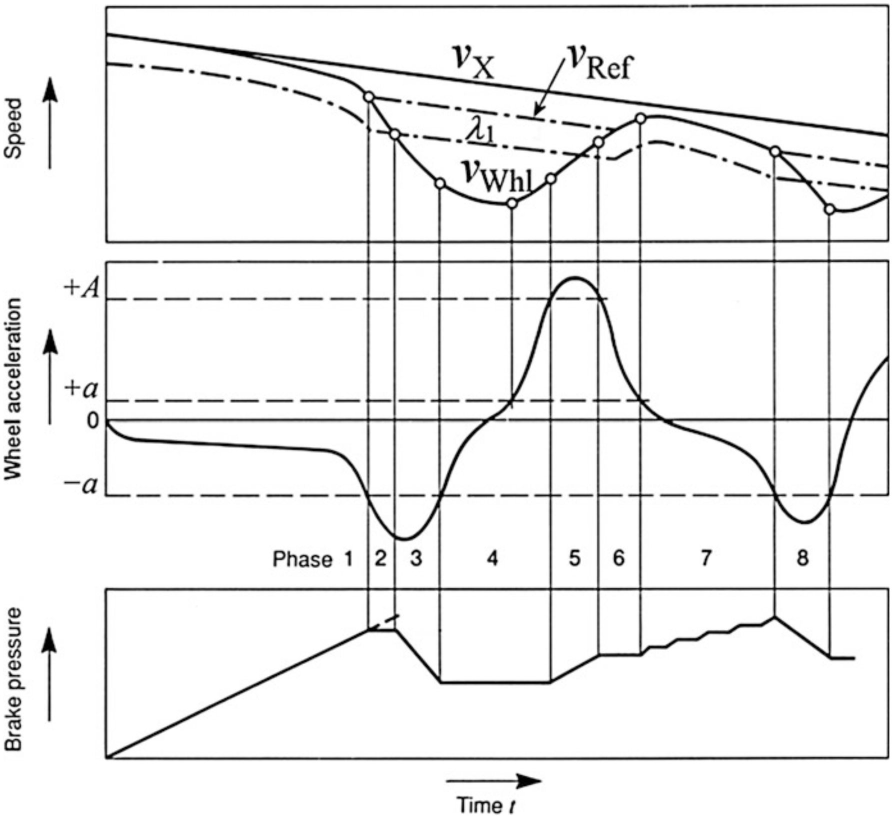


Fig. 7 Control concept at the beginning of ABS control

compares it with limit values. Therefore, the ABS control concept is called control logic. The limit values are chosen such that the slip remains close to the slip at the maximum of the μ-slip curve, λ_T. The control concept of ABS is sometimes also called the optimizer principle.

In Fig. 7, the first part in time of a typical ABS control history of one wheel is shown (Robert Bosch GmbH 2004). In phase 1, the brake pressure is shown as applied by the driver. Because of the increasing brake torque at the wheels, the vehicle and the wheels decelerate. If the wheel deceleration has reached the limit value of $-a$, further increase of brake pressure at the wheel is stopped, and the brake pressure is kept constant. The brake pressure is not reduced yet, since the increasing brake force from the road on the wheel makes the wheel axle move in a longitudinal direction of the vehicle relative to the vehicle chassis because of the compliance of the wheel suspension. This motion results in a fast increase of the wheel deceleration although the wheel slip has not yet reached its value at the maximum of the μ-slip curve λ_T. The brake pressure is reduced only after the wheel velocity v_{Whl} has dropped substantially below the so-called reference velocity v_{Ref}. The reference velocity is an artificial signal which is supposed to be approximately equal to the wheel velocity if the wheel slip were λ_T. From the beginning of the brake application, the reference velocity follows the wheel velocity until the wheel deceleration has reached the limit value $-a$. Then the reference velocity is extrapolated with a certain rate, which is -0.3 g at the beginning of the ABS control. As soon as the wheel velocity has also dropped below the limit value λ_1, the brake pressure is reduced by the ABS control.

In phase 3, the brake pressure at the wheel is continuously reduced as long as the wheel deceleration is below the limit value $-a$. In phase 4, the brake pressure at the wheel is kept constant which results in an acceleration of the wheel. In phase 5, the acceleration increases above a limit value $+A$. The brake pressure is increased as fast as possible, and the increase is continued as long as the acceleration is larger than the limit value $+A$. Otherwise the brake pressure is kept constant (phase 6). If the acceleration has dropped below the limit value $+a$, it is assumed, that the brake force is nearly maximal. The slip has now almost reached a stable value near λ_T. Since the brake force is almost maximal, it suffices to increase the brake pressure slowly which is realized by a stepwise increase of the brake pressure. Thus, the brake force is almost maximal and stays maximal for a longer time. If the wheel deceleration drops again below the limit value $-a$, the brake pressure is reduced immediately in phase 8 since it is supposed that the brake force from the road on the wheel is almost constant because of the ABS control, and therefore the axle motion relative to the chassis is almost zero. The first step of the stepwise pressure increase is variable and chosen such that after the first step increase of the brake pressure, the tire slip is already close to λ_T.

If during propulsion the driven wheels spin, the lateral force on the vehicle cannot be influenced by changing the slip angles of the driven wheels. ASR, the traction slip control system, should prevent the wheels from spinning and controls the traction slip of the driven wheels by modulation of the engine torque and, if necessary, also by active braking of the driven wheels. The μ-slip curve for traction slip looks very similar to the μ-slip curve for brake slip. However, ASR cannot adopt a control logic similar to that of ABS, which is based on the wheel acceleration. The reason is that the clutch is engaged so that the effective moment of inertia at the driven wheels is very large, particularly in the first gear. It is then not possible to clearly differentiate between stable and unstable slip values on the basis of the wheel acceleration. Moreover, the engine torque also depends on the engine speed. Therefore, the ASR control concept is different from the control concept of ABS.

ASR controls the wheel traction slip. For the evaluation of the wheel traction slip, the velocity of the nondriven wheels is used as the vehicle velocity. For all wheel-drive vehicles, this strategy cannot be used. For these vehicles only a traction control, which synchronizes the velocities of the driven wheels, was offered without engine torque modulation. However, with the introduction of ESC, the vehicle velocity is also estimated during ASR, so that the ASR control also became possible for four-wheel-drive vehicles.

During straight-line acceleration on a homogeneous road surface, the traction forces and the traction slips at the driven wheels on the same axle are approximately

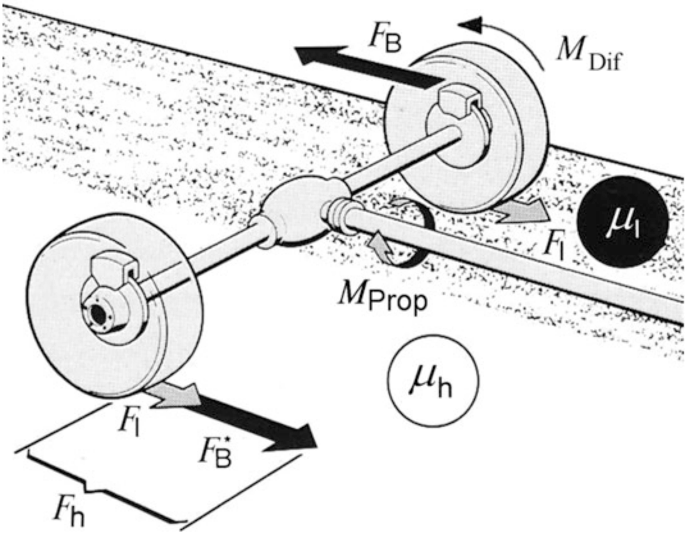


Fig. 8 ASR traction control on an asymmetric road

equal. Both driven wheels may spin if the traction slip reaches λ_T. ASR then reduces the engine torque to reduce the wheel traction slips. With gasoline engines, the throttle valve is closed, while with diesel engines, the control lever of the injection pump is retracted.

During straight-line acceleration on a μ-split road surface, the wheel on the low-μ road surface (μ_l) may spin first (Fig. 8). Without traction slip control, the traction forces at both wheels are equal to F_l. Traction slip control actively brakes the spinning wheel on the low-μ road surface with a force F_B on the brake disk to keep the wheel slip close to λ_T. This results in an increase of the traction force at the wheel on the high-μ road surface by F_B*. If the torque at the propulsion shaft M_{Prop} is so large, that the wheel at the high-μ road surface also starts to spin, i.e., $F_h = F_l + F_B$* has reached its maximum value, the engine torque is also reduced by ASR control. Ideally, ASR keeps the wheel velocities of both wheels approximately equal, since a difference between the velocities together with a high torque at the axle will stress the differential considerably.

If the accelerator pedal is released, the resulting braking torques at the driven wheels may be so high that in particular on low-μ road surfaces, the engine runs more or less in idle velocity. The brake slips at the driven wheels may then be much larger than λ_T. The driver may lose control of the vehicle. ASR can also be used in such situations to increase the engine torque in order to reduce the brake slips of the driven wheels. This control is called engine drag torque control (MSR). The ASR and MSR control will be described in more detail in Sect. 4.

Brake regulations (ECE13-H) require that during straight-line braking with vehicle decelerations below 0.85 g, the rear wheels must not lock before the wheels at the front wheel lock. This requirement can be achieved by appropriate

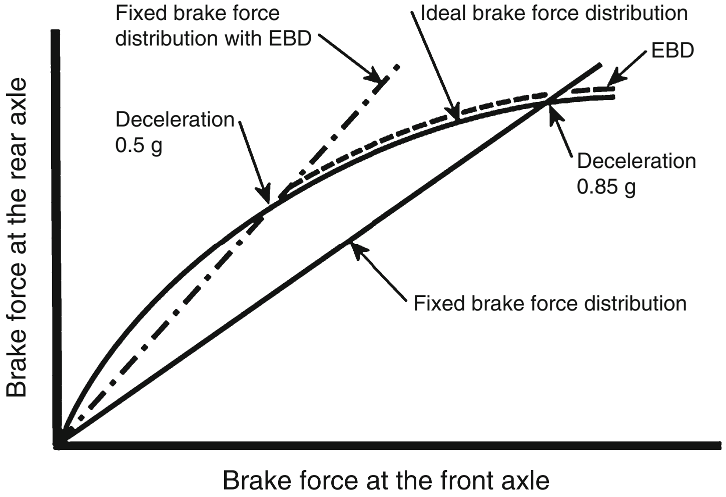


Fig. 9 Principle of the electronic brake force distribution, EBD

dimensioning of the brakes at the front wheels and at the rear wheels. The brake force distribution front to rear is then fixed (Fig. 9), with the result that for most brake applications where the vehicle deceleration is below 0.3 g, the rear wheels hardly contribute to the total brake force on the vehicle. The brake force distribution is called ideal, if the coefficient of friction between the tires and the road surface at the front and the rear wheels are equal. The fixed brake force distribution can be modified to approach the ideal brake force distribution for small vehicle decelerations. However, the regulations may then be violated for vehicle decelerations below 0.85 g, e.g., for vehicle decelerations above 0.5 g (Fig. 9). Therefore, in the past, mechanical components have been installed to limit or reduce the brake pressure at the rear wheels for larger vehicle decelerations. Wear, corrosion, and other influences on the function of these mechanical components may again lead to violation of the regulations after some time.

The function of the mechanical components can also be implemented by an extension of ABS which is called the electronic brake force distribution (EBD). EBD monitors the wheel velocities and controls the brake pressure at the rear wheels such that the velocity at the rear wheels is not by much lower than the velocity at the front wheels. If the front and the rear wheel velocities are equal, then the slips of all wheels are equal, and the coefficients of friction between the tires and the road are equal at all wheels. Thus, EBD makes the brake force distribution to follow the ideal brake force distribution closely. For most brake applications, the vehicle deceleration is small (in Fig. 9 below 0.5 g), and the brake force at the rear wheels of the fixed distribution is smaller than the brake force of the ideal distribution. Thus, for most brake applications, the velocity of rear wheels is larger than the velocity of the front wheels, and EBD will not intervene. EBD interventions may produce some noise and some brake pedal feedback which may irritate the driver. However, for most brake applications with vehicle decelerations below 0.5 g where EBD does not intervene, there will be no irritation of the driver.

The EBD function is the following: If the velocity of the rear wheels is below the velocity at the front wheels by a certain first limit value, then the brake pressure at the rear wheels is kept constant. This first limit value depends on the vehicle velocity. If the velocity of the rear wheels is below the velocity at the front wheels by a certain second limit value which is larger than the first limit value, then the brake pressure at the rear wheels is reduced. The brake pressure reduction at the rear wheels is continued as long as the velocity of the rear wheels is below the velocity at the front wheels by the second limit value. Thus, the difference in the velocities between the front and rear wheels is kept small. If the difference becomes smaller than the first limit value, then the brake pressure at the rear wheels is increased in a stepwise manner. In case of a failure of ABS and if the EBD function fails, the driver is informed by a red light since ECE13-H regulation is violated.

4 ESC

4.1 Requirements

The requirements on ESC relate to the handling of the vehicle at the physical limit of lateral dynamics. Handling performance is judged by experts objectively as well as subjectively. However, the tuning of ESC on the vehicle largely depends on the expert and on the philosophy of the company. There is hardly any correlation between the subjective and objective rating of the vehicle handling performance.

At the physical limit, the tire forces on the road cannot be increased in magnitude. During full braking, a compromise must be found between the desire for maximum longitudinal forces on the tires for the shortest stopping distance on the one hand and for maximum lateral forces on the tires for smallest deviations of the vehicle from the desired path on the other hand. During free rolling of the vehicle, a compromise must be chosen between best steerability and stability of the vehicle on the one hand and unintended deceleration of the vehicle on the other hand.

The requirements of ABS and ASR also apply to ESC. Additional requirements on ESC describe the compromises mentioned above (Breuer and Bill 2013).

4.2 Sensors

For the evaluation of the vehicle state, cheap sensors for automotive applications are used. These are an inertial angular rate sensor for the yaw velocity and an accelerometer for the lateral acceleration of the vehicle. For the evaluation of the nominal vehicle motion, an angle sensor for the steering wheel angle and a pressure sensor for the pressure in the brake master cylinder (MC) are used. Furthermore, wheel velocity sensors as used for ABS and ASR are also used for the evaluation of the wheel velocities. These sensors are described in ▶ Chap. 15, "Vehicle Dynamics Sensors for DAS" of this book.

4.3 Control Concept of ESC

ESC was developed on the basis of ABS and ASR with which the pressure in the individual brake wheel cylinders (WC) can be modulated, the wheels can be actively braked, and the engine output torque can be controlled. The control concept of ESC relies on the effect that the lateral force on a tire can be modified by tire slip (Fig. 2). Thus, the lateral dynamics of the vehicle can be influenced by modification of the slip value of each individual tire. For this reason, ESC uses the wheel slip as the vehicle dynamics control variable (van Zanten et al. 1994). Basically, the yaw moment from the road on the vehicle can be influenced by the slip value of each of the four wheels. However, a change of the slip value at a wheel also implies a change of the longitudinal force from the road on the wheel and with that an unintentional change in the vehicle acceleration.

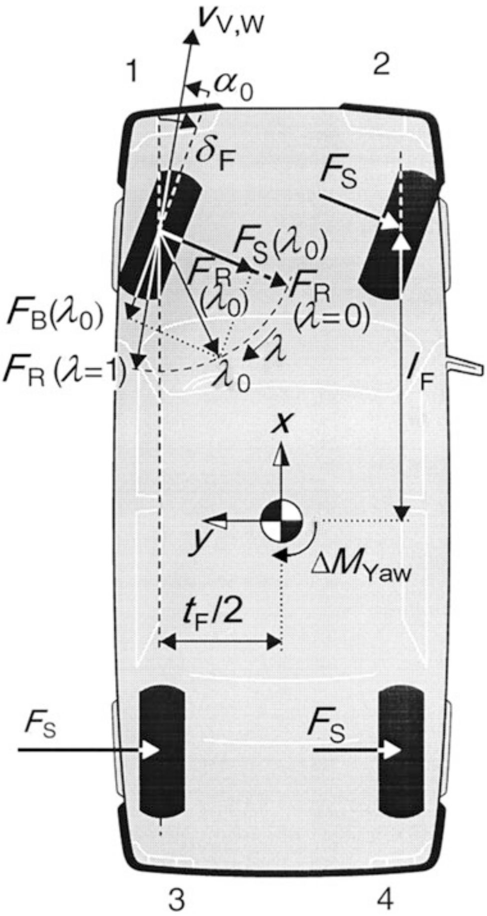
The effect of a slip change of the wheel is a rotation of the resulting horizontal force from the road on the wheel. This is shown in Fig. 10. In the figure, a car is shown in a right turn with no braking forces nor traction forces on the wheels, and it is supposed that the lateral wheel forces have reached their maximum values, i.e., that the slip angles of the wheels are those where the slip angle curves have their maximum values. On the front left wheel, the resulting force from the road is initially $F_\mathrm{R}(\lambda = 0)$ which is equal to the side force $F_\mathrm{S}(\lambda = 0)$. If ESC applies a brake slip of λ_0 to the front left wheel, then the lateral force will be reduced to $F_\mathrm{S}(\lambda_0)$, and a brake force from the road on the wheel results which depends on the wheel slip $F_\mathrm{B}(\lambda_0)$. The geometrical sum of these forces is then the resultant force $F_\mathrm{R}(\lambda_0)$. Usually, the friction circle between the tire and the road is assumed (Schindler 2007). This means, that the absolute values of $F_\mathrm{R}(\lambda = 0)$ and $F_\mathrm{B}(\lambda_0)$ are equal. As a result of the brake slip application, the resultant wheel force from the road on the tire is rotated, and the yaw moment of the wheel about the center of mass of the car is modified by ΔM_Yaw.

$$\Delta M_\mathrm{Yaw} = \frac{\partial F_\mathrm{S}}{\partial \lambda} \cdot \Delta\lambda \cdot (l_\mathrm{F} \cdot \cos\delta_\mathrm{F} - 0.5 \cdot t_\mathrm{F} \cdot \sin\delta_\mathrm{F}) - \frac{\partial F_\mathrm{B}}{\partial \lambda} \cdot \Delta\lambda \cdot (0.5 \cdot t_\mathrm{F} \cdot \cos\delta_\mathrm{F} + l_\mathrm{F} \cdot \sin\delta_\mathrm{F}) \tag{6}$$

The rotational angle of the resultant force increases with increasing brake slip but is limited by the maximum brake slip $\lambda = 1$. If the brake slip is 1, then the direction of the resultant brake force of the road on the wheel is opposite to the direction of the horizontal velocity of the car at the wheel center $v_\mathrm{V,W}$.

In Fig. 11, the influence of changes in the values of brake slip on the yaw moment on the car at each wheel is shown for all brake slip values, for a dry asphalt

Fig. 10 Rotation of the resultant tire force F_R by a slip change from 0 to λ_0



road (μ ≈ 1.0), as well as for a packed snow road surface (μ ≈ 0.3) as computed by simulation of a simple car/tire model. It shows that by changing the brake slip of the front left wheel, the yaw moment on the center of mass of the car may be changed considerably, from approximately +5000 Nm at $\lambda = 0$ to −4000 Nm at $\lambda = 1$ on a dry asphalt road. On the packed snow road surface, the result is qualitatively the same; however, the values are smaller because of the lower coefficient of friction of packed snow. Variation of the brake slip of the front right wheel shows little influence on the yaw moment. The reason is that the lever arm of the resulting wheel force to the center of mass of the car increases with increasing brake slip values first, but decreases again for larger brake slip values. As a result, the yaw moment curve shows a small maximum. Changing the brake slip at the front axle of the wheel on the outside of the turn has a much larger influence on the yaw moment of the car than changing the brake slip at the front axle of the wheel on the inside of the turn in this vehicle situation. This has to be considered with the control of the

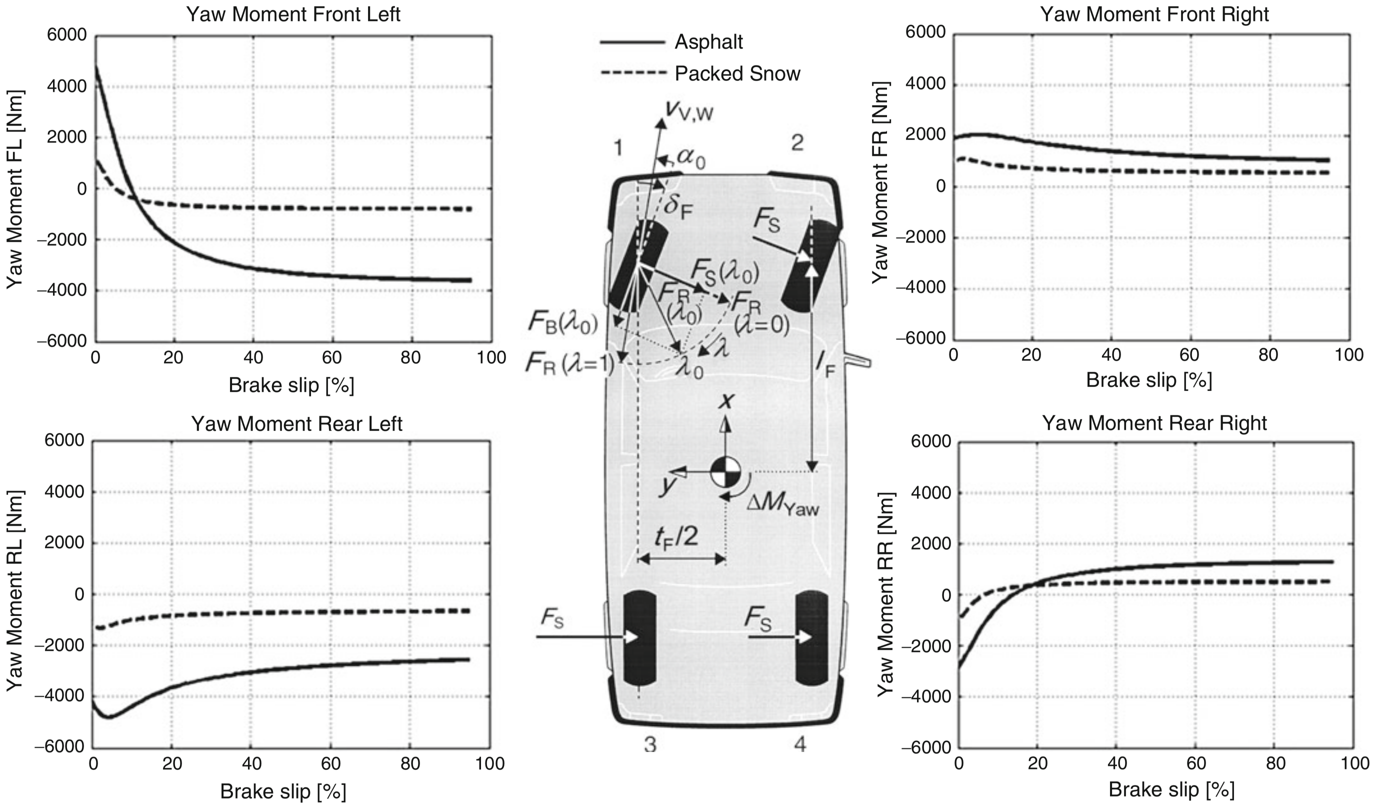


Fig. 11 Influence of a wheel individual brake slip change of the yaw moment on a dry asphalt road and on a packed snow road during stationary cornering with limit velocity and with radius of the turn 100 m

yaw moment on the car. During heavy cornering, the rear wheel on the inside of the turn may lift off, and brake slip interventions at this wheel will have no effect on the yaw moment on the car. This has to also be considered with the control of the yaw moment on the car. If in the free rolling situation of the car at the physical limit its behavior is oversteer during cornering, increasing the brake slip of the wheel at the front axle on the outside of the turn reduces the yaw moment and thus reduces the oversteer behavior of the car.

In Fig. 12, the influence of changes in the value of traction slip on the yaw moment on the car at each wheel is shown for all traction slip values, for a dry asphalt road ($\mu \approx 1.0$), as well as for a packed snow road surface ($\mu \approx 0.3$) as computed by simulation of a simple car/tire model. It follows that by changing the traction slip value of the front left wheel of a front-wheel-drive car, the yaw moment on the center of mass of the car does not change by much and shows a small maximum like explained for the front right wheel with brake slip above. On the packed snow road surface, the result is qualitatively the same; however, the values are smaller because of the lower coefficient of friction of packed snow. For a rear-wheel-drive car, the influence of changes in the value of traction slip on the yaw moment on the car is also shown. In particular, traction slip changes at the rear axle on the outside of the turn change the yaw moment on the car considerably from -4000 Nm at $\lambda = 0$ to +2000 Nm at $\lambda = -1$ (for historical reasons, traction slip is defined to be negative).

Because of the clear relationship between wheel slip and yaw moment, a hierarchical structure of the control is very useful, in which the control of the vehicle motion determines the required yaw moment change and in which the control of the wheel slips determines the rotation of the forces on the wheels. Control of the wheel slips includes the basic functions of ABS and ASR. Since ASR controls the wheel slips of the driven wheels, the ASR control can be used as wheel slip control for traction slip. However, since standard ABS control is based on wheel acceleration and not on wheel slip, ABS control cannot be used for wheel slip control during ABS braking. Instead, a novel ABS control is used which controls the wheel slip during ABS braking. Both the traction slip control and the brake slip control will be explained below. However, ESC control structures exist in which both an acceleration controller for the ABS function and a brake slip wheel controller for the modification of the yaw moment are used. This control structure is called modular.

The hierarchical control structure is shown in Fig. 13 with the vehicle as the process to be controlled and with the central elements the state variable control for the control of the vehicle motion, the brake slip control, and the traction slip control. An important component of the ESC control is an observer, with which the vehicle motion is analyzed and estimated. Another important component is the determination of the nominal motion of the vehicle expressed by its yaw velocity and its slip angle. These values are evaluated by using the driver input (the steering wheel angle δ, the brake master cylinder pressure p_{Circ}, and the driver-requested engine torque $M_{\mathrm{M,Pre}}$ as obtained from the engine management system), the lateral acceleration of the vehicle a_{y}, the longitudinal acceleration a_{X} of the vehicle, and the

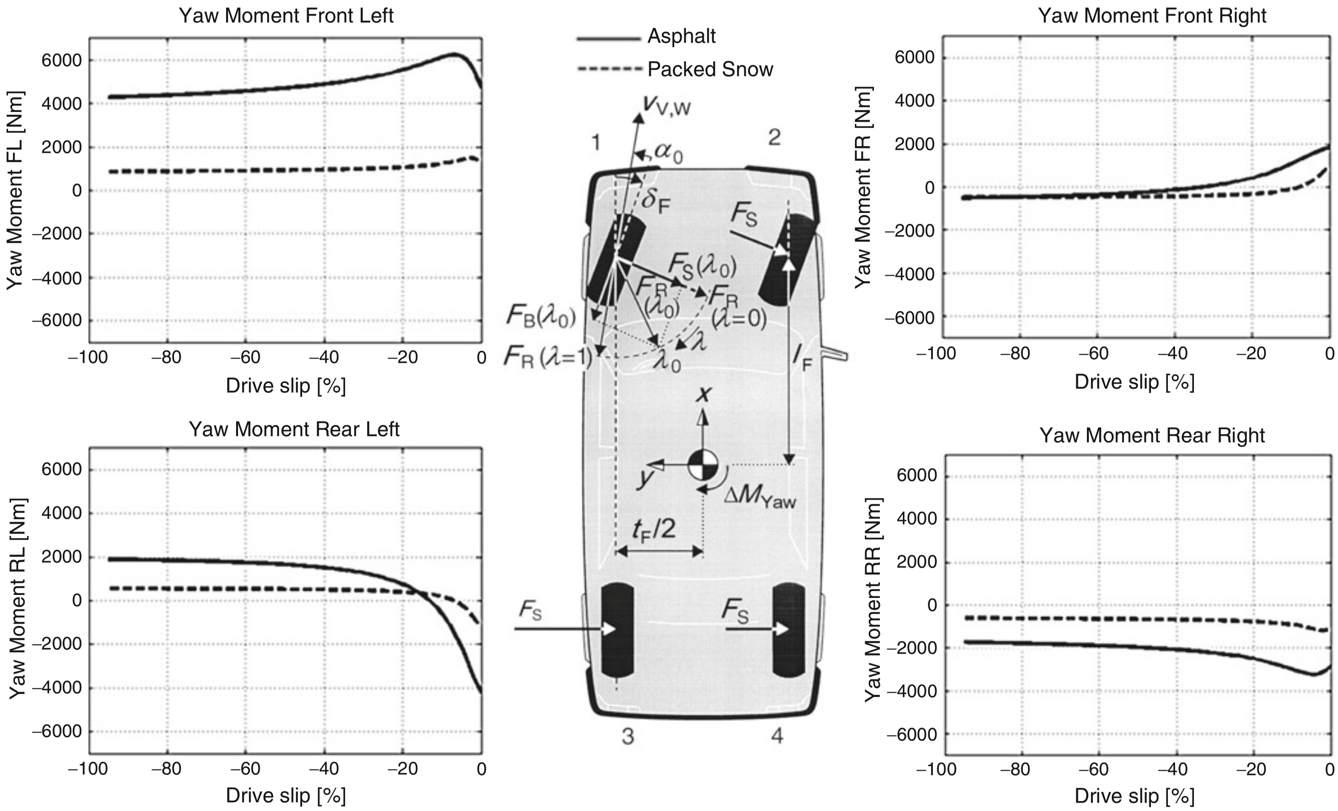


Fig. 12 Influence of wheel individual traction slip on the yaw moment on a dry asphalt road and on a packed snow road during stationary cornering with limit velocity and radius of the turn 100 m

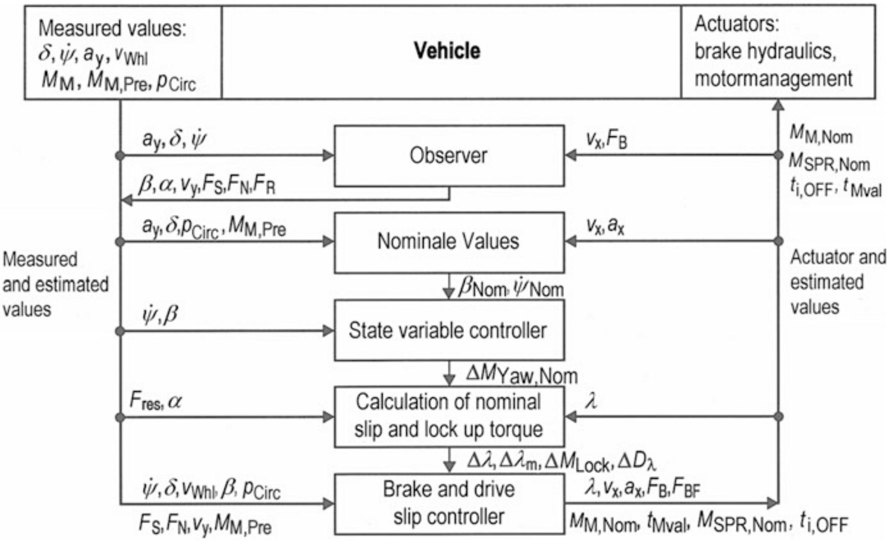


Fig. 13 Simplified block diagram of the ESC control with input and output quantities

longitudinal velocity of the vehicle v_X. The output of the state variable controller is the required change in the yaw moment on the vehicle $\Delta M_{\mathrm{Yaw,Nom}}$. From the required change in the yaw moment on the vehicle $\Delta M_{\mathrm{Yaw,Nom}}$, the required changes in each wheel slip $\Delta\lambda$ can be computed, e.g., from Eq. 6 for the left front wheel. Since changes in the slip value at one wheel may have a much larger effect on the yaw moment on the vehicle than changes in the slip value at another wheel as explained by Fig. 10, the wheels that are chosen to change the yaw moment must be carefully selected. The brake and traction slip controllers compute the target slip values λ_T for maximum brake forces and maximum traction forces, respectively, dependent on the coefficient of friction of the road and of the vehicle velocity at each wheel. Each slip value λ_T is merely modified by a slip value change $\Delta\lambda$ required for the change in the yaw moment on the vehicle and results in the nominal slip value $\lambda_{\mathrm{Nom}} = \lambda_T + \Delta\lambda$ for each wheel. Therefore, the structure of the slip controllers is independent of the yaw control, which demonstrates the hierarchical structure of the controller shown in Fig. 13. A detailed description of the control structure can be found in Isermann (2006).

The brake slip controller structure is shown in Fig. 14. The task of the brake slip controller is on the one hand the ABS function for shortest stopping distance and on the other hand the control of the wheel slip changes $\Delta\lambda$ which are demanded by the vehicle motion control. In order to combine both tasks in a unified manner, a novel brake slip controller was designed by which the ABS function is realized by brake slip control instead of wheel acceleration control as is used with standard ABS. For the ABS function, the brake slip controller determines the target slip, λ_T, which is the slip at the maximum of the μ-slip curve.

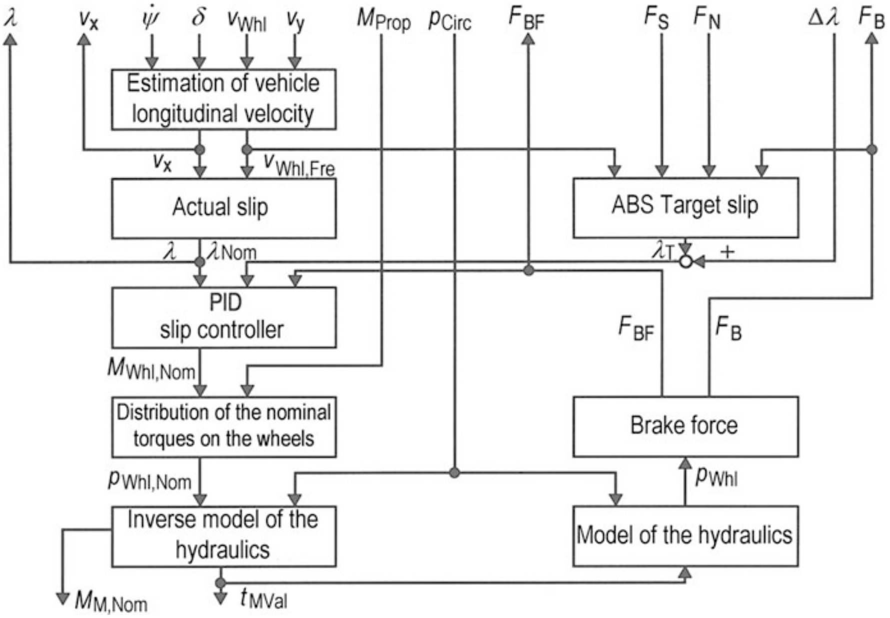


Fig. 14 Block diagram of the brake slip control with the most important modules and their input and output quantities

In order to compute the actual wheel slip, the longitudinal vehicle velocity must be known. Since this velocity is not measured, it must be estimated. The estimation procedure is described in Sect. 4.4. The nominal wheel slip is computed as described above.

$$\lambda_{\mathrm{Nom}} = \lambda_{\mathrm{T}} + \Delta\lambda \tag{7}$$

The traction slip controller is used to control the traction slip of the driven wheels during propulsion. Active slip interventions at the nondriven wheels which are required for changes in the yaw moment are controlled by the brake slip controller. The traction slip controller structure (Fig. 15) is described for a rear-wheel-drive vehicle.

Because of the differential, the transmission and the engine with the clutch the two driven wheels are mechanically coupled. This means that the equations of motion of the driven wheels are also coupled. By change of variables, the equations of motion become independent. The new variables are the mean velocity of the two driven wheels $V_{\mathrm{Prop}} = (v_{\mathrm{Whl,RL}} + v_{\mathrm{Whl,RR}})/2$ and the difference velocity of the two driven wheels $V_{\mathrm{Dif}} = v_{\mathrm{Whl,RL}} - v_{\mathrm{Whl,RR}}$ in which $v_{\mathrm{Whl,RL}}$ is the velocity of the left rear wheel and $v_{\mathrm{Whl,RR}}$ is the velocity of the right rear wheel (Isermann 2006). For this reason, the slip of each wheel is not controlled by itself, but instead the mean velocity V_{Prop} and the wheel velocity difference V_{Dif} are controlled.

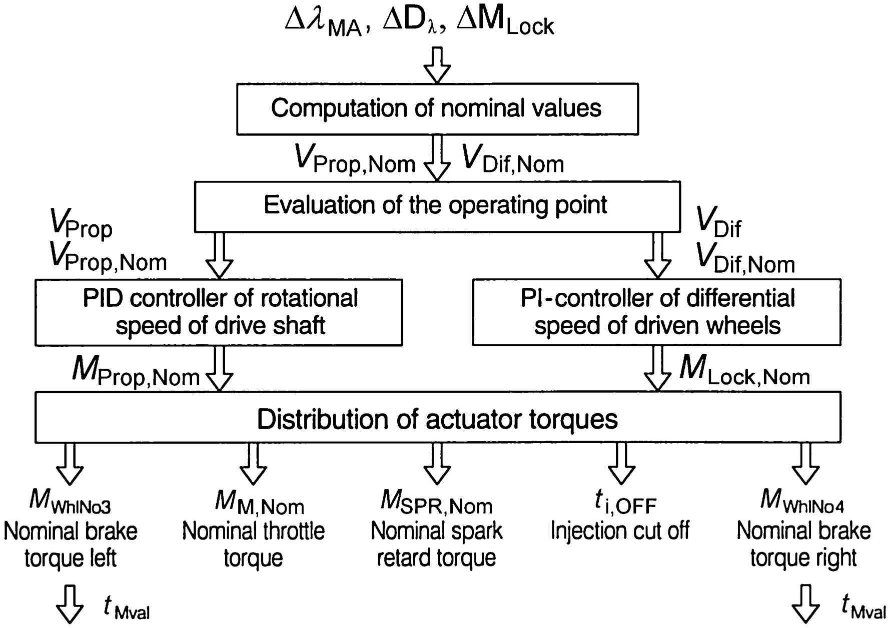


Fig. 15 Block diagram of the traction slip control with the most important modules and their input and output quantities

For the determination of the nominal mean velocity $V_{\mathrm{Prop,Nom}}$, a symmetrical nominal slip value λ_m is computed (where both driven wheels have the same nominal slip value λ_m). For the determination of the nominal value of the wheel velocity difference $V_{\mathrm{Dif,Nom}}$, an asymmetric slip value D_λ is computed, where the difference in the nominal slip values between the two driven wheels is D_λ. Using the longitudinal vehicle velocity v_X, the nominal values of the mean wheel velocity $V_{\mathrm{Prop,Nom}}$ and the wheel velocity difference $V_{\mathrm{Dif,Nom}}$ can be computed from the nominal symmetric slip value λ_m and the nominal asymmetric slip value D_λ.

The dynamics of the two equations of motion depend on the state of the controlled process. Therefore, the operating point is evaluated (which considers, e.g., the influence of the gear ratio i_G of the transmission) and used to adjust the controller gains. Control variables for the mean velocity are the engine torque and a symmetric brake torque of the driven wheels. The control variable for the difference velocity is the asymmetric brake torque control of the driven wheels.

As mentioned above, besides the hierarchical control structure of ESC, a modular control structure also exists (Rieth et al. 2001) where a brake slip controller is used for changes in the yaw moment, and a standard ABS controller is used for the ABS function. If ABS and the yaw moment controller work simultaneously, then the access of the controllers to the magnetic valve stimulations must be coordinated by a so-called arbitrator. Instead of the vehicle velocity v_X, the reference velocity v_{Ref} is used by the yaw moment controller (Fig. 7). Because of the arbitration,

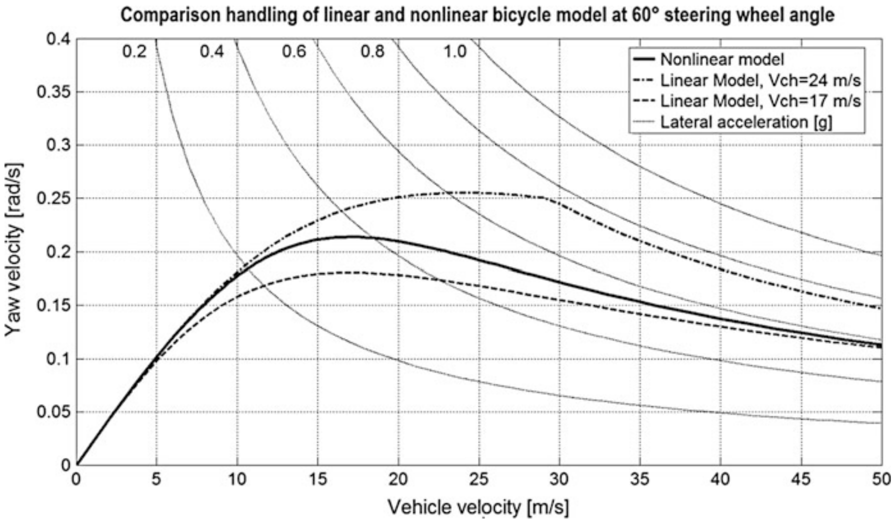


Fig. 16 Approximation of the nonlinear vehicle model by weighted interpolation between two linear bicycle models

ABS control must be stopped once in a while. The same applies to the yaw moment control. Therefore, the control logic of ABS and of the yaw moment controller must be modified (e.g., for the determination of the integrals of ESC and for the "learning" of ABS). Also heuristic elements like the recognition of positive μ-jumps and adaptive modifications of gains must be modified. This is also to be considered during traction control.

4.4 Determination of Nominal Values and Estimation of Vehicle Dynamic Quantities

The nominal value of the vehicle slip angle β_{Nom} is defined by the requirements on ESC. The nominal value of the yaw velocity is defined by a model of the vehicle. This model is based on the linear steady-state bicycle model as described in Sect. 2.1. However, the yaw velocity of the linear steady-state bicycle model does not represent the yaw velocity of the vehicle with sufficient accuracy. In particular, the model must be extended to improve the accuracy of the yaw velocity if the vehicle dynamics approaches the physical limit. Otherwise, the yaw moment changes may occur late, when the vehicle is already skidding, or early, before the physical limit is reached. Expert drivers complain most if the latter situation occurs. Instead of one, linear steady-state bicycle model ESC uses two of them (Fig. 16). The difference between the two models is a different value of the characteristic velocity v_{ch}. Each of the linear steady-state bicycle models produces a yaw velocity for a certain vehicle velocity and steering wheel angle. The characteristic velocities are chosen such that the yaw velocity of the vehicle lays in between the yaw velocities of the two models. A weighted average is computed from these two yaw velocities to result in the nominal yaw velocity of the vehicle. The weights are chosen dependent on the lateral acceleration of the vehicle and of the driving situation.

Using a four-wheel model, the observer estimates from the measured yaw velocity, the steering wheel angle, and the lateral acceleration and from the estimated values of the braking and driving forces on the wheels, the slip angle of each wheel, the slip angle of the vehicle, and the vehicle velocity (Fig. 13). Also, the side and vertical forces on the wheels are estimated, and the resultant forces are computed. The four-wheel model also computes the transients and considers special situations like μ-split and banked roads.

On a horizontal, homogeneous road, the differential equation for the slip angle of the vehicle β is as follows:

$$\dot{\beta} = -\dot{\psi} + \frac{1}{v_{\mathrm{X}}} \cdot (a_{\mathrm{Y}} \cdot \cos\beta - a_{\mathrm{X}} \cdot \sin\beta) \tag{8}$$

in which a_{X} and a_{Y} are the longitudinal and the lateral acceleration of the vehicle, respectively. For small values of the longitudinal acceleration and of the slip angle of the vehicle, the equation can be reduced to

$$\dot{\beta} = \frac{a_{\mathrm{Y}}}{v_{\mathrm{X}}} - \dot{\psi}, \;\; \beta(t) = \beta_0 + \int\limits_{t=0}^{t} \left(\frac{a_{\mathrm{Y}}}{v_{\mathrm{X}}} - \dot{\psi} \right) \mathrm{d}t \tag{9}$$

Since the measured values of the lateral acceleration and of the yaw velocity as well as the estimated vehicle velocity include some errors, the error in the integral grows rapidly in time so that the computed slip angle of the vehicle cannot be trusted after a while.

For large values of the longitudinal acceleration of the vehicle, a Kalman filter can be used as an observer for the lateral velocity of the vehicle. The Kalman filter uses the equations of motion of the lateral velocity and the yaw velocity of a four-wheel model of the vehicle. Included in the observer is an estimate of the road inclination (Isermann 2006). Included in the observer are also simple estimates like those of the wheel normal forces which are based on the longitudinal and lateral acceleration of the vehicle.

For the control of the wheel slip, the longitudinal vehicle velocity must be known. Since this velocity is not measured for cost and other reasons like weather conditions, the velocity of the vehicle is derived from the velocities of the wheels. This is particularly difficult if all wheels show some slip like with ABS. Therefore, the derivation of the longitudinal vehicle velocity will be described for the case in which all wheels are controlled by ABS.

For the derivation of the longitudinal vehicle velocity during ABS, a single wheel is selected for which the ABS slip control to the nominal value λ_{Nom} is interrupted for some time period, called the adaptation time period or adaptation phase (Fig. 17). At the beginning of the adaptation time period, the wheel brake

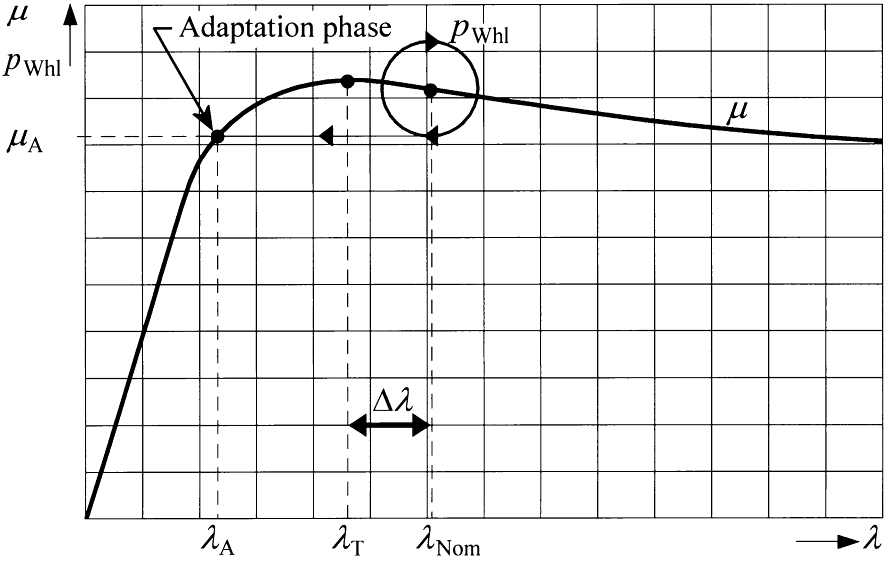


Fig. 17 Adaption phase during ABS brake slip control for the computation of the free rolling wheel velocity (the *circle* pR symbolically indicates the pressure modulation of ABS control)

pressure is lowered by some amount and kept constant such that a stable wheel slip λ_{A} results. The free rolling wheel velocity is then

$$\begin{aligned}\mu_{\mathrm{A}} &= \frac{F_{\mathrm{B,A}}}{F_{\mathrm{N,A}}} = c_{\lambda} \cdot \lambda_{\mathrm{A}} = c_{\lambda} \cdot \frac{v_{\mathrm{WhlFre,A}} - v_{\mathrm{Whl,A}}}{v_{\mathrm{WhlFre,A}}} \qquad \rightarrow v_{\mathrm{WhlFre,A}} \\ &= v_{\mathrm{Whl,A}} \cdot \frac{c_{\lambda}}{c_{\lambda} - \dfrac{F_{\mathrm{B,A}}}{F_{\mathrm{N,A}}}} \end{aligned} \tag{10}$$

where μ_{A} is the coefficient of friction between the tire and the road, $F_{\mathrm{B,A}}$ is the brake force, $F_{\mathrm{N,A}}$ is the normal force from the wheel on the road, c_{λ} is the slope of the μ-slip curve at its origin, $v_{\mathrm{WhlFre,A}}$ is the free rolling wheel velocity, and $v_{\mathrm{Whl,A}}$ is the measured wheel velocity, all during the adaptation time period. The free rolling wheel velocity is equal to the vehicle velocity in the direction of the wheel plane. Using the values of the yaw velocity, the steering wheel angle, the lateral velocity of the vehicle, and the geometrical data of the vehicle, this free rolling wheel velocity can be transformed to the center of mass. Using the transformed velocity in the longitudinal direction of the vehicle as the input to a Kalman filter, the longitudinal velocity v_{X} of the vehicle can be estimated. The estimated longitudinal vehicle velocity can then be transformed back to the locations of the four wheels to result in the free rolling wheel velocities of all four wheels. Thus, the actual slip values can be computed for all four wheels during ABS control. Usually, the left and right rear wheels of the vehicle are selected in an alternate manner. Then the stopping distance of the vehicle is not so much increased by the lower brake pressure during the adaptation time period, and the vehicle stability is improved.

For the Kalman filter, the differential equation of the longitudinal velocity of the vehicle is used in which the very small Coriolis acceleration due to the lateral and yaw velocity of the vehicle is neglected.

$$\begin{aligned}\dot{v}_{\mathrm{X}} = {} & \frac{1}{m} \\ & \cdot \left\{ \left(F_{\mathrm{S,FL}} + F_{\mathrm{S,FR}}\right) \cdot \sin\delta - \left(F_{\mathrm{B,FL}} + F_{\mathrm{B,FR}}\right) \cdot \cos\delta - \left(F_{\mathrm{B,RL}} + F_{\mathrm{B,RR}}\right) \right\} \\ & - \frac{c_{\mathrm{W}} \cdot A \cdot v_{\mathrm{X}}^2 \cdot \rho}{2 \cdot m} - \dot{v}_{\mathrm{X,\,offset}}.\end{aligned} \tag{11}$$

$$\ddot{v}_{\mathrm{X,\,offset}} = 0 \tag{12}$$

in which the subscripts FL, FR, RL, and RR denote the front left wheel, the front right wheel, the rear left wheel, and the rear right wheel, respectively, c_{W} is the drag coefficient, A is the projected area of the vehicle, ρ is the air density, and the time derivative of $v_{\mathrm{X,offset}}$ is an offset in the vehicle acceleration induced by the inclination of the road. Equation 12 indicates that it is assumed that the inclination of the road changes slowly in time. The inclination of the road can then also be estimated by the Kalman filter (Isermann 2006).

Figure 18 shows measured time histories of ESC during braking with ABS control in which the alternating adaptation time periods at the rear wheel pressures are clearly seen.

4.5 Safety Concept

ESC is a complex mechatronic system which influences the brake system which is a safety system of the vehicle. Therefore, the requirements on ESC with regard to failure rate and reliability are very high. Redundancy of components to increase the system safety and reliability must be avoided for cost reasons. Instead, components with inherent safety and component monitoring are used. Safety in this context does not mean the decrease of accidents in traffic situations by ESC but the safety of the vehicle in traffic situations in case of an ESC-component failure. Figure 19 shows the components considered in the total system driver-vehicle-ESC (Isermann 2006).

In order to increase the safety of the system, the actions of the driver are included in plausibility considerations. For instance, sudden changes in the measured value of the steering wheel angle can be compared with those a driver can maximally physically apply. The engine and the transmission are included in the system together with their controllers, since both of them are influenced by ESC. As shown in Fig. 19, the connections between the components are part of the total system and must also be monitored.

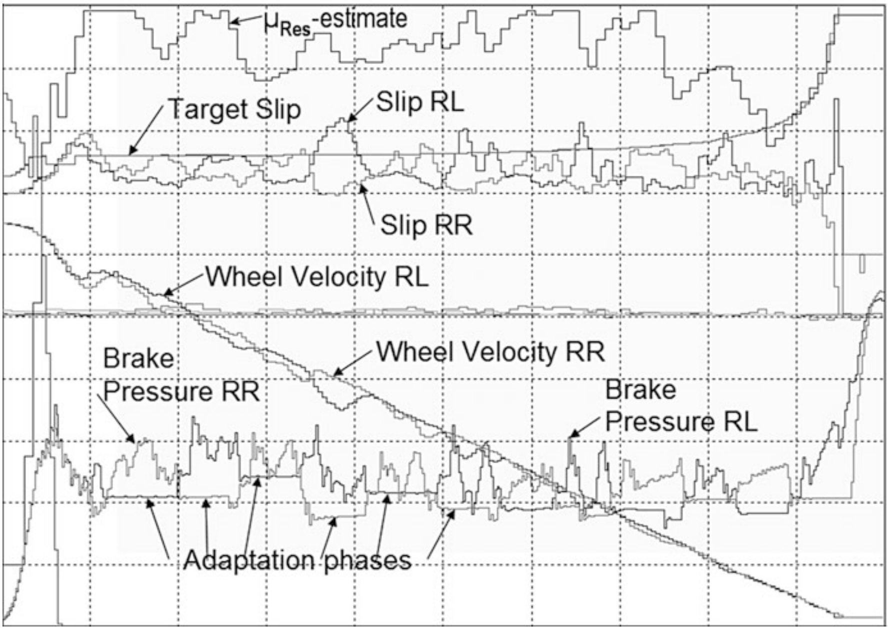


Fig. 18 Straight-line ABS braking with ESC from 120 km/h initial vehicle velocity on a smooth dry asphalt road. Plotting limits: time, 0–4.2 s; slip, −0.7 to +0.3; wheel velocity, 0 to 50 m/s; brake pressure, 0 to 250 bar

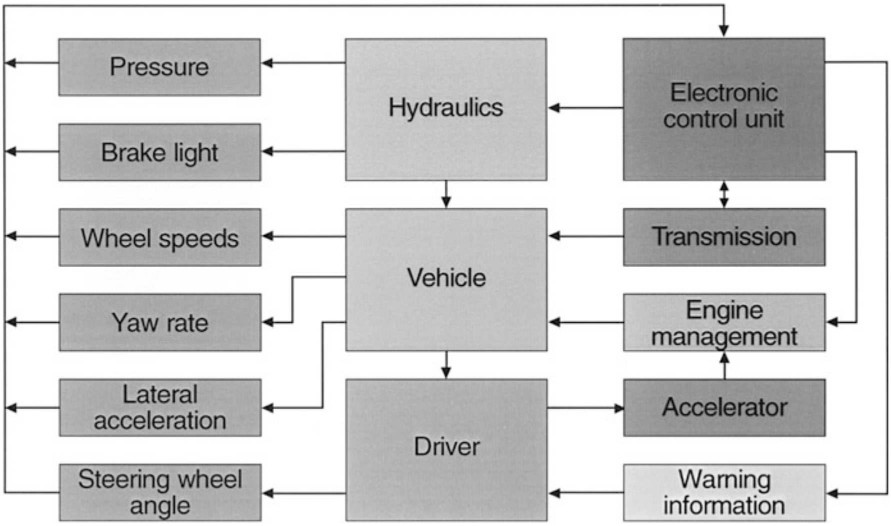


Fig. 19 Total system driver-vehicle-ESC considered for system safety

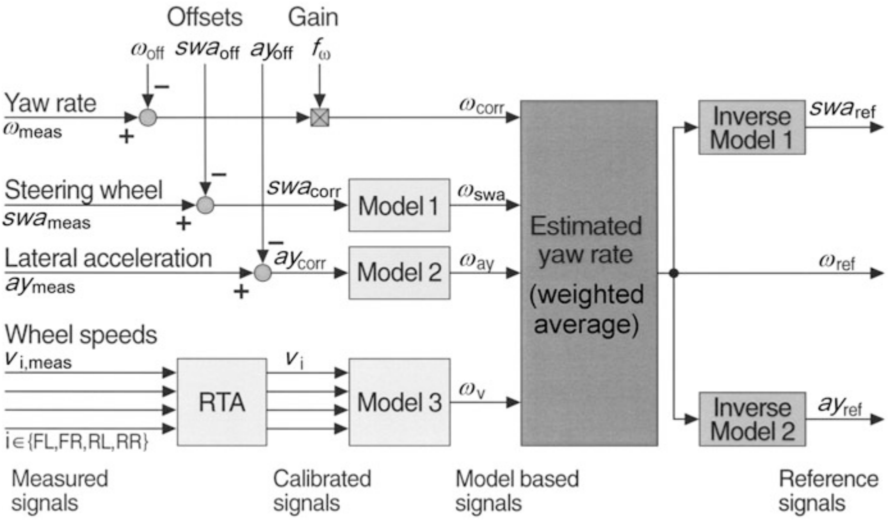


Fig. 20 Model-based monitoring of the yaw rate sensor, the steering wheel angle sensor, and the lateral accelerometer

For the development of the safety concept, several methods are used. Such methods are the failure mode and effect analysis (FMEA) and fault tree analysis (FTA). A new method for monitoring and calibration of the sensors was included with the introduction of ESC. This method uses models for the monitoring of the steering wheel angle sensor, yaw rate sensor, and the lateral accelerometer and is called model-based sensor monitoring (Isermann 2006). This method is also called analytic redundancy.

Figure 20 shows the concept of the model-based sensor monitoring. The measured and during driving continuously calibrated sensor signal values of the steering wheel angle swa_{corr}, the lateral acceleration ay_{corr}, and the wheel velocities v_{i} are used as inputs for models whose outputs are the yaw velocity of the vehicle. Model 1 is the bicycle model of the vehicle, model 2 is $a_{\mathrm{y}}/v_{\mathrm{X}}$, and model 3 is $(v_{\mathrm{WhlFre,FL}}\text{-}v_{\mathrm{WhlFre,FR}})/t_{\mathrm{F}}$ for the front wheels and $(v_{\mathrm{WhlFre,RL}}\text{-}v_{\mathrm{WhlFre,RR}})/t_{\mathrm{R}}$ for the rear wheels, where t_{F} is the tread of the front axle and t_{R} the tread of the rear axle. Model 3 for the front axle with consideration of the steering angle is used for a rear-wheel-drive vehicle, model 3 for the rear axle is used for a front-wheel-drive vehicle, and a weighted mean value of the model outputs of the front and rear axles is used for four-wheel-drive vehicles.

Before model 3 can be used, a tire tolerance compensation (RTA) must be finished, which assures that during straight-line free rolling, all wheels show the same velocity. For the calibration of the other sensor signal values, calibration values can be found in EEPROM. The measured values of the yaw velocity, the steering wheel angle, the lateral acceleration, and the wheel velocity are ω_{meas}, swa_{meas}, ay_{meas}, and $v_{\mathrm{i,meas}}$, respectively, where the subscript “i” denotes the

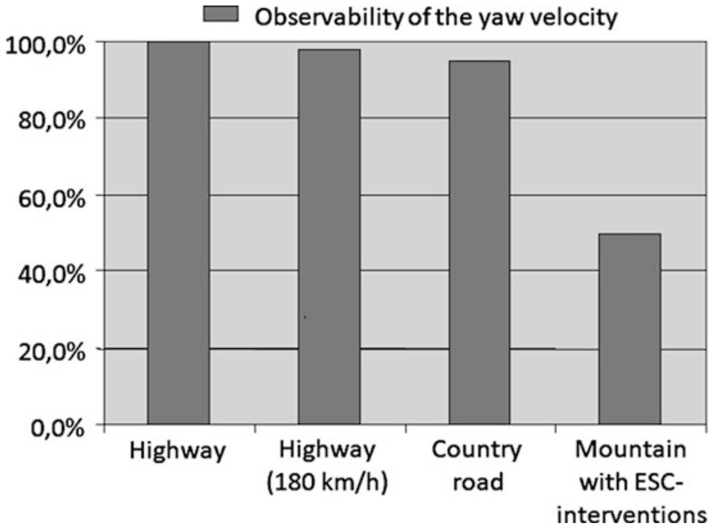


Fig. 21 Observability of the yaw rate sensor dependent on the route

different wheels. The offset values of the yaw velocity, the lateral acceleration, and the steering wheel angle are ω_{off}, ay_{off}, and swa_{off}, respectively. The sensitivity error of the yaw rate sensor is f_{ω}, which is stored in EEPROM together with the offset values. The subscript "corr" refers to the calibrated signal values.

The signal values ω_{swa}, ω_{ay}, and ω_{v} are estimates of the yaw velocity on the basis of the calibrated steering wheel angle sensor signal, the calibrated lateral accelerometer signal, and the compensated wheel velocity sensor signals, respectively. The computation of a weighted average of the four yaw velocity values in which the distance between the values and the distance between the time rate of change of the values are considered to compute the weight of each value results in the reference yaw velocity value ω_{ref}. This value of the yaw velocity is very reliable and delivers very good estimates for the true yaw velocity of the vehicle even during yaw moment control. From the reference yaw velocity, reference values of the steering wheel angle swa_{ref} and of the lateral acceleration ay_{ref} can be computed using an inverse model 1 and an inverse model 2 for the steering wheel angle and the lateral acceleration, respectively. These reference values can be used for monitoring the steering wheel angle sensor and the lateral accelerometer.

If the vehicle approaches the physical limit, the accuracy of the models loses precision. However, if the vehicle motion is still stable, the models compute yaw velocity values which can still be used for sensor failure monitoring. If the computed values of the yaw velocity ω_{swa}, ω_{ay}, and ω_{v} differ little, i.e., the weights of these values are large, then the signal of the yaw velocity sensor is called "observable," and sensor failure monitoring using the models can be continued. During driving, this is very often the case (Fig. 21).

The offset in the yaw velocity sensor signal is very important, because the outside of the turn is determined by the sign of the yaw velocity. As long as the offset value is not accurately determined, the dead zone of the vehicle motion controller is increased which makes the control less sensitive but which reduces the possibility of unintended ESC interventions.

Failures cannot always be detected early in time. The effect on the vehicle behavior because of some failures which are not timely detected can be limited by the following:

- **Banked turn logic** (Isermann 2006). As an example, if the lateral accelerometer signal is stuck at zero, then it might be that before the failure is detected, cornering of the vehicle is interpreted as "skidding on ice" and an ESC intervention should occur. However, this is a situation which can also occur while cornering in a banked turn. If the driver does not countersteer, then cornering of the vehicle is interpreted as cornering in a banked turn and no unintended ESC intervention occurs. After detection of the failure in the lateral accelerometer signal, ESC is switched off.
- **Monitoring the time duration of the ESC interventions**. Since the ESC interventions are of short-time duration (usually less than 0.5 s), then a failure is suspected if the time duration of the ESC intervention occurs for a longer time period and then ESC is switched off.
- **Monitoring the time duration of ABS control**. Also, ABS control is physically limited in time duration. Therefore, a failure may be expected, and the system may be switched off if the ABS control lasts longer than a certain time period.
- **Plausibility between the signals of the brake light switch and the brake pressure sensor**. If the brake light switch is on for a longer time period although the pressure sensor signal remains zero, then the system is switched off.

5 Value-Added Functions

5.1 Special Stability Support

This group of value-added functions aims at discovering the tendency of the vehicle for instability and modifying the brake pressure, like with rollover mitigation.

5.1.1 Extended Understeer Control, EUC

If the handling of the vehicle becomes too much understeer, then ESC corrects the behavior with a slip intervention at the rear wheel on the outside of the turn. This makes the vehicle behave less understeer and the lateral forces at the rear axle increase because of the larger slip angle. Since this function is not always sufficient to keep the vehicle on track, EUC can reduce the engine torque and in addition also actively brake all wheels and thus reduce the vehicle velocity and thus also the centrifugal force on the vehicle considerably. The function starts if the driver wants to reduce the radius of the turn by more than that which is physically possible according to the coefficient of friction of the road and the momentary vehicle velocity.

5.1.2 Load Adaptive Control Mode for LCV/Vans, LAC

LAC includes an analysis of the vehicle handling (characteristic velocity v_{ch} of the bicycle model) and an estimation of the vehicle mass. The latter is based on Newton's second law of the vehicle motion in longitudinal direction during propulsion. Since the traction forces are known (estimated by the engine management system) and the vehicle acceleration is derived from the wheel velocities, the vehicle mass can be computed. With this information, some of the basic ESC functions as well as the function ROM can be adjusted and improved if the vehicle is loaded.

5.1.3 Roll Movement Intervention, RMI

Vehicles that do not tend to roll over during stationary cornering may use a reduced function ROM. For these vehicles, the stationary parts of the ROM function can be omitted, while those parts that deal with dynamic vehicle maneuvers are kept. The reduced function is called RMI and is simpler to apply than ROM.

5.1.4 Rollover Mitigation, ROM

Most vehicles do not roll over in daily traffic. Rollover is typical for vehicles that show a combination of a high center of mass and a soft suspension, as is the case with off-road and light commercial vehicles. There are two different situations in which rollover may occur:

- The rotation of the steering wheel is very dynamic (like with fast lane changes).
- The steering wheel angle continuously increases during limit cornering.

The key issue is the detection of an incipient rollover motion of the vehicle. Detecting rollover instability just on the basis of the lateral acceleration is not sufficient. However, by using additional signals, it is possible to reliably detect rollover critical vehicle situations. The rollover detection includes a prediction which uses the time rate of changes of the steering wheel angle and the lateral acceleration. For the detection, allowances are computed (also called "offsets") which are added to the measured lateral acceleration to result in an "effective lateral acceleration" $a_{\mathrm{Y,eff}}$ (Fig. 22).

If the effective lateral acceleration exceeds some limit value, an incipient rollover is suspected, and, for instance, brake slip interventions at the front and rear wheels on the outside of the turn follow. Both the side force on the vehicle and the velocity of the vehicle are thus quickly reduced. Even if an accident still occurs, the damage will be reduced because of the reduced vehicle velocity. Since the rollover motion is fast, there is not much time left for rollover detection and intervention.

At highly dynamical situations like fast lane change maneuvers, the rollover risk is larger than for stationary maneuvers. Therefore, the time derivative of the lateral

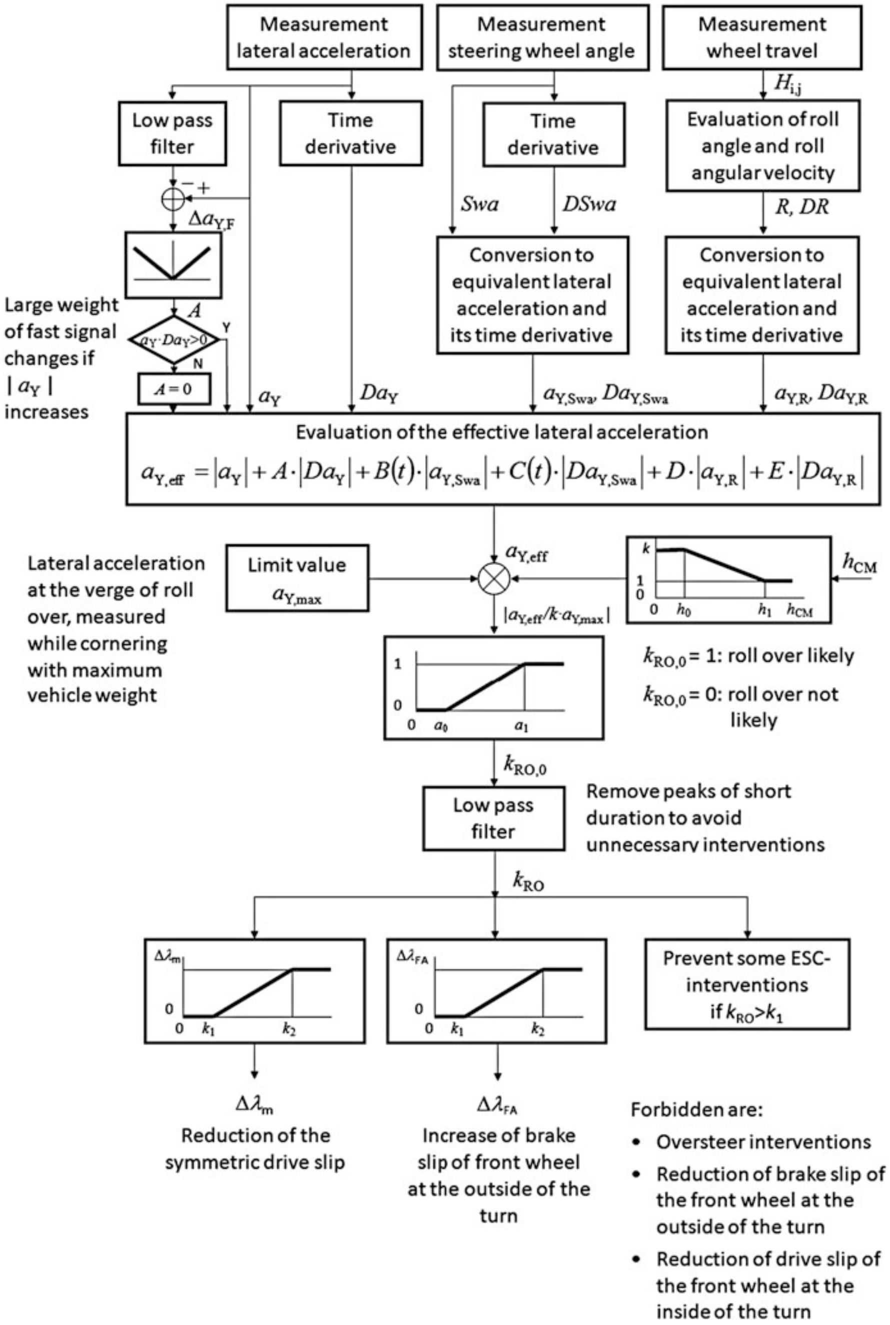


Fig. 22 Block diagram of the rollover function

acceleration Da_Y and the time derivative of the steering wheel angle $DSwa$ are included in the determination of the value of the rollover risk index k_{RO}. If the lateral acceleration a_Y is large and increasing, then the risk of rollover increases in time. In this situation, the product of a_Y and Da_Y is positive. Only if this product is positive, the time derivative of Da_Y is included in an offset. Fast changes of the lateral acceleration are indicated by the difference $\Delta a_{Y,F}$ between the lateral acceleration a_Y and its low-pass filtered value.

The steering wheel angle Swa and its time derivative $DSwa$ are both used to generate offsets. A linear bicycle model is used in order to obtain equivalent values in the lateral acceleration $a_{Y,Swa}$ and $Da_{Y,Swa}$. These equivalent values are multiplied by parameters $B(t)$ and $C(t)$, respectively, which are time dependent such that the offsets decay in time.

Off-road vehicles sometimes have sensors with which the travel of each wheel suspension $H_{i,j}$ can be measured. Here, the subscript i denotes the front or rear axle, and the subscript j denotes the left or the right wheel on the axle. Using this information, the roll angle of the vehicle R and its time derivative DR can be evaluated. Using the roll stiffness of the vehicle, the equivalent lateral acceleration $a_{Y,R}$ and its time derivative $Da_{Y,R}$ can be evaluated. These values are multiplied by parameters D and E, respectively, and then used as offsets for the computation of the effective lateral acceleration $a_{Y,eff}$.

The effective lateral acceleration $a_{Y,eff}$ is compared with a limit value of the lateral acceleration $a_{Y,max}$ at which the vehicle is on the verge of rollover. This limit value $a_{Y,max}$ is obtained by experiment at stationary cornering with the fully laden vehicle. The vehicle is laden such that the height of its center of mass h_{CM} is maximal, h_1. If the vehicle is only partly laden, then h_{CM} will be less than h_1, and the risk of rollover will be less than that with the fully laden vehicle, and $a_{Y,max}$ can be increased by a factor k for the partly laden vehicle. The vehicle mass is estimated during driving and from this h_{CM} is estimated. Using a linear relationship, the factor k is evaluated dependent on the height of the center of mass h_{CM}.

The ratio between the effective lateral acceleration $a_{Y,eff}$ and the weighted maximum acceleration $k{\cdot}a_{Y,max}$ results in a first indication of the rollover risk. If this ratio is smaller than a_0, then rollover is unlikely. If this ratio is larger than a_1, then rollover is highly likely. In between those two values a_0 and a_1, the likelihood of rollover increases linearly with the ratio. Thus, a first value of the risk of rollover $k_{RO,0}$ is obtained. However, this value is truncated to result in a value between 0 and 1. Furthermore, since the signal from the accelerometer is corrupted by noise, the value $k_{RO,0}$ is filtered by a low-pass filter to finally result in the rollover risk factor k_{RO}.

The value of the risk factor k_{RO} is now used to control the slip values at the wheels. If the value of k_{RO} exceeds the value k_1, then the symmetric slip of the driven wheels λ_m is reduced proportional to the value $k_{RO}-k_1$ by the value $\Delta\lambda_m$, while the reduction is limited for values of k_{RO} larger than k_2. The reduction of the symmetric slip will reduce the engine torque. Similarly, if the value of k_{RO} exceeds the value k_1, then the brake slip of the front wheel at the outside of the turn is increased proportional to the value $k_{RO}-k_1$ by the value $\Delta\lambda_{FA}$, while the increase is

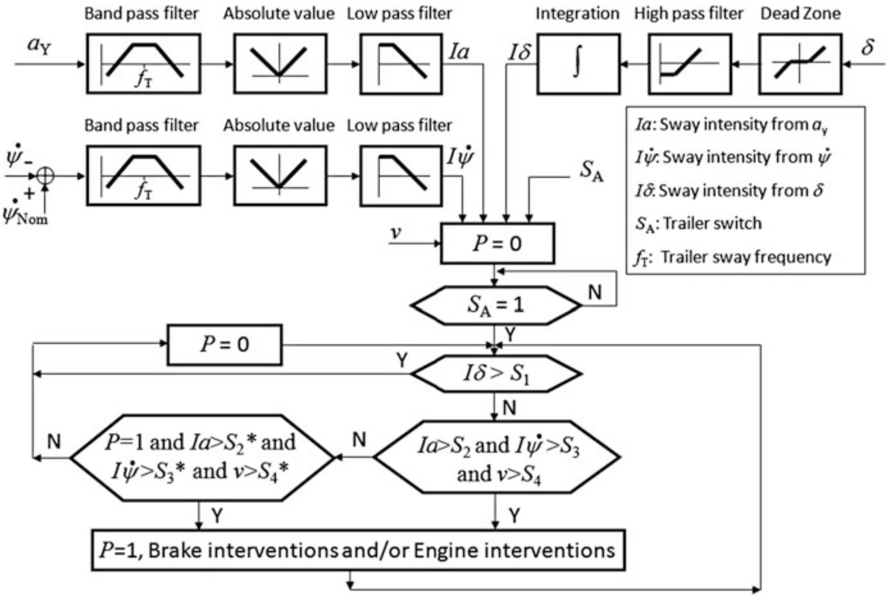


Fig. 23 Block diagram of trailer sway stabilization TSM

limited for values of k_{RO} larger than k_2. This slip increase has an understeer effect on the vehicle. Destabilizing ESC interventions like oversteer interventions, reduction of brake slip of the front wheel at the outside of the turn, and reduction of traction slip of the front wheel at the inside of the turn are forbidden if $k_{RO} > k_1$.

5.1.5 Trailer Sway Mitigation, TSM

Trailers that are towed by a vehicle may show sway oscillations if a certain critical velocity v_{crit} (e.g., 80 km/h) is exceeded. These trailer sway oscillations impose yaw moments on the towing vehicle which then shows oscillations in the yaw velocity of the same frequency as the trailer sway oscillation f_T. With increasing velocity of the train, the sway oscillation as well as the yaw oscillation becomes stronger. The yaw velocity of the vehicle may become so large that ESC interventions on the vehicle result. The detection of trailer sway oscillations is based on a frequency analysis of the yaw velocity of the vehicle. If trailer sway is detected, then the train is decelerated by active braking of the vehicle. The trailer sway oscillation decays by itself if the velocity of the train is lower than the critical velocity v_{crit}. The principle of TSM is shown by the block diagram in Fig. 23.

If trailer sway oscillations are detected, a sway indicator P is set to the value 1. At vehicle start, the sway indicator P is set to the value 0 (no sway oscillations). If the value of the trailer switch S_A is not 1, then it is presumed that there is no trailer connected to the vehicle, and further processing is stopped. Otherwise, the steering wheel angle is observed. If the driver steers violently, an oscillation in the yaw velocity of the vehicle may result which must not be interpreted as a trailer sway oscillation. This is the case if steering indicator $I\delta$ is larger than the limit value S_1. Otherwise, it is checked if the train velocity is larger than the activation limit value S_4, since at low velocities, trailer sway oscillations do not occur. If this is the case, the intensity values from the yaw velocity oscillation and of the lateral acceleration oscillation of the vehicle are checked to be larger than their activation limit values S_3 and S_2, respectively. If that applies, then the sway indicator is set to the value $P = 1$, and the engine torque is reduced in order to decelerate the train with a deceleration of -0.3 g. If the required deceleration cannot be achieved by an engine intervention alone, then all wheels are actively braked by the same brake pressure. The interventions in the engine torque and in the wheel brake pressures continue as long as the intensity values of the yaw velocity, the lateral acceleration, and the train velocity are larger than deactivation limit values S_3*, S_2*, and S_4*, respectively. If during the train deceleration of -0.3 g the intensity indicators of the yaw velocity and of the lateral acceleration increase further, then the pressure in the brake wheel cylinders is increased until the deceleration has reached the value of -0.5 g.

The suppression of trailer sway can be improved further by a sidewise modulation of the wheel brake pressures at the axles. The resulting yaw moments on the vehicle then counteract the yaw moments from the trailer on the vehicle and thus reduce the oscillation of the yaw velocity directly. It has been shown that then the trailer sway oscillation is also reduced. It is important that the delay in the yaw moment modulation is small. Otherwise, the yaw velocity oscillation may increase instead of decrease. A delay in the yaw moment occurs if during the brake pressure modulation the pressure in the brake wheel cylinder is reduced to zero. In that case, dead times are introduced in the brake torque generation. These dead times occur, because if the pressure at a brake wheel cylinder is reduced to zero, the brake pad is retracted from the brake disk. If the pressure is increased again, the brake pad has to be moved back again to the brake disk which requires some time and which causes the delay.

5.1.6 Secondary Collision Mitigation, SCM

Vehicle crashes often result in vehicle motions which are very difficult to be handled by the driver and which may lead to a follow-up collision. On the one hand, the driver may be caught by surprise so that a definite reaction cannot be expected for a longer period of time. On the other hand, the driver may be injured and may not be able to react appropriately. A case study has shown, that in almost all crashes, except for frontal crashes, it is appropriate to fully apply the brakes. This can be realized by coordinating the airbag with ESC. ESC can then by active braking and active yaw moment generation reduce the severity of the secondary collision or even avoid it. The intervention is an automatic emergency braking maneuver with which the vehicle comes to a stop in a stable manner. The intervention occurs on the basis of the information “collision intensity” and “collision situation” from the airbag ECU. Using this information, the required vehicle deceleration level is evaluated in the airbag ECU. In the ESC ECU, the required brake pressures in the wheel brakes are computed in order to achieve this level of deceleration. If the driver applies the brake pedal, then the intervention is interrupted. However, it must then be assured that the driver does not apply the brakes unintentionally, e.g., by inertia forces of his body during the collision. The analysis of this situation is based on the time histories of the application of the brake pedal and accelerator pedal.

5.1.7 Side-Wind Assist, SWA

Sudden side-wind gusts may influence the lateral vehicle motion and lead to substantial deviations in the vehicle path. Such situations can be detected by using the sensors of ESC. All ESC sensor signals are used for the evaluation (yaw velocity, lateral acceleration, brake pressure, wheel velocities). In this evaluation, the yaw moment is computed which is required to reduce this unintentional lateral dynamics of the vehicle. The yaw moment is generated by ESC by brake slip interventions that are already described in Sect. 4. Through the reduction of the lateral dynamics of the vehicle, handling becomes easier, and the effort required from the driver to keep the vehicle in lane is reduced. Furthermore, the driver has more time for his own maneuvering tasks (Keppler et al. 2010).

5.2 Special Torque Control

This category of value-added functions supports the driver with stabilization, lane keeping, and propulsion of the vehicle and improves the handling of the vehicle.

5.2.1 Dynamic Center Coupling Control, DCT

Four-wheel-drive vehicles with a controllable multi-disk clutch as a longitudinal lock between the driven axles offer an alternative to braking an axle. Instead of braking both spinning wheels at one axle to reduce the propulsion torque on that axle, the multi-disk clutch can be closed or opened, which is also of advantage from an energy economy point of view. Additional features of the controlled multi-disk clutch are:

- Improved performance at straight-line driving, on μ-split road surfaces, and at off-road.
- The clutch must be open during active ESC interventions (otherwise, slip interventions at one wheel may lead to slip changes at other wheels), during braking (otherwise, the electronic brake force distribution is corrupted), and during the adaptation phases for the computation of the vehicle velocity.
- If a visco-clutch is used, the unintended lock torque (which may occur in lifetime) must not be larger than 100 Nm. Otherwise, the computation of the vehicle velocity is corrupted.
- For rear-wheel-drive vehicles with optional four-wheel drive, the lock torque must be reduced if the vehicle understeers and increased if the vehicle oversteers.

- For front-wheel-drive vehicles with optional four-wheel drive, the lock torque must be increased if the vehicle understeers and reduced if the vehicle oversteers.
- For comfort reasons, the lock torque is modulated by a ramp function. The interventions are more comfortable than brake interventions. Therefore, the control of the vehicle dynamics using the multi-disk clutch can be tuned more sensitive than if the brakes were applied. Brake interventions are then less frequently required.

5.2.2 Off-Road Detection and Measures, ORD

It is well known that on loose road surfaces like gravel, the largest traction and braking forces are reached at large wheel slips, which is often the case off-road. For those roads, the stopping distance during ABS may be reduced if the target slip λ_T is increased as compared with that on normal roads (Fischer and Müller 2000). In the function, it is assumed that the road surface is loose if off-road is detected. Therefore, the off-road detection function was introduced. In the off-road detection function, the frequency spectra of the wheel velocities are analyzed. If the wheel velocities show spectral components of high frequency and if the amplitudes of these spectral components are also high, then off-road is presumed. Short-time periods with large high-frequency components in the wheel velocity must not immediately lead to off-road detection. For this reason, an off-road counter is used. The counter is increased in time if large high-frequency components exist and decreased in time if they vanish. If the counter reaches a limit activation value, then off-road is detected. For safety reasons, the target slip λ_T of the front wheels is increased from their normal values only for vehicle velocities below some limit value (e.g., 50 km/h). Sometimes, it is difficult to differentiate between ABS on scraped ice and ABS off-road. The differentiation is based on the relation between wheel slip and vehicle deceleration. If the vehicle deceleration is small but the wheel slip is large, then ice is detected, and the target slip λ_T is not increased. If the relation is not clear, the target slip λ_T is increased only at one front wheel. If the driver turns the steering wheel, the target slips are immediately reduced to the normal ones. The target wheel slips λ_T at the rear axle are not increased from their normal values.

Figure 24 shows an acceleration of a vehicle on a gravel road with a small slope with a successive full braking with ABS. During the acceleration, large oscillations occur in the wheel velocities which indicate an off-road situation. During the acceleration, the oscillations are analyzed and used in the off-road detection. After approximately one second, the off-road counter reaches the activation limit value, and off-road is detected. In this special maneuver, the relation between the wheel slip and the vehicle deceleration was not clear, so that it was not clear if the situation “scraped ice” or “off-road” applies. Therefore, the target slip λ_T was increased only at the right front wheel and not at the left front wheel. The increase $\Delta\lambda_{\mathrm{Whl,FR}}$ in the target slip λ_T of the right front wheel can be seen in the upper part of Fig. 24.

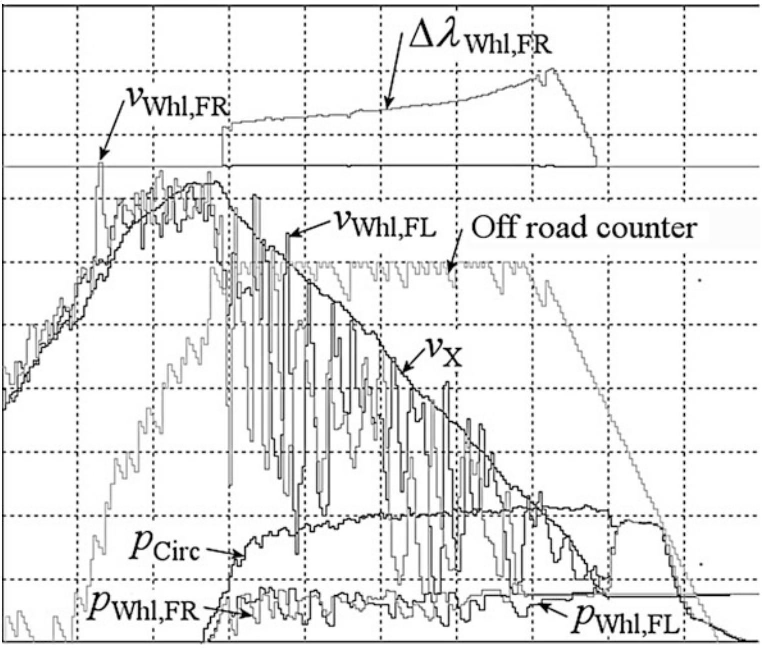


Fig. 24 Full braking off-road with off-road detection and increase of the target brake slip at the right front wheel. Plotting limits: time, 0–5 s; wheel/vehicle velocity, 0 to 10 m/s; brake pressure, 0 to 500 bar; off-road counter, 0 to 100; slip, −150 % to +50 %

5.3 Brake and Boost Assist

In this category of assistance functions, the brake pressures and the brake boost functions are adjusted to the driving and system situations, for instance, with the brake assistant.

5.3.1 Hydraulic Brake Assist, HBA

Investigations with the driving simulator of Mercedes-Benz showed that normal drivers brake reluctantly in frightening situations (Fig. 25). The full application of the brake pedal by the driver after an initially fast but low brake pedal force application occurs after some delay. This delay has a large influence on the stopping distance since at the beginning of the brake application the vehicle velocity is the highest during braking. The brake assistant (BA) overcomes the delay and thus reduces the stopping distance considerably. Key to the brake assistant is the detection of an emergency situation and the immediate increase of the brake pressure at all wheels beyond that induced by the driver up to ABS operation at all wheels.

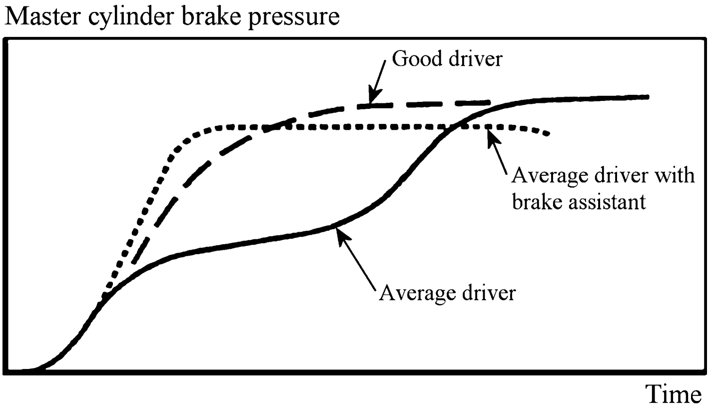


Fig. 25 Support of the driver at the start of braking by the brake assistant

The most important functional requirements on the brake assistant are the following (Robert Bosch GmbH 2004; Breuer and Bill 2013):

- Support of the driver in emergency braking situations and reduction of the stopping distance to those values that can usually be achieved by expert drivers only.
- Interruption of full braking as soon as the driver reduces the pedal force by a substantial amount.
- Conservation of the conventional brake boost function. Pedal feel and comfort should remain untouched during normal braking.
- Activation of the function only in real emergency situations, so that the driver does not get used to the function.
- No impairment of the conventional brake if the brake assistant function fails.

The hydraulic brake assistant uses the hydraulic unit of ESC to actively increase the brake pressure to values beyond that induced by the driver. Using the pressure sensor signal of ESC, the brake pedal application by the driver is analyzed. The detection of an emergency situation is based on the value of the pressure and on the value of its time derivative (Table 1). The HBA can be easily adapted to the properties of the vehicle and of the brake system by the choice of the activation limit values of the brake pressure and of its time derivative. The activation limit values are dynamically adjusted to the actual situation which includes the vehicle velocity, brake master cylinder pressure, state variables of the pressure control in the brake wheel cylinders, and an analysis of the time history of the braking maneuver. Also, the vehicle velocity must exceed a least velocity to activate the brake assistant function.

As soon as the activation conditions are fulfilled, the brake assistant is activated (number 1 in phase 1 in Fig. 26). Now, the brake assistant increases the brake

Table 1 HBA-Logic (See Fig. 26)

Situation	Detection logic
Phase 1 Emergency situation Panic braking	Brake pedal actuated and MC pressure gradient larger than activation limit value and MC pressure larger than activation limit value and vehicle velocity larger than activation limit
Phase 2 Reduction of brake pressure request	Pedal force (derived from MC pressure) below deactivation limit
Renewed activation	MC pressure gradient larger than activation limit
Standard braking	Brake pedal not actuated or MC pressure below deactivation limit or vehicle velocity below deactivation limit or pedal force large enough

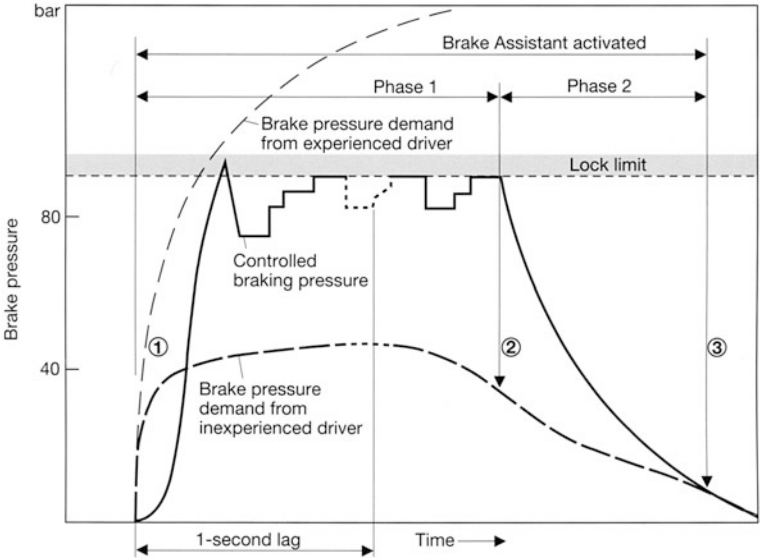


Fig. 26 Concept of the hydraulic brake assistant, HBA

pressure beyond the level induced by the driver at all four brake wheel cylinders up to ABS control at all wheels. The active brake pressure increases, and the brake pressure control occurs in a similar manner as with active interventions of ESC. As soon as the brake pressure passes the locking level, the brake slip controller ABS is started, and the brake slip is controlled for optimal brake forces.

If the driver releases the brake pedal and if the brake master cylinder pressure has dropped below a certain deactivation level (number 2), the system recognizes the driver's desire and can reduce the pressure in the brake wheel cylinders (Fig. 26, phase 2). From the beginning of phase 2 on the control strategy is changed. Now, the pressure in the brake wheel cylinders must follow the signal from the pressure sensor in a smooth and comfortable transition back to the standard brake function. The brake assistant is switched off as soon as the enhanced pressure in the brake wheel cylinders reaches the value intended by the driver or falls below a deactivation limit value (number 3). The driver can now continue braking without the additional support of the brake assistant.

5.3.2 Brake Disk Wiping, BDW

Wet brake disks show a coefficient of friction between the brake lining and the brake disk which is lower than that of dry brake disks. If the brake disks are wet because it rains, then a small pressure in the brake wheel cylinders is actively generated and increased up to a very small level (approx. 1.5 bar) for a short period of time (approx. 3 s). The brake pad then removes the water film from the brake disk and thus improves the coefficient of friction between the brake lining and the brake disk. This procedure is repeated regularly (approx. every 3 min). During this procedure, the induced vehicle deceleration is so small that it is not noticed by the driver. Rain is signaled by the rain sensor. Another indication of rain is the operation of the wipers. The function is interrupted if the driver pushes the brake pedal. For this function, the hydraulic unit must be equipped accordingly, e.g., with precise control valves.

5.3.3 Electronic Brake Prefill, EBP

If the driver applies the brake pedal or if ESC actively generates pressure at the brake wheel cylinders, the brake pads move to the brake disks first before the pressure can be increased. During this time period, the brake torque at the wheels is zero. The brake force generation shows a delay from the beginning of braking (see also TSM), and the stopping distance increases with the delay. The ESC intervention is also delayed and stabilization of the vehicle may not be possible.

The situation can be improved if the brake pads are already in contact with the brake disks at the time the driver applies the brake or at the time the ESC intervention starts. Indicative for an expected fast brake pedal application is a fast release of the accelerator pedal. If this happens, a small pressure of approx. 3 bar at the brake wheel cylinders is generated actively by the ESC hydraulic unit. This pressure will already move the brake pads to the disks before the driver applies the brake.

An example of the use of this function with active ESC interventions shows the following situation. The driver steers the vehicle fast to the left, e.g., to avoid an obstacle. During this fast steering maneuver to the left, the pressure in the brake wheel cylinder at the front left of the vehicle is increased actively up to a small level of approx. 3 bar. This pressure value will push the brake pad at the front left wheel to the brake disk. If the driver steers the vehicle subsequently fast to the right, then it is highly likely that the vehicle behavior becomes oversteer to the right and unstable. In this situation, an ESC intervention at the left front wheel is required. Since the brake pad at the left front wheel is already pushed to the brake disk, there is no delay in the brake torque generation at that wheel brake. This improves the performance of ESC. However, if the driver does not rapidly steer to the right after the fast steering maneuver to the left, then the ESC intervention at the right front wheel is not required and may irritate the driver. Because of this ambiguity, the pressure in the front left brake wheel cylinder must be kept small and just sufficient to push the brake pad to the brake disk. For this function, the hydraulic unit must be equipped accordingly, e.g., with precise control valves.

5.3.4 Hydraulic Brake Boost, HBB

By far, the most brake actuations occur with brake pressures below 30 bar. For these situations, a small vacuum booster is sufficient to support the driver. A small vacuum booster is of advantage for packaging reasons. However, brake boosting must also insure that the driver is sufficiently supported during emergency braking with high brake pressures. In order to satisfy both requirements, the small vacuum booster is augmented with ESC boosting. ESC boosting starts if the small vacuum booster has reached saturation. Like with the hydraulic brake assistant, the ESC hydraulic unit can actively increase the brake pressure at the wheels. The harder the driver pushes the brake pedal (which is seen from the pressure sensor signal), the longer the recirculation pump is activated, the more brake fluid flows to the brake wheel cylinders, and the more the pressure in the brake wheel cylinders increases. For most brake actuations (brake pressure <30 bar), the function is not active, since the small booster supports the driver sufficiently. Only for very few brake actuations with large brake pressures, the function may become active, and the driver may notice some pedal feedback. Moreover, the function can compensate situations in which boosting is limited because of a low level of vacuum or even because of a complete failure of the vacuum booster.

5.3.5 Hydraulic Boost Failure Compensation, HFC

If the brake booster fails, the hydraulic unit of ESC can be used to support the driver with braking the vehicle with sufficient brake pressure. Similar to the HBB function, the pump of the ESC hydraulic unit is used to provide the brake wheel cylinders with sufficient brake fluid and to achieve the vehicle deceleration required by the driver.

5.3.6 Hydraulic Fading Compensation, HFC

During braking, the temperature at the wheel brakes increases and may reach very high values from which the brake efficiency and the vehicle deceleration suffer (brake fading). In order to keep a constant vehicle deceleration at a constant pressure in the brake master cylinder, the pressure in the brake wheel cylinders must be increased in such situations. For this pressure increase, the pump of the hydraulic unit of ESC is used. HFC supports the driver if at high brake pedal forces (evaluated from the pressure sensor signal) at which usually an ABS control results, the full vehicle deceleration is not achieved. The pump supplies brake fluid to the brake wheel cylinders continuously until the full vehicle deceleration is achieved, i.e., until ABS control at all wheels. If the pressure value in the brake master cylinder has fallen below a certain level, the function is stopped.

5.3.7 Hydraulic Rear Wheel Boost, HRB

Normal drivers tend to keep the force on the brake pedal constant if they feel the start of ABS brake pressure modulation. Because of the stable brake force distribution, ABS modulation often starts at the front wheels first. This occurs, for instance, with straight-line braking on homogeneous roads at vehicle decelerations that are below a certain critical value (see Fig. 9). Thus, the brake force potential at the rear axle is not fully exploited although the situation requires this. Full exploitation of the brake forces at the rear axle can be obtained if the pressure at the brake rear wheel cylinders can be increased above that at the front wheels. This is possible by using the pump of the ESC hydraulic unit. HRB checks if the pressure at the brake front wheel cylinders is controlled by ABS. If that is the case but the pressure at the brake rear wheel cylinders is not controlled by ABS, then the pressure at the brake rear wheel cylinders is continuously actively increased by the pump delivery like with ESC interventions during partial braking. As a result, the rear wheel brakes will also start ABS control. The active pressure increase at the brake rear wheel cylinders is stopped if the front wheel brakes are no longer controlled by ABS or if the brake master cylinder pressure has dropped below a certain deactivation value.

5.3.8 Soft Stop, SST

At very low vehicle velocities, the coefficient of friction between the brake lining and the brake disk is larger than at higher vehicle velocities. A jerk can be felt by the driver just before the vehicle stops because of braking. This jerk can be avoided by the driver if the brake pedal is somewhat released just before the vehicle comes to a standstill. Using the control valves of ESC, this procedure can be realized without the action of the driver. Shortly before the vehicle stops, the pressure in the brake wheel cylinders is reduced from that value in the brake master cylinder induced by the driver.

5.4 Standstill and Speed Control

This category of value-added functions supports the driver at road inclination and with standing start, e.g., hill hold control and ACC stop&go. They allow the driver a comfortable ride.

5.4.1 Hill Descent Control, HDC

Off-road vehicles with engaged reduction gear can drive downhill with steep inclinations without braking, just with engine drag torque, and without a significant increase of the vehicle velocity. Off-road vehicles without this reduction gear use automatic braking to achieve the same effect (Fischer and Müller 2000). Part of the automatic braking is the function CDD-B.

HDC can be activated and deactivated by a tip switch at the dashboard. If HDC is activated and the vehicle velocity is below a certain level (e.g., 35 km/h) and the driver gives a little gas (accelerator position <20 %) and a road inclination is detected, then HDC is ready for operation. The time derivative of $v_{X,\text{offset}}$ which is estimated by Eqs. 11 and 12 is used for the evaluation of the road inclination.

Target for HDC is a constant vehicle velocity of 8 km/h. If the driver pushes the accelerator pedal, then HDC increases the target velocity to a higher value which, however, is limited to a maximum of 35 km/h. If the driver applies the brake, HDC reduces the target velocity to a lower value which is limited to a minimum of 6 km/h. Like with CDD-B, the brake lights are switched on during HDC operation.

HDC control is interrupted if the vehicle velocity increases above 35 km/h but continued again if the vehicle velocity drops below 35 km/h again. HDC is deactivated automatically if the vehicle velocity increases above 60 km/h.

During HDC control, the brake temperature may increase to high values which may damage the brake lining. If the temperature of the brakes at both wheels on one axle increases beyond 600 °C, the brake pressure is gradually reduced until the temperature has decreased below 500 °C. If the temperature is below 500 °C, the brakes may be activated again. The brake temperature is estimated using a thermal brake model. In the simulation, the heating time duration is considered as well as the cooling time period during which the brakes are not applied. Input to the simulation is the brake torque which is used to evaluate the generated thermal energy.

In off-road situations with uneven grounds, the normal forces on the wheels may vary significantly during driving. One or two wheels may even lift off from the ground. Because HDC actively applies the brakes, some wheel slips may increase very fast and start ABS control. This may induce large yaw moments on the vehicle which must be compensated by the driver through steering. In order to keep the velocity of the vehicle constant, HDC must then increase the brake pressure at the other wheels which are not yet under ABS control. However, the increase of the brake pressure at the other wheels will increase the yaw moment on the vehicle even further, and the steering task of the driver becomes increasingly difficult. The brake pressure increase may also initiate ABS control at the other wheels. However, the driver can fully concentrate on only his steering task since HDC takes over his task of keeping the vehicle velocity downhill constant.

5.4.2 Automatic Vehicle Hold with Acceleration Sensor, AVH-S

This driver assistance function is used to apply the brakes of the standing vehicle with a hold pressure so that it keeps standing and does not roll away. Using the ESC hydraulic unit, the pressure in the brake wheel cylinders is actively increased up to the hold pressure. Contrary to the function HHC-S, which can keep the hold pressure for only 2 s, the vehicle can be kept standing for several minutes without brake application by the driver. After some time, the hold function is taken over by the automatic parking brake.

For the pressure generation, the pump as well as the separation valve between the brake master cylinder and the brake circuit is stimulated (Breuer and Bill 2013). By reduction of the electrical current, the valve functions as an orifice. The brake fluid flow of the pump generates a pressure at the valve and in the brake wheel cylinders. Since the electrical current of the valve can be varied, it is also possible to vary the pressure in the brake wheel cylinders. Thus, a minimal pressure at the wheels can be set which can be variably adjusted to the longitudinal acceleration sensor signal and which stresses the ESC hydraulic unit minimally. If the pressure is sufficiently high to hold the vehicle, the electrical current of the valve is increased to its maximum value so that it closes and the pump motor can be switched off. The AVH-S function must be activated by the driver by pushing a switch or a tip switch. The brake must be released if after the stop the vehicle should accelerate. If the brake pressure is held by the ESC hydraulic unit, then the pressure is reduced by a reduced electrical current of the valve. As soon as the driver hits the accelerator pedal, the pressure in the brake wheel cylinders is reduced to a value which depends on the actual engine torque and the engaged gear.

5.4.3 Automatic Vehicle Release, AVR

This function allows the controlled reduction of the hold pressure at standstill. It is contained and described in the function AVH-S.

5.4.4 Cruise Control Basic, CCB

In the adaptive cruise control with environment sensor (ACC), the vehicle velocity is first reduced by a reduction of the engine torque. If the vehicle deceleration required by the ACC cannot be achieved by this single intervention, an active brake application is added using the hydraulic unit of ESC (see function CDD-B). For this basic function of the ACC brake pressures, up to 40 bar are required. Since the brake function must comply with high requirements on comfort, special, accurate and continuously controllable separation valves between the brake master cylinder and the brake circuits are required.

5.4.5 Cruise Control Touch Activated, CCT

This function also uses the hydraulic unit of ESC to decelerate the vehicle smoothly. In contrast with CCB, the function CCT offers the driver the possibility to choose arbitrary acceleration and deceleration levels by using control elements at the steering wheel. In addition, it is possible to decelerate the vehicle up to a standstill and to keep it standing by, e.g., AVH-S. This function poses high requirements concerning stress and low-noise emission of the hydraulic unit of ESC.

5.4.6 Controlled Deceleration for DAS Basic, CDD-B

Many assistance functions require a definite vehicle deceleration, e.g., TSM, HDC, ACC, and the automatic partial braking in case of an expected crash. CDD-B is designed for cruise control systems and realizes vehicle decelerations of up to -3.5 m/s^2 at vehicle velocities larger than 35 km/h. Input to CDD-B is the required nominal vehicle deceleration, and output is the actual vehicle deceleration which is realized by active braking of all wheels. For active braking, the pump motor of the hydraulic unit of ESC is stimulated, and the separation valves are controlled by a variable current as described at AVH-S to achieve the required vehicle deceleration. Also, here the requirements concerning noise and comfort are very high so that high-precision separation valves must be used (Breuer and Bill 2013).

5.4.7 Controlled Deceleration for DAS Stop&Go, CDD-S

Vehicles are relatively often driven in the lower velocity range (0–30 km/h), i.e., in about 32 % of the total vehicle operation time. The traffic jam assistant helps the driver to avoid collisions in traffic jam situations in which the vehicle velocity is below 30 km/h. The system requires a short distance sensor (e.g., a RADAR sensor) for low velocities in order to recognize obstacles in front of the vehicle. In addition, a high-performance brake system is required in order to decelerate the vehicle at low velocities up to a standstill in a comfortable manner. If required, the vehicle will be decelerated by the traffic jam assistant up to a standstill. Like with CDD-B, the function CDD-S is used in cruise control systems to set the vehicle deceleration. CDD-S covers the complete velocity range, including stop&go. CDD-S can achieve high deceleration values up to -6 m/s^2. The vehicle can be kept standing by hydraulic means or by the mechanical parking brake. Because of the frequent operation of CDD-S, enhanced high-performance hydraulic units of ESC are required. If the vehicle in front stops, the driver can be warned by visual, by acoustical, as well as by haptical means, e.g., by AWB, in order to stimulate him to apply the brake. If the driver does not brake in time, then the system decelerates the vehicle up to a standstill.

5.4.8 Controlled Deceleration for Parking Brake, CDP

CDP can be used in vehicles that possess an electromechanical parking brake. This brake replaces the conventional hand brake lever: the cables of the parking brake are pulled by an electromotor. With the engine running, the ESC hydraulic unit takes over the task of the parking brake until the vehicle reaches a standstill and for some time period thereafter until the parking brake is activated and holds the vehicle. CDP is the interface to the ECU of the electromechanical parking brake and brakes the vehicle by active pressure generation at the wheels. All ESC functions remain available during braking.

5.4.9 Hill Hold Control with Acceleration Sensor, HHC-S

Accelerating a vehicle from a standstill on an inclined road requires a complicated coordination of releasing the brake pedal, engaging the clutch, releasing the handbrake, and actuating the accelerator pedal such that the vehicle does not roll back while the brake pedal is released. This process can be simplified to a normal acceleration on a horizontal road using the ESC hydraulic unit. The function holds the pressure in the brake wheel cylinders induced by the driver for up to 2 s after release of the brake pedal. An active increase of the brake pressure is not implied. The driver has now sufficient time to change from the brake pedal to the accelerator pedal. The brake pressure is reduced as soon as the acceleration procedure is detected. In order to determine the right starting time for the pressure release, the equilibrium situation of the vehicle is analyzed. The analysis uses the engine torque and the downhill force on the vehicle because of the vehicle weight. The downhill force is derived from the longitudinal acceleration sensor signal. The HHC-function is automatically activated. In order to avoid the driver leaving the vehicle while HHC is active, additional signals are monitored (e.g., clutch signal).

5.5 Advanced Driver Assistance System Support

In this category of value-added functions, the ESC interventions are modified based on sensor signals of other active and passive safety systems. For instance, the automatic warning brake helps to alert the driver in critical situations (see also ▶ Chap. 47, "Development Process of Forward Collision Prevention Systems").

5.5.1 Adaptive Brake Assist, ABA

The stopping distance in an emergency situation may be shortened by early brake application. HBA helps the driver to continue fast brake application in emergency situations up to ABS control at all wheels. However, HBA starts operation only after some conditions are fulfilled, which takes some time. Moreover, because the brake pads must first be moved to the brake disks before the brake pressure can increase and the brake torque can be generated, there is some delay between brake application and vehicle deceleration. If sensors are used that scan the area around the vehicle, the emergency situation may be detected before the driver applies the brake pedal. If an emergency situation is detected this way, the conditions for activating the HBA can be reduced so that the function starts earlier. Moreover, the brake pads can be actively moved to the brake disks by the function ABP ahead of the brake application by the driver. If the driver then applies the brake, the HBA function starts earlier and the vehicle deceleration starts earlier and the stopping distance is substantially reduced. This function is also called "predictive brake assist" (PBA). If the vehicle has a brake by wire system, then the booster gain can also be increased. Even if a collision cannot be avoided, the function can reduce the severity of the accident. In an enhanced function, the signals of the sensors that scan the area around the vehicle can be used to estimate the required brake pressure in order to avoid a collision. If the driver then applies the brake, this required pressure is automatically generated immediately.

5.5.2 Automatic Brake Prefill, ABP

If an emergency situation is detected from the sensor signals that scan the area around the vehicle which can lead to a collision, then the brake pads are moved to the brake disks by a small active brake pressure. If the driver then applies the brake, the vehicle deceleration will follow without time delay. For the movement of the brake pads, the function electronic brake prefill (EBP) is used. The function is used, e.g., in the function ABA.

5.5.3 Automatic Emergency Brake, AEB

This function automatically generates full braking of the vehicle with ABS control at all wheels, even if the driver does not apply the brake on time. A prerequisite for this function is a reliable detection of the emergency situation. The function uses information from a sensor for long distance, as is used with ACC, and information from sensors that scan the near area around the vehicle (e.g., video camera). As with CDD-B, an active braking of the vehicle is initiated and continued up to vehicle standstill like with CDD-S. The pressure in the brake wheel cylinders is not increased up to a level at which a desired vehicle deceleration is reached but as fast as possible up to a level at which all wheels are ABS controlled like with HBA.

5.5.4 Automatic Warning Brake, AWB

Various possibilities are available to attract the attention of the driver to emergency situations. The driver can be alerted by optical or acoustical signals if on the basis of sensor signals that scan the area around the vehicle potential danger is detected. Effective are haptic signals which can be felt by the driver as, for instance, a jerk on the vehicle which occurs with a fast change in the vehicle acceleration. With AWB, this jerk is obtained by a small active brake pressure impulse of approx. 10 bar. The active brake pressure impulse is generated by running the pump of the hydraulic unit of ESC. The separation valves between the brake main cylinder and the brake circuits are current controlled with an electrical current that corresponds to 10 bar. The pressure of 10 bar is kept for 250 ms at the wheels. Then the pressure is reduced by opening the separation valves, and the pump motor can be stopped.

5.6 Monitoring and Information

To this category belong functions that are based on ESC and that provide the driver with important information, e.g., the tire inflation pressure.

5.6.1 Tire Pressure Monitoring System, TPM

If the tire inflation pressure is lower than the nominal value, then the tire wear increases. At high vehicle velocities, the tires with low inflation pressure become hot and may burst because of the increased roll resistance and deformation work, in particular if the vehicle is loaded and at a high environment temperature. It is recommended that the driver checks the tire inflation pressure regularly. Often, the driver forgets to check the tire inflation pressure. An investigation in the USA showed that more than half of the vehicles in traffic drive with low tire inflation pressure. TPM monitors the tire inflation pressure during driving continuously and informs the driver if the pressure is too low. After many severe accidents in the USA which were caused by loss of tire inflation pressure, an automatic tire inflation monitoring is required on all new vehicles (cars and pickups) in the USA since 2008. The function warns the driver if the loss in inflation pressure exceeds 25 %.

In the function TPM, the tire inflation pressure is not measured directly (like with the so-called direct method, TPM-C) but deduced from the wheel velocity signals (this is called the indirect method). In TPM, the four-wheel velocities are compared during straight-line driving with constant velocity. The function performs well if only one tire loses pressure. However, it is also possible to warn the driver if all four tires or if two tires at one axle lose pressure. The detection is based on the fact that if a tire loses pressure, then the roll radius decreases and the wheel rotates faster. The velocity difference is very small, in particular with low-profile tires, and must be checked for values of 0.25 %. Therefore, a very slow filtering and mean value computation of the wheel velocities is necessary. After a change of a tire, the function has to be reset, e.g., by pushing a tip switch, and all tires must be inflated with their nominal pressure. Besides the evaluation of the wheel velocities, the frequency spectrum of the wheel velocities can be analyzed to detect a loss of inflation pressure of each tire by itself (TPM-F).

6 Outlook

Shortly after the introduction of ESC on the market in 1995, the important brake assistant was introduced. Since then, the number of assistance functions has exploded. In the beginning, the integration of ESC with other active vehicle dynamics control systems like active steering, active suspension, and active propulsion distribution was the center of attention (Isermann 2006). This development has been pushed since 2008, but the combination of the active safety system ESC with systems that are based on sensors that observe the area around the vehicle and with passive safety systems has also been pushed hard. Of key importance are the reliable detection of emergency situations and the safety of the integration of active systems. Safety determines the pace of progress in these fields. Therefore, it will take some years before integration and combination of the systems have been completely realized. Furthermore, it is difficult to integrate and couple systems which are produced by competitors. The exchange of highly confidential information like specifications and safety-relevant data between competitors (e.g., information on failure rates and risk numbers), which are a prerequisite for the safety of the total system, is a big challenge.

References

Breuer B, Bill KH (eds) (2013) Bremsenhandbuch. Vieweg Verlag, Wiesbaden

Burkhardt M (1993) Radschlupf-Regelsysteme. Vogel Buchverlag, Würzburg

Fischer G, Müller R (2000) Das elektronische Bremsenmanagement des BMW X5. ATZ 102 (9):764–773

Isermann R (ed) (2006) Fahrdynamik-Regelung. Vieweg Verlag, Wiesbaden

Keppler D, Rau M, Ammon D, Kalkkuhl J, Suissa A, Walter M, Maack L, Hilf KD, Daesch C (2010) Realisierung einer Seitenwind-Assistenzfunktion fuer PKW. Tagungsband des 11. Braunschweiger Symposiums AAET 2010 (Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel), 10. And 11. February 2010, IST Niedersachsen e.V. Braunschweig 2010:178–199

Rieth P, Drumm S, Harnischfeger M (2001) Elektronisches Stabilitätsprogramm. Bibliothek der Technik, Bd. 223. Verlag Moderne Industrie, Landsberg/Lech

Robert Bosch GH (ed) (2004) Fahrsicherheitssysteme. Vieweg Verlag, Wiesbaden

Schindler E (2007) Fahrdynamik. Expert Verlag, Renningen

van Zanten A, Erhardt R, Pfaff G (1994) FDR – Die Fahrdynamikregelung von Bosch. ATZ96 (11):674–689

Vehicle Dynamics Control Systems for Motorcycles

40

Kai Schröter, Raphael Pleß, and Patrick Seiniger

Contents

Abstract

Motorcycling is a fascinating kind of transportation. While the riders' direct exposure to the environment and the unique driving dynamics are essential to this fascination, they both cause a risk potential which is several times higher than when driving a car.

K. Schröter (✉)
Formerly Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: kaischroeter@gmail.com

R. Pleß (✉)
Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: pless@fzd.tu-darmstadt.de

P. Seiniger
Referat F1, Bundesanstalt für Straßenwesen, Bergisch Gladbach, Germany
e-mail: seiniger@bast.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_42

This chapter gives a detailed introduction to the fundamentals of motorcycle dynamics and shows how its peculiarities and limitations place high demands on the layout of dynamics control systems, especially when cornering. The basic principles of dynamic stabilization and directional control are addressed along with four characteristic modes of instability (capsize, wobble, weave, and kickback). Special attention is given to the challenges of braking (brake force distribution, dynamic over-braking, kinematic instability, and brake steer torque induced righting behavior).

It is explained how these challenges are addressed by state-of-the-art brake, traction, and suspension control systems in terms of system layout and principles of function. It is illustrated how the integration of additional sensors – essentially roll angle assessment – enhances the cornering performance in all three categories, fostering a trend to higher system integration levels.

An outlook on potential future control systems shows exemplarily how the undesired righting behavior when braking in curves can be controlled, e.g., by means of a so-called brake steer torque avoidance mechanism (BSTAM), forming the basis for predictive brake assist (PBA) or even autonomous emergency braking (AEB). Finally, the very limited potential of brake and chassis control to stabilize yaw and roll motion during unbraked cornering accidents is regarded, closing with a promising glance at roll stabilization through a pair of gimbaled gyroscopes.

1 Introduction

In 2010, the probability of being killed in a motorcycle accident in Germany was more than 12 times that of being killed in any other form of road accident, for the same distance traveled (DESTATIS 2013). The combination of longitudinal, lateral, and vertical dynamics at play when leaning into a bend on a motorcycle holds a great fascination but, at the same time, places high demands upon the design of vehicle dynamics control systems.

This is why, for many years, the only systems on the market were brake and traction control systems for straight travel and therefore of limited use for bends. The first antilock braking system (ABS) for motorcycles appeared on the market in 1988 (Stoffregen 2010) and the first traction control in 1992 (Tani et al. 1993). Traction Control Systems that could detect curved trajectories by means of sensors and allow for them in control first became available in 2009 (Landerl et al. 2010) – and similar ABS in 2013 (Bosch 2013a). Since 2012 there have also been semi-active suspensions on the market (Böhringer 2013), promising further improvements by interacting with the existing systems.

Although the market penetration of vehicle dynamics control systems for motorcycles is limited compared with that for cars, there has been a huge increase in their acceptance and fitting rates over the last few years (for ABS as an example, cf., e.g., DAT 2014; ADAC 2010; Bosch 2013b). The legislator is now providing some impetus for these systems by requiring that, throughout Europe, all newly developed motorcycles over 125 cm^3 are equipped with ABS as from 2016 and all new vehicles of this capacity as from 2017 (EU 2013).

This chapter outlines the limits of vehicle dynamics control systems for motorcycles, provides an overview of the functioning of existing systems, and looks at what the future holds for vehicle dynamics control systems.

2 Dynamic Stability

The most obvious difference between motorcycles (for which the technically accurate term is single-track vehicles) and cars (two-track vehicles) is the stability of the vehicle – especially at standstill. A motorcycle is an unstable system; without stabilization it tips over and falls (for which the technically accurate term is capsize); it is stabilized by various dynamic mechanisms. But it is precisely this instability that allows a type of driving that makes motorcycling a fascinating mode of travel: Corners are taken in a leaning position. The lean angle of the vehicle is known as the roll angle λ and, in a steady-state circular course, is such that the resultant of centrifugal force and mass force of the vehicle intersects the tire contact line. There is no rolling moment around the tire contact line and the vehicle moves – like an inverted pendulum – in so-called unstable equilibrium. The force equilibrium of steady-state cornering is shown in Eq. 1 and Fig. 1.

The arising theoretical (physically active) roll angle λ_{th} is

$$\lambda_{\mathrm{th}} = \arctan\frac{F_{\mathrm{C}}}{G} = \arctan\frac{m \cdot a_{\mathrm{y}}}{m \cdot g} = \arctan\frac{a_{\mathrm{y}}}{g} = \arctan\frac{v^2}{R \cdot g} \tag{1}$$

with mass force of the vehicle G, centrifugal force F_{C}, mass m, lateral acceleration related to the lane a_{y}, driving speed v, and curve radius R. The roll angle is therefore only dependent upon lateral acceleration. With the maximum lateral friction coefficient of the tires,

$$\mu_{\mathrm{lat,\,max}} = \frac{a_{\mathrm{y}}}{g} \tag{2}$$

the maximum roll angle is obtained as

$$\lambda_{\mathrm{th}} = \arctan\mu_{\mathrm{lat}} \leq \arctan\mu_{\mathrm{lat,\,max}} = \lambda_{\mathrm{th,\,max}} \tag{3}$$

On dry, gripping road surfaces, the lateral friction coefficients of modern motorcycle tires reach values of around 1.2. This means that physical roll angles of up to 50° are drivable.

However, the theoretical roll angle and the roll angle equation (Eq. 1) are only valid for idealized tires without width. With real tires, an additional roll angle is required to maintain equilibrium, because the tire contact patch does not lie in the symmetry plane of the vehicle; see Fig. 1. The so-called tire-related additional roll

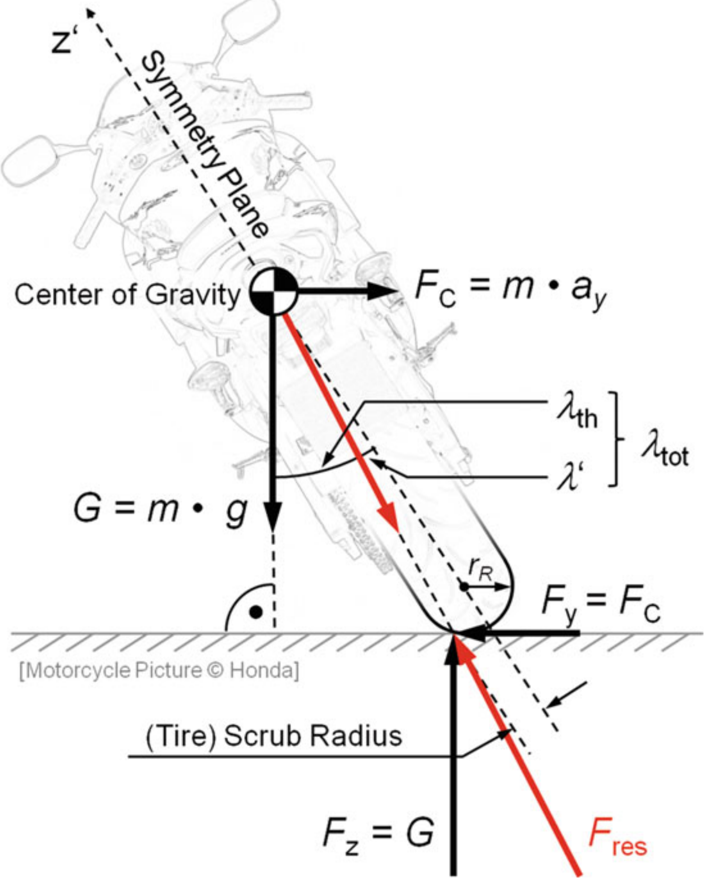


Fig. 1 Force equilibrium during cornering with additional roll angle determined by tire width

angle λ' is approximately 10 % of λ_{th}, depending upon the tire width (more) and height of the center of gravity (less).

Other additional roll angles (λ'', λ''') are one or two more orders of magnitude below this first tire-width-related additional roll angle and are negligible for the purpose of understanding the peculiarities of the driving dynamics of motorcycles in practice (Weidele 1994). The overall (total) roll angle for standard tire widths and center of gravity heights of modern motorcycles is therefore

$$\lambda_{tot} = \lambda_{th} + \lambda' \approx 1,1 \cdot \lambda_{th} \tag{4}$$

Under ideal conditions (of $\mu_{lat,max} = 1.2$), geometric roll angles of up to 55° can therefore occur for typical motorcycles. As a rule, the roll angle is limited to values of around 50° because of add-on components such as exhaust system and footrests, so that ideally a small safety margin remains.

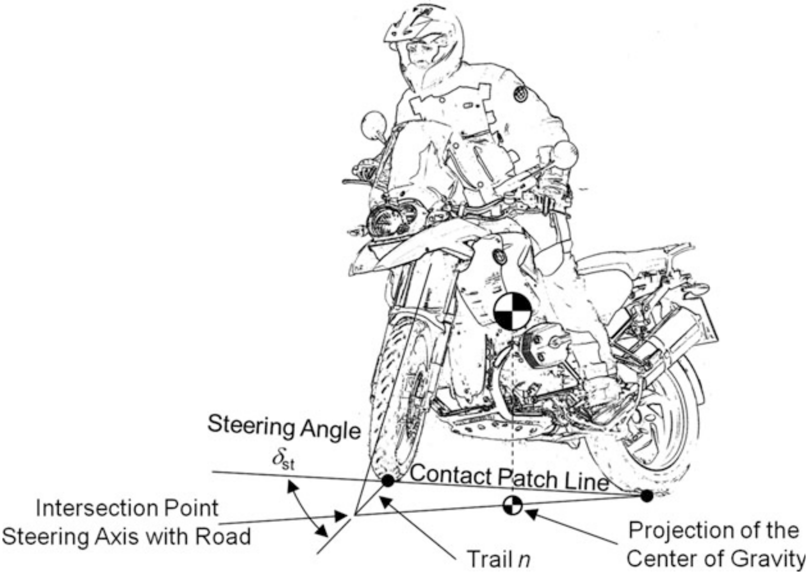


Fig. 2 Stabilization by steering adjustments and possibly shifting body weight

The previously described unstable equilibrium ceases to return to the position of equilibrium with the smallest deflections; this attribute is called instability. Motorcycles are stabilized by two mechanisms:

- At low speeds below approx. 30 km/h, for example, the rider mostly stabilizes the motorcycle by making steering adjustments, which, like the balancing of bicycles, can be supported by the rider shifting his/her weight.
- At higher speeds above approx. 30 km/h, the gyroscopic effect of rotating masses stabilizes the motorcycle. In particular, the rotating front wheel makes a significant contribution and is of utmost importance for gyroscopic stabilization.
- There is a fluid transition between these two mechanisms.

Figure 2 shows the system of motorcycle and rider with tire contact patches and projection of the center of gravity. It illustrates how, by moving the handlebars, it is possible to control the horizontal distance between center of gravity and tire contact line – at first approximation this is the roll axis. In addition to the steering adjustments, the rider can also move his/her weight relative to the vehicle to control the lever arm between the center of gravity and roll axis, thereby stabilizing the rolling motion. However, it is the steering adjustments that are the most important, because even cabin motorcycles, where there is limited opportunity to shift body weight, can be driven stably at low speeds.

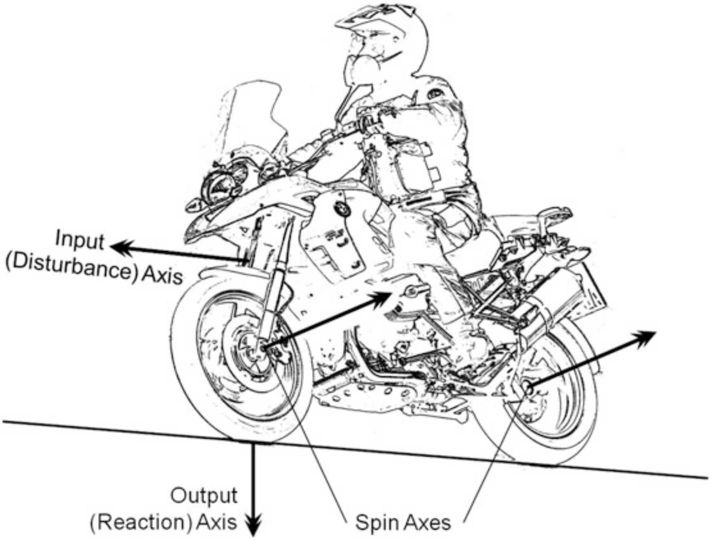


Fig. 3 Stabilization due to gyroscopic effect at the front wheel

From speeds of approximately 30 km/h upwards, the angular moments of the wheels (and other rotating parts of the motorcycle) reach such high values that the tipping motion or capsize of the vehicle is stabilized by their gyroscopic effect. The stabilization mechanism is illustrated in Fig. 3.

A gyroscope that is disturbed perpendicular to its rotational axis responds with a reaction moment perpendicular to the rotational and disturbance axis. This mechanism couples the equation of motion of the motorcycle around the longitudinal axis with the equation of motion of the steering system. A roll motion of the vehicle (e.g. to the right) causes the steering system to skew in the same direction. The lateral force on the front wheel generated by the steering angle produces equivalent centrifugal force at the center of gravity, which re-rights the vehicle (in this example, towards the left). In the same way as the lateral force with the trail as lever arm, the generated gyroscopic reaction moment exerts a resetting steering torque (cf. Fig. 2).

Since the coupling gyroscopic moments are a function of the respective disturbance speed in the steering or rolling direction, they also serve to damp the stabilization process. In an infinite sequence of the previously described effect chain, the stabilization process can be visualized as driving in wavy lines, which become smaller and smaller with increasing speed. From approximately 30 km/h, the tipping motion of the capsize mode is largely damped and the vehicle travels without any apparent deflections in steering and roll angle.

With increasing speed, the gyroscopic effect of the wheels increases once more; from speeds of approximately 130 km/h, the system can again become unstable, depending upon its stability characteristics. The so-called (high-speed) weave eigenmode of the motorcycle is a coupled yawing, rolling, and steering oscillation of the entire vehicle, which increases with speed and in extreme cases can lead to a fall by exceeding the traction limits at the front and/or rear wheel (Bayer 1986). Depending upon the vehicle, weave frequencies are between 2 and 4 Hz. The most effective remedy for incipient weave is to reduce speed. The main influences on the excitation of weave oscillations are the torsional stiffness between the front and rear wheel and the inertial characteristics of the vehicle. Minimizing weave is an important aspect in the development of modern motorcycles. Therefore, it typically only occurs in exceptional cases.

Another technically relevant eigenmode – which is similarly minimized during the development of a new vehicle – is the so-called wobble , a rotational oscillation of the steering system that is also known as flutter or tank-slappers. The usual frequencies of wobble oscillations are in the region of 10 Hz. This frequency corresponds to the rotation frequency of standard front wheels at approximately 60–80 km/h, the wobble being excited by wheel imbalances and irregularities. Usually this can be countered by gripping the handlebars more firmly, in order to increase the moment of inertia around the steering axis by coupling the rider's body and so push the oscillating system down to a lower eigenfrequency.

Another form of oscillation of the steering system is the so-called kickback, which is not a natural mode but a parametrically excited oscillation with many different influencing variables. A prerequisite for the occurrence of kickback is wheel load fluctuation – for example, due to a bump in the ground – at the front wheel with prevailing steering torque. If the wheel load drops rapidly, the prevailing steering torque turns the steering system in and the sideslip angle of the front wheel increases. If the wheel load subsequently increases, there is excessive sideslip and therefore excessive side force on the front wheel, which turns the handlebars back into direction of the neutral position. With corresponding excitation, these handlebar movements can even cover the entire range between the two end stops. The usual remedy against kickback is to use hydraulic steering dampers. In order to reconcile the conflicting aims of easy handling with light steering at low speeds and controlling kickback, semi-active steering dampers with electronically adjustable damping – e.g., the Honda Electronic Steering Damper (HESD) – have been successfully used in series production vehicles since 2004 (Wakabayashi and Sakai 2004). The use of this technology to influence the eigenmodes of weave and wobble is the subject of ongoing research (De Filippi et al. 2011a).

3 Braking Stability

The additional roll angle introduced in the previous section has a particularly marked effect when braking in curves: The steering axis of a motorbike is normally in the symmetry plane. Braking forces, which act in the tire contact patch, therefore maintain a lever arm to the steering axis when cornering; see Fig. 1. Via this so-called (tire) scrub radius (SR), the braking forces induce a misaligning steering torque, the so-called brake steer torque (BST). It is the riders' task to compensate for this disturbance and stay on course. If they do not succeed, the steering system turns into the curve, sideslip angle at the front wheel and lateral acceleration increase, and

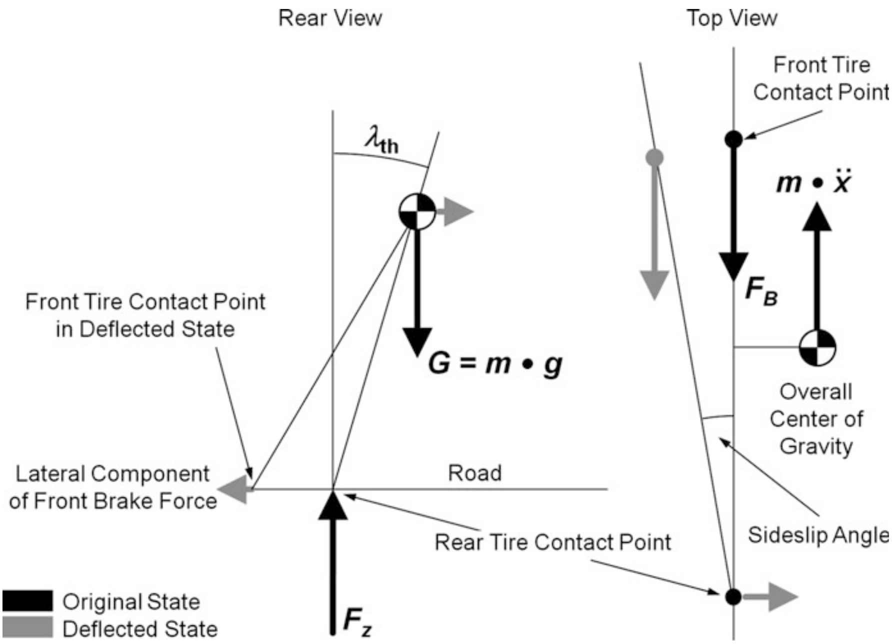


Fig. 4 Kinematic instability of yaw and roll motion (explanation as per Funke (2007) and Seiniger (2009))

the vehicle straightens up in interaction with the gyroscopic forces of the inward-turning front wheel and – often unexpectedly for the rider – forces the bike into a wider radius (Schröter et al. 2013). In extreme cases, the brake steer torque reaches values of approx. 90 Nm, which follow the buildup in braking pressure with virtually no delay. If the braking force also pulses, for example, due to a "roughly" controlling ABS on the front wheel, it becomes practically impossible for the rider to stay on course. Although the righting behavior due to brake steer torque is actually a whole chain of effects, it is frequently referred to simply as "righting moment."

The behavior of motorcycles in the event of wheel locking also differs significantly from that of two-track vehicles. For the latter, locking of both front wheels does not compromise the directional stability of the vehicle – in contrast to locking of the rear wheels. In the case of motorcycles, on the other hand, front wheel locking almost inevitably leads to a fall: The reasons for this are the loss of gyroscopic stabilization and, even more importantly, the kinematic instability of the vehicle. For a two-track vehicle, front wheel locking is stable up to a certain sideslip angle limit – for standard cars this is approximately 45°. In motorcycles, even small deviations in sideslip angle or roll angle are sufficient to cause self-amplification of yawing and rolling motion; see Fig. 4. A locked front wheel (slip $s = 1$) now only transfers a force against its direction of motion determined by the value of μ_{slip} and wheel load, but no longer any cornering force. If this force has a lever arm around the center of gravity, this causes sideslip or yaw rotation; if the rotation increases the lever arm, this is an unstable movement.

Since the motorcycle is an unstable vehicle and is constantly stabilized by the gyroscopic effects or steering movements, there is constant transverse force acting on the tire contact patches. A braking force acting on the front wheel against the direction of motion (as in the case of a locked front wheel) always causes a self-amplifying yawing motion – the tire contact line turns away under the center of gravity. Measured times between locking of the front wheel and crashing are between approximately 0.2–0.7 s; if the vehicle is already cornering, these times are much shorter (Funke 2007). The ideal distribution of braking force on the front and rear wheels is very different for motorcycles than it is for cars: The ratio between the height of the center of gravity and the wheelbase is much greater in motorcycles than in cars. Therefore, the wheel load transfer during deceleration is also greater. In conjunction with today's high-grip tires, modern motorcycles can reach the brake flip-over point. The maximum deceleration is often limited by the position of the center of gravity and the wheelbase, that is to say the geometric data of the vehicle, and no longer by the braking system or tires.

Figure 5 shows the ideal brake force distribution of a typical car (Opel Astra H) and a typical supersport motorcycle (Honda CBR 600 RR), disregarding changes in chassis geometry that are related to brake pitch and dive. The ideal brake force distribution of the motorcycle intersects the abscissa at a braking rate of 1.0 (that is to say at a deceleration of 9.81 m/s^2 corresponding to gravitational acceleration). Greater decelerations would only be possible with the rear wheel lifting off ground, so that the vehicle would no longer be stable. The illustrated braking force distributions are only valid for traveling without lateral acceleration. During corner braking, dynamically variable cornering forces also have to be supported at the tire contact patches. This reduces the transferable braking forces and consequently changes the ideal brake force distribution (BFD) (cf. Weidele 1994; Schröter et al. 2012, 2013). Moreover, the curves shown in Fig. 5 are only valid for steady-state deceleration. The brake pitch or dive process delays wheel load transfer significantly – in contrast, there is practically no time lag in the transmission of brake force to the front wheel. Especially in the case of motorbikes with negative kinematic brake dive compensation (e.g., in the case of telescopic fork bikes) with consequently greater dive movement, there is a danger of front wheel locking, even at low brake pressures that the rider does not perceive as critical – with the consequence that a fall is practically inevitable. This phenomenon is known as dynamic over-braking of the front wheel (Weidele 1994).

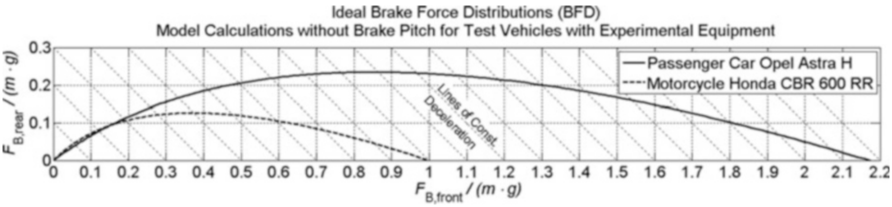


Fig. 5 Ideal brake force distribution for Opel Astra H car and Honda CBR 600 RR motorcycle (2010 C-ABS model), both fitted with experimental equipment and therefore retarded rear axle/wheel lift-off points, calculated on the basis of own measurements of centers of gravity

4 Motorcycle Accident Scenarios Relevant for Vehicle Dynamics Control Systems

Whereas the number of motorcyclists killed each year in Germany had remained more or less constant for 15 years at about 800–1,000, it fell significantly for the first time in 2008 to 656. With only a seasonal exception in 2011, when, due to a long spell of fine weather with an increase in motorcycle traffic, the accident figures unfortunately rose to 708 fatalities, this downward trend continued in subsequent years and in 2012 reached its lowest level of 586 fatalities. Despite this positive trend, over the long term the number of motorcyclists killed is falling much more slowly than the total number of road accident fatalities. Due to the huge increase in motorbike use in Australia and the USA over the last 10–15 years, the number of people killed is actually increasing in these countries (IRTAD 2013). The problem of front wheel locking in motorcycles described in the previous section – in combination with the danger of dynamic over-braking – suggests that a high proportion of accidents are braking related. Although the data available from Germany's Federal Office of Statistics are not sufficiently detailed, this assumption is supported by numerous detailed studies over long periods of time (cf. Hurt et al. 1981; ACEM 2009). For example, German insurance company databases also include detailed descriptions of a large number of motorcycle accidents which, in terms of various criteria, are representative of the accident scenarios in Germany. The German Insurance Association (GDV) database includes a study (Sporner and Kramlich 2000) of 610 motorcycle/car collisions: There was evidence of braking in 239 of these accidents and in 45 cases the motorcycle fell over prior to the collision. In approximately 7 % of the cases evaluated, wheel locking contributed significantly to the course of the accident; even in the evaluation of single-vehicle accidents, falling off the bike was the primary cause of approximately 40 % of them. Overall, at least 20 % of motorcycle accidents can apparently be influenced by having ABS. In the database analysis conducted by Allianz insurance group (Reissing et al. 2006), it was similarly found that between 8 % and 17 % of the accidents investigated could have been prevented by ABS. In a total of 87 motorcycle accidents, DEKRA accident research (DEKRA 2010) determined that between 25 % and 35 % of accidents could have been prevented by ABS, and Bosch (Yildirim and Mörbe 2013) states that 26 % of accidents involving injuries or fatalities could have been prevented. This means that, translated to the current accident figures, between 46 and 205 fatal accidents could be prevented in Germany alone by the universal use of ABS on motorcycles. The latest studies from the USA (IIHS and HLDI 2013) confirm this prediction and show that, in a comparison of otherwise equivalent motorcycles, vehicles with ABS are generally involved 20 % less frequently in collisions – and 31 % less frequently in fatal accidents – than vehicles without ABS; motorcycles with a combined ABS (ABS and CBS, Sect. 5) were involved in collisions as much as 31 % less frequently. According to Bosch (Schneider 2013), 67 % of all cornering accidents, which constitute approximately 16 % of all motorcycle accidents, are potentially preventable with the curve-adaptive combination of an ABS with a traction control system, which is referred to as Motorcycle Stability Control (MSC, Sect. 6). Since riders braked too little or too late in many cases, the use of braking aids to increase the pressure more quickly, or even predictive systems such as “predictive brake assist” (PBA), promise further improvements. DEKRA tentatively assumes that between 50 % and 60 % of all relevant accidents could have been prevented (DEKRA 2010) and also that the severity of the unavoidable accidents could have been reduced by drastically reducing collision energy (Roll et al. 2009; Roll and Hoffmann 2010).

However, because of the uncertain nature of the data available, it is difficult to accurately determine the potential for future vehicle dynamics control systems that go beyond this. A related study (Gwehenberger et al. 2006) found that unbraked accident in curves were potentially preventable, which account for approximately 8 % of accidents. These are addressed in more detail in Sect. 8.2.

5 State-of-the-Art Brake Control Systems

Figure 6 gives an overview of the operating principle of hydraulic braking systems. Hydraulic motorcycle braking systems are based on a dual-circuit standard braking system with separate controls for front and rear brake. Operation of the handbrake generates a hydraulic pressure that is transmitted to the front wheel brake via hydraulic lines where it is converted into a clamping force. The same process is repeated when the rear wheel brake is operated via the footbrake or a second handbrake. Nowadays, disk brakes are primarily used as wheel brakes: Such braking systems are technically mature and universally used; however, without additional measures, they are not sufficient to meet the requirements of a modern braking system for motorcycles when it comes to avoiding wheel lock. In order to achieve a short stopping distance, it is up to the rider to modulate the pressure in the braking system, i.e., build up the braking pressure on the front wheel as quickly as possible, in keeping with the ideal brake force distribution – without causing the wheel to lock – and similarly build up the pressure as quickly as possible on the rear wheel, but then reduce it again because of the dynamic wheel load transfer during braking. This is the only way to guarantee a short stopping distance while

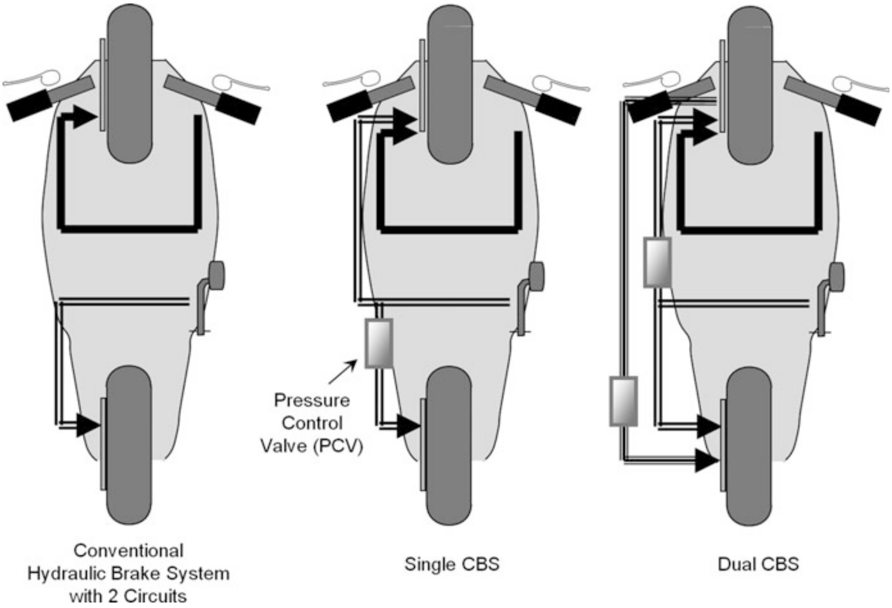


Fig. 6 Operating principles of hydraulic motorcycle braking systems

maintaining the stability of the motorcycle. However, such a control task often proves too much for a motorcyclist, especially in emergency situations. This means that the vehicle is not slowed down optimally – either braking pressure is generally too low, built up too late, or with insufficient gradient – or the wheels are overbraked or even locked. This compromises the stability of the vehicle, and if a wheel locks up (especially the front wheel), this almost inevitably leads to a fall. In order to better replicate an ideal brake force distribution on the front and rear wheel, motorcycles with so-called combined brake systems (CBS) are commercially available.

These are available in two versions (Fig. 6):

- Single CBS, whereby manual control acts on the front wheel, foot control (or the second manual control) acts on the front and rear wheel; this means that relatively high deceleration can be achieved by using only one control element.
- Dual CBS, whereby both wheels can be slowed down using either brake lever (the hand- or the footbrake).

Such systems have relatively complex hydraulics: Dual CBS uses a floating front brake caliper with an additional actuating cylinder, a so-called secondary master cylinder. By means of an additional hydraulic connection, this serves to build up pressure in the hydraulically partitioned rear wheel caliper. In both systems the front wheel caliper is hydraulically partitioned – e.g., five pistons connected to the manual control and one piston connected to the foot control – which pushes up the cost of the overall system even further. By supplementing this brake system with so-called pressure control and/or delay valves (PCV, DV), it is possible to adapt the brake pressure and force buildup as well as limitation on both wheels more exactly to the desired brake force distribution. Despite the general nature of the term PCV for all kinds of valve systems used to control the brake pressure, it is typically used to describe pressure cut-off valves in the rear brake circuit, while the term DV describes a special type of PCV used to delay pressure buildup at the front upon activation of the rear brake.

5.1 Hydraulic Antilock Braking Systems (ABS)

However, preventing locking up of the wheels and therefore maintaining stability can only be guaranteed by a system which modulates brake pressure by sensing traction so that if a braked wheel is threatening to lock, it can be speeded up again to maintain the lateral force. Figure 7 provides an overview of the operating principles of hydraulic ABS: This type of antilock braking system (ABS) has been available for cars since 1978. The first motorcycle ABS was introduced in 1988 on the BMW K100 and, following initial skepticism, has found increasing acceptance amongst motorcyclists, which is reflected in rising installation rates. In a dual-circuit braking system, the ABS is switched between actuation and wheel brake; it has wheel speed sensors to monitor the wheel speed. If the rotation speed of a wheel drops excessively during braking, this is sensed and the brake pressure is reduced via the brake pressure control. Once the wheel has regained the reference speed of the vehicle, the brake pressure is increased again to slow the vehicle down further. Nowadays, dual-channel systems using valves are very common; they are lighter and less expensive than antilock integral braking systems (cf. Sect. 5.2). The same principle applies to a single CBS-ABS, exept that an additional modulator circuit is needed to couple rear wheel actuation to the front wheel.

Such systems therefore require a total of three control channels that are controlled independently of each other. Dual CBS-ABS features the previously mentioned Dual CBS supplemented by ABS modulators. The system uses a total of four control channels, each being required to control brake pressure from manual actuation to the front wheel, foot actuation to the front and rear wheel, and from the secondary cylinder of the front wheel to the rear wheel, respectively. In the outlined antilock systems, pump/valve configurations, and occasionally also plunger systems, are used to regulate brake pressure.

Many vehicles on the emerging Asiatic market only have a hydraulic disk brake on the front wheel, so that inexpensive single-channel ABS are now available.

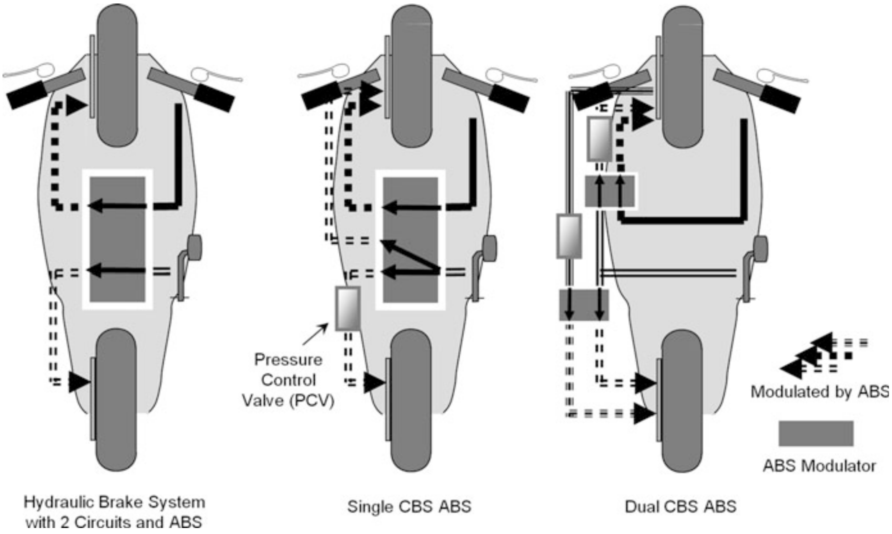


Fig. 7 Operating principles of hydraulic motorcycle ABS braking systems

5.2 Electrohydraulic Integral Braking Systems

Pure ABS are passive, as they cannot generate a higher brake pressure than that applied by the rider. However, there are units from the automobile sector that are able to supplement ABS functionality by actively, i.e., autonomously, generating pressure on individual wheels. Based on this technology, electrohydraulic integral braking systems have been developed for the motorcycle sector; Figure 8 provides an overview of their operating principles.

Just like a CBS, when one brake circuit is activated, these systems can actively generate braking pressure in the other braking circuit without additional hydraulic connections or special measures in the caliper. Partially integral systems are restricted to acting on one brake circuit, while fully integral systems can actively operate on both brake circuits.

5.2.1 Integral Braking Systems Without Power Assistance

The state of the art is to use valve technology known from the automobile sector, which has been miniaturized over the last few years for use in motorcycles. A specialized version is the partially integral braking system, whereby the braking pressure is actively built up only at the rear; i.e., such a system creates an integrated function from the handbrake to the rear wheel. Below, the partially integral braking system from Continental serves as a functional example: The system consists of a total of six hydraulic valves (two for the front wheel circuit, four for the rear wheel circuit), three pressure sensors, a low pressure accumulator, and a hydraulic pump for each wheel circuit and an ECU (electronic control unit). The two pumps of each wheel circuit are jointly driven by an electric motor. A system overview is provided

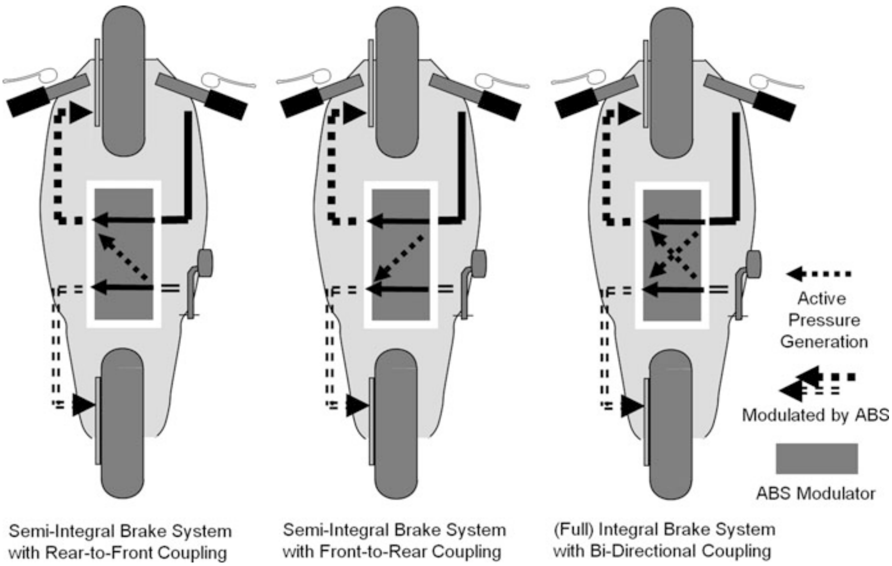


Fig. 8 Operating principles of electronic integral braking systems

in Fig. 9. If the rider operates the handbrake, the pressure is hydraulically transmitted to the front wheel brake; at the same time the pressure sensor measures the pressure rise and transmits the information to the ECU. The pump motor (M) is controlled in accordance with prescribed characteristics, operating conditions or other variables. To actively build up the pressure on the rear wheel, the isolating valve (IV-RW) is closed and the electrical switch valve (SV-RW) is opened.

The pump can then transfer brake fluid from the reservoir into the rear brake caliper and build up pressure. If the rider also operates the footbrake at the same time, the SV-RW is closed again when wheel braking pressure is reached and the IV-RW reopened, so that the rider once again has direct access to the rear wheel brake from the footbrake. In terms of valves, the front wheel circuit is designed as a simple ABS circuit.

5.2.2 Integral Brake Systems with Power Assistance

In order to reduce the brake operating forces necessary to reach high decelerations in the ABS control range even on heavier motorcycles, in 2000 BMW Motorrad produced the first braking system with “Integral ABS” (CORA BB) manufactured by FTE, which not only had the integral function but also brake power assistance (Stoffregen 2010). While, for a time, Piaggio and Peugeot used modified systems without locking protection at the rear wheel (CORA) in scooters, further development of valve-based ABS technology soon made power assistance obsolete. For example, at BMW Motorrad the valve-based technology of second-generation

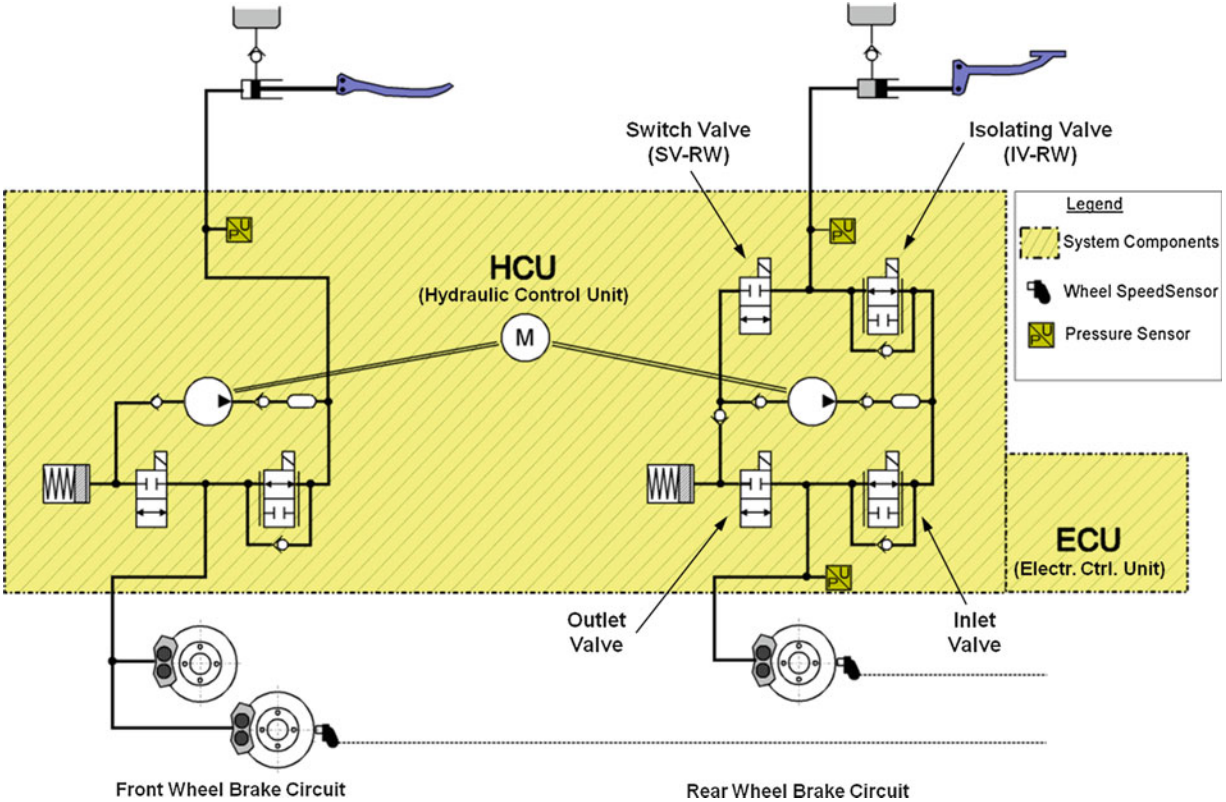


Fig. 9 Motorcycle integral braking system MIB with partially integral function

integral ABS (Sect. 5.2.1 and Stoffregen (2010)) succeeded the first as early as 2006, and in the last 2009 model, this was only available in the K1200LT.

Depending upon the stage of development of the FTE system, the hydraulics of the control elements are largely separated from the wheel brakes, and in an intact system, operation takes place in a simulator or control room. A hydraulic pump is activated by each actuation – even partial braking – so that the pressure can be built up in the wheel brake cylinder, at least in accordance with an amplification factor prescribed by hydraulic ratios. As a fallback in the event of system failures, the master cylinders of hand- and footbrake have a direct hydromechanical connection to the wheel brake cylinders. The absence of the power assist function consequently requires much higher operating forces, similar to those of a conventional brake. The ABS function works on the plunger principle, whereby a control piston is continuously displaced against the operating pressure in the control room by an electromagnet to modulate the wheel brake pressure. The integral function is produced via an additional hydraulic input, coming from the actuating element of the other brake circuit. This pressure acts on the control piston via a dividing piston, and the minimum integral brake pressure is adjusted at the respective wheel brake by the geometric relationships, as in a normal braking operation. While the electromagnet can be used to counteract the actuation at the control piston – i.e., in order to diminish the resulting brake pressure in accordance with dynamic wheel load transfer and in approximation of an ideal brake force distribution – the pump is used to generate additional brake pressure, all electronically monitored by pressure sensors.

5.2.3 Honda Combined ABS: “Brake by Wire”

Honda is following its own route with the Combined ABS (C-ABS) presented for the supersport sector in 2008. The center of gravity of supersport motorcycles is quite high relative to their short wheelbase. Hard braking maneuvers give rise to correspondingly large wheel load transfers and consequently considerable diving movements with a rapid tendency towards braking-related pitching, which has a destabilizing effect upon vehicle dynamics, especially when reducing speed coming into curves. Experimental investigations show, first of all, that the disruptive suspension reaction can be predicted from the brake pressure gradient set by the rider and the wheel speed information and, secondly, can be minimized by briefly increasing the front wheel slip and by early triggering of ABS control (see Nishikawa et al. (2008), cf. also Sect. 5.3). A “brake-by-wire” architecture was chosen to implement this strategy with rapid brake pressure buildup, independently of the rider.

The system is divided into five components: Apart from the electronic control unit (ECU), an equivalent valve unit and power unit are integrated into the hydraulic line of the handbrake and footbrake, respectively. Although this means that they can be mounted on various types of vehicles in a way that is favorable to the center of gravity, it comes at the price of a comparatively high system mass of around 10 kg.

For braking maneuvers when switched off, the system works like a conventional dual-circuit brake system, which also serves as a fallback in the event of failure. In the active system, when a low brake pressure threshold is exceeded, the hydraulic connection of the brake lever to the wheel brakes is isolated by switchover of valves and diverted to force/displacement simulators that imitate the feel of a conventional brake at the lever. The rider's desired deceleration is detected by brake pressure sensors and processed in the ECU. Via spur gears and ball screws, electric motors in the power units drive separate master cylinders, which build up the pressur at the wheel brakes. ABS control is based on conventional wheel speed sensors but is continuous, without the otherwise characteristic pulsing.

The system allows to produce arbitrary brake force distributions and possibly even an amplification function with many degrees of freedom: For example, the rear wheel is always braked in advance, and when the brake is released, a different brake force distribution is applied than during actuation (Nishikawa et al. 2008), which is not immediately possible with a conventional hydraulic CBS.

Although the system does not have a roll angle sensor, the control of large pitching motions and the smooth ABS control already allow an astoundingly good performance during corner braking, even at larger roll angles (cf. Sects. 8.1 and 5.3.1). This is impressively confirmed by its successful use in racing events (Tani et al. 2010).

5.3 Additional Functions

The so-called rear wheel lift-off protection (RLP, often also referred to as "rear wheel lift-off/lift-up mitigation") effectively reduces the risk of brake flip-over and is already used in many simple dual-circuit ABS. RLP compares the wheel speed signals and signals derived from both wheels during braking. In addition, pressure information from the individual control circuits – and, in the latest systems, even the pitch rate and longitudinal acceleration (Bosch 2013a) – can be processed to give a lift-off tendency and limit deceleration as a function of the driving situation. There is no direct sensing of the distance between the wheel and the road surface. The pressure control algorithm of the front wheel reduces braking pressure – even below the ABS control threshold – in such a way as to guarantee a minimum contact force of the rear wheel as surely as possible.

The "active brake pressure distribution" (ABD, also called eCBS – electronic CBS) is responsible for distributing the rider's desired braking to both wheels. This happens in interaction with the brake pressure directly induced by the rider via the two control elements, whereby the individual allocations – from the handbrake to the rear wheel and from the footbrake to the front wheel – are being converted by the software. The basic characteristic can be based on the ideal brake force distribution and then altered in keeping with the riding situation: It uses input variables such as vehicle speed and also the signals describing the rider's braking profile. In this way, for example, the integral action of the rear wheel brake on the front wheel brake can be reduced at very low speeds to prevent deflection of the steering during a turning maneuver. However, an ABD also requires an active braking system (e.g., Bosch ABS 9 ME, Continental MIB, FTE CORA BB, or Honda C-ABS). Consequently, with these, it is also possible to have a function such as motorcycle Hold & Go (MHG), to actively assist the rider with hill starts.

Apart from providing special operating modes for racing and off-road use, there is a trend towards extending the function by adding additional sensor information about driving conditions (Sect. 5.3.1) and interaction with other control systems such as traction control (Sect. 6) or suspension regulation (Sect. 7).

5.3.1 Curve Adaptive Braking System

The central requirement for brake control in curves is to safeguard vehicle stability, which is particularly delicate in this situation (Sect. 3). At the same time as achieving high decelerations, it is always necessary to ensure that there is an adequate lateral force reserve available. Sudden friction variations from high to low and low friction coefficients in general place physical limits on this (cf. Seiniger (2009) and Sect. 8.2). Also, because of the coupling of steering and rolling dynamics, it is important to control brake steer torque (Sects. 3 and 8.1).

In particular by taking account of the roll angle as a characteristic parameter, the braking strategy can be adapted to the specific requirements of corner braking. Although it is not possible to expand the vehicle dynamic potential imposed by the abovementioned limits, it can be made usable to a greater extent by making it easier to manage for the rider.

The resulting safety gain is illustrated below using the example of corner braking with conventional integral ABS, followed by a presentation of various control strategies and the first curve-adaptive brake system that was introduced in 2013 as part of Motorcycle Stability Control (MSC) (Bosch 2013a) in Bosch and KTM standard production.

Figure 10 shows the course over time and the consequences of impending wheel lock during corner braking: At $t = 0$ s, a roll angle of approximately 20° and a driving speed of 65 km/h, the front wheel displays a clear drop in rotation speed. Because excessive demands are placed on traction, its slip angle increases, while the yaw rate of the vehicle and the curvature of its course decrease. At the start of incipient wheel locking and the subsequent ABS control action, braking force drops. The outward steering torque applied by the rider to balance the formerly higher level of brake steer torque turns the handlebars further outwards. At the end of the control operation, the maximum braking force is reestablished at the front wheel and therefore also a strong inward brake steer torque, which – with the steering effort obviously taken back by the rider – again turns the handlebars significantly inwards. Temporarily, the handlebars start to oscillate, and when sufficient amplitudes are reached, a rapid drop in roll angle and associated high roll rates are observed. This is followed by obvious yawing and rolling oscillations of the vehicle during the rest of the braking process. Under certain circumstances in real road traffic, this can result in the vehicle departing from its designated lane. The steering and rolling oscillations are caused by the effect of brake steer torque

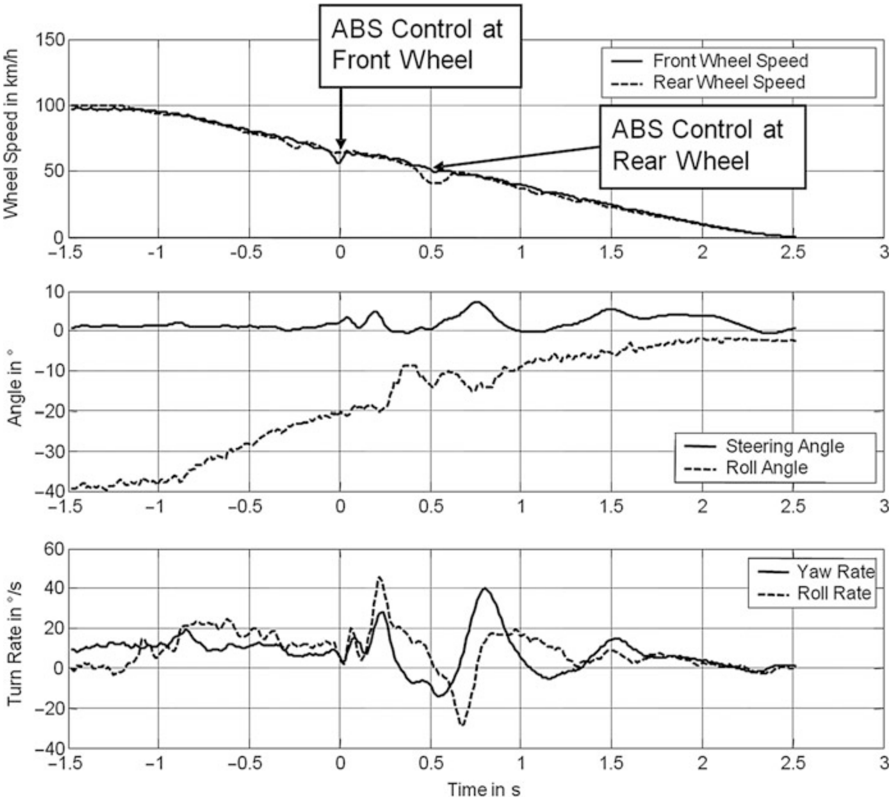


Fig. 10 Sequence of ABS control during corner braking with conventional integral ABS (BMW R1150RT, cf. Seiniger et al. 2006)

(Sect. 3) in combination with the kinematic instability (Sect. 3) of the vehicle and control of its course by the rider.

Apart from the steering, rolling, and course disruptions caused by brake steer torque, the example also shows the tendency towards destabilization by placing excessive demands on the available traction. While research is currently being carried out to develop a chassis with dynamically adjustable steering axis to combat brake steer torque (Sect. 8.1), the measures described in the following section are also known for improving corner braking with a conventional chassis.

In order to avoid dynamic front wheel over-braking, which is particularly critical when cornering, to control brake pitch and, above all, to give the rider a bit more time at the start of braking to compensate for brake steer torque, it lies at hand to limit the brake force buildup gradients and possibly also the maximum braking pressure, as a function of the roll angle. The kinematic instability can furthermore be taken into account by using roll-angle-dependent slip thresholds (Roll and Hoffmann 2010): These facilitate more sensitive ABS control, including a compensation of so-called pseudoslip, which results from the typically different widths and contours of the front and rear tires. A brake force distribution that is increasingly directed towards the rear wheel with increasing roll angle (Weidele 1994) not only increases the lateral force reserves at the stability critical front wheel by reducing its braking force, but simultaneously reduces brake steer torque. If, in addition, the rear wheel is over-braked early, it is possible to estimate the prevailing traction level. Nevertheless, due to premature ABS intervention at the back end, this can also trigger yawing, steering, and course interferences. However, these are typically not critical and are offset by the advantage that limiting the maximum brake pressure at the front means that ABS-related steering torque disruptions and their consequences can be avoided. Finally, the brake system should prevent the rear wheel from lifting off or even a brake flip-over, even during cornering.

Following the introduction of a roll angle sensor for traction control systems into standard production in 2009 (cf. Sect. 6), using this information for curve-adaptive braking control was only a matter of time. The braking system presented by Bosch in 2013, in conjunction with KTM, in connection with Motorcycle Stability Control (MSC, Bosch 2013a), uses a sensor cluster, which measures two rotation rates and accelerations in all three spatial directions. Because its installation position is turned through 45° around the transverse axis, one rotation rate sensor measures the pitch rate, while the other measures a combination of roll and yaw rate. Information can therefore be mathematically obtained about all six degrees of freedom of movement of the vehicle, especially the roll and pitch motion, taking into account the brake force distribution (eCBS, at KTM currently only with active pressure buildup at the rear), ABS control, as well as rear wheel lift-off detection and mitigation (Bosch 2013a; Yildirim and Mörbe 2013; Willig and Lemejda 2012).

Because of the current lack of published information about the practical implementation of the MSC strategy, the last sentence is deliberately tentative. While the abovementioned roll-angle-dependent braking strategies in MSC can principally be refined in several ways, because of the additional sensor information available, this is not always necessary to represent a particular function. For practical application, it is necessary to decide whether, for example, the improvement in rear wheel lift-off mitigation by taking account of pitch rate and longitudinal acceleration (Bosch 2013a) justifies the additional efforts to assure functional safety under all circumstances, as opposed to the conventional approach.

As the term Motorcycle Stability Control already suggests, the overall system goes beyond the function of corner braking control alone. By including engine control, it also allows curve-sensitive traction control (motorcycle traction control, MTC), optionally with additional functions such as launch or wheelie control (Sect. 6). Special off-road mappings not only work with adapted slip thresholds but also without taking account of the inertially measured roll angle, since this suggests reduced traction potential when riding through banked curves, whereas traction potential is in fact increased by the centrifugal force. Also, in the sense of a scalable system architecture, networking with additional control systems, such as a semi-active suspension (Sect. 7), is already prepared (Yildirim and Mörbe 2013).

The accident statistics relating to road use show that the gain in stability during braking achieved with measures such as MSC is more important than the loss of maximum deceleration theoretically associated with it. Initial practical trials show that the righting motion during corner braking with MSC is well suited to the deceleration and, thanks to the improved stability during emergency braking when cornering at large roll angles, a rider can most probably even achieve higher average decelerations than with a conventional braking system (Schneider 2013, 2014).

In multitrack tilting vehicles such as the Piaggio MP3 scooter, brake controls based on electronic stability control (ESC) from the automobile sector are also feasible, ideally also taking influence via engine control (Roll and Hoffmann 2010).

6 State-of-the-Art Traction Control Systems

In view of the high performance of modern motorcycles, a traction control system, is an expedient addition to the brake control systems that have now become established (Reissing et al. 2006). The primary aim is to prevent excessive rear wheel spin, in order to assist the driver during acceleration – especially on roads with varying or reduced friction coefficients – and at the same time to maintain vehicle stability. Especially when cornering, it is important to prevent uncontrolled sideslip with the risk of a highsider accident (see also Sect. 8.2).

Honda pioneered the series production of traction control systems with their TCS in 1992. BMW delivered further milestones with its Automatic Stability Control (ASC) in 2006 and Dynamic Traction Control (DTC) in 2009. Other manufacturers have since caught up. Because the operating principle of the various systems is essentially the same, this is explained below, initially using the example of ASC and DTC and supplemented as necessary.

Figure 11 shows an overview of the DTC system components networked via CAN bus. In addition to the throttle valve actuators of "eGas" (or "ride by wire"), DTC is enhanced relative to the ASC system to include a sensor box. By measuring the roll and yaw rate, as well as the lateral and vertical acceleration, this allows the driving conditions, which are largely characterized by the roll angle, to be captured by sensors and included in the control.

In both systems, the traction control algorithm runs on the engine control unit and, in DTC, so does evaluation of the sensor box signals. In order to determine the timing and intensity of a control intervention, the control unit receives the signals from the ABS wheel sensors. The prevailing drive slip is determined from the speed difference between the front and rear wheel – and also, in the case of DTC, a correction based on roll angle.

Specific vehicle parameters – including the characteristics of the wheel-tire pairings approved for the respective vehicle – are stored in the control unit for this purpose. Since tires from different manufacturers differ slightly in terms of rolling radius and contour, manufacturing tolerances and increasing wear with use,

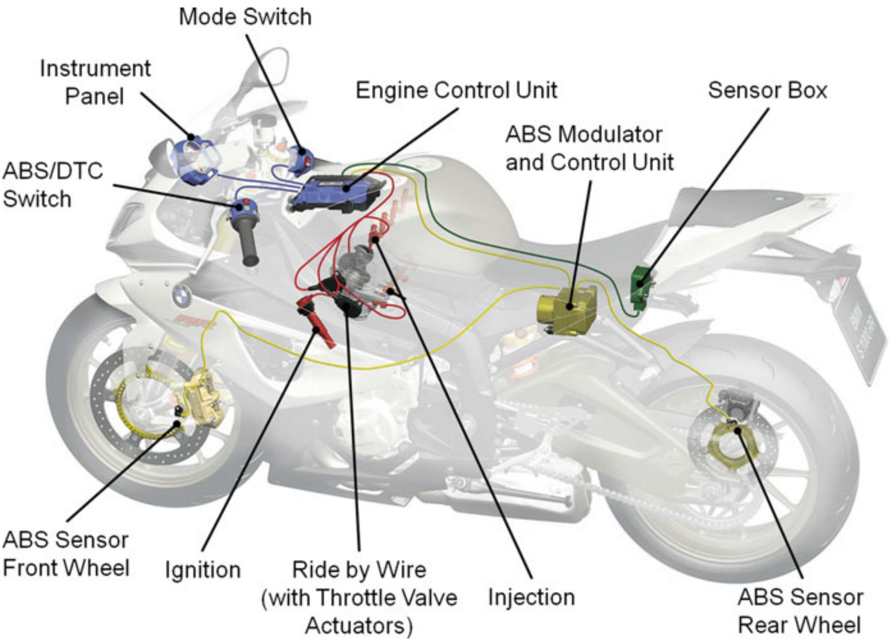


Fig. 11 System overview of DTC using the example of BMW S1000RR

deviations are automatically adjusted to the basic data in defined driving conditions by comparing the wheel speeds.

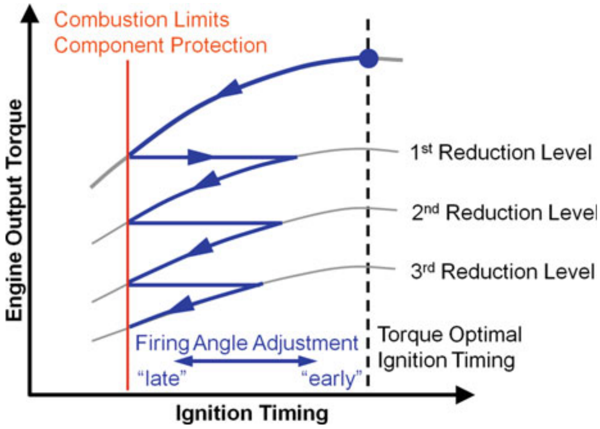
If the determined drive slip exceeds a reasonable value for ensuring vehicle stability, the control system intervenes by reducing the drive torque. Apart from the cycle time of the calculation algorithm (usually 10 ms) and recognition time (approximately 50 ms), the response times primarily depend upon the time between two work cycles (short for a large number of cylinders and high engine speed) and the actuator used for the control intervention (see below). They are typically in the range of approximately 50–160 ms, but can be shorter for an inline four-cylinder during racing or longer for a two-cylinder boxer revving at a leisurely pace.

Whereas, thanks to the sensing of driving conditions, DTC can operate very close to the physical limits as suitable for racing, establishing the ASC thresholds requires a greater compromise between sporty and safe control. Implementation is via speed-dependent threshold values, which operate reliably for all roll angles encountered during riding. The consequence of this is that, at greater leans ($\lambda > 40°$), acceleration capacity can decline noticeably with ASC.

Although, in principle, an intervention to reduce drive torque at the rear wheel can be made or supported by an active braking intervention by the ABS (Roll and Hoffmann 2010), in the commercially available systems, it is done exclusively by reducing engine torque.

The basic control strategies for this are illustrated in Fig. 12.

Fig. 12 Schematic diagram of torque reduction



Starting from an ideal ignition point for the given engine load, in a first step, the firing angle is retarded, reducing the engine torque by up to 25 %. The retarding of firing angle increases the exhaust temperature and is limited by the engine's combustion limit. Further retarding would mean that the fuel is no longer completely burnt. Therefore, a maximum limit value for ignition retarding is stored in the control unit for each operating point. If the drive slip at the rear wheel is still too high, despite maximum ignition retard up to the combustion limit, fuel injection is then restricted. This is done for selected cylinders using special restrictor patterns in various reduction stages. A further continuous reduction in engine torque is possible within the reduction stages by varying the ignition point to retarded ignition points. When the combustion limit is reached once more, the engine control unit switches to the next reduction stage: This means that further injections are prevented for each work cycle (second and third reduction stage). In the final reduction stage, injection is completely suppressed, so that the engine only continues to run in towing mode. In the case of vehicles with eGas, it is also possible to superimpose these interventions by adjusting the throttle valves and intake airflow as a slightly less reactive control channel. To prevent the engine from stopping and the rear wheel from locking up, depending upon the engine design, the torque reduction is suppressed below an engine speed of approx. 1,200–1,800 min^{-1}. In this speed range (approx. 5–15 km/h, depending upon active gear), keeping the engine running is just as important to maintain vehicle stability and is therefore prioritized over a reduction in drive torque.

The transitions into the reduction stages to reduce drive slip are adapted to the riding and slip conditions. Resetting, on the other hand, takes place as quickly as possible so as not to restrict acceleration capacity unnecessarily.

Additional features already possible with the ASC or similar traction control systems are the detection and prevention of acceleration-related front wheel lift off (so-called wheelies) or even backward flip-overs, as well as the adaptation to off-road vehicles. If the rider causes a wheelie by accelerating hard, the front wheel necessarily slows down relative to the rear wheel. The ASC recognizes this as rear wheel slip and reduces the drive torque. The slip thresholds for roads are often not suited to off-road use. Therefore, additional off-road adjustments have been developed to take into account the particular slip characteristics of loose terrain such as sand and gravel with higher threshold values. It is possible to switch between the setups or even turn the system off.

The control function of the DTC system that is suitable for racing is adjustable between four different modes (“rain,” “sport,” “race,” and “slick”). The latest version of DTC (such as in the BMW S1000RR HP4) provides the so-called Launch Control via a sensor with additional measurement of longitudinal acceleration in combination with a gear shifting assistant. This allows maximum acceleration from standstill with “suspended” front wheel, as in a racing start. In addition, the “Race Calibration Kit,” which is offered as an accessory, provides the option for individual fine-tuning.

While sensory measurement of the driving conditions in high-end systems such as DTC or Bosch MSC/MTC (Sect. 5.3.1) provides the preconditions for further additional functions – such as targeted drift control or even a “Wheely Automatic” – there are far simpler systems on the market. These are typically offered for racing purposes, sometimes even as an add-on, and are often restricted to sole monitoring of the rear wheel. From the rotation speed, gear, throttle position, and increase in rotation speed, the stored algorithm detects an unusually large increase in rear wheel rotation speed and, in the same way as ASC, intervenes via the engine management system.

7 State-of-the-Art Suspension Adjustment Systems

Unlike in a car, rider, passenger, and additional loads can easily represent 50 % or more of the overall mass of the motorcycle system, which has a significant influence upon the center of gravity and driving behavior. Simpler conventional suspensions therefore offer a manual adjustment of the rear spring preload, while more elaborate designs include adjustable spring preload and damping in the rebound and compression stages on both wheels. Systems such as BMW’s Electronic Suspension Adjustment (ESA) provide this electromechanically at the push of a button. A second generation (ESA II) even offers variable spring stiffness by series connecting the steel spring with an elastomer spring. Whereas, for safety reasons, load adjustment can only be made at standstill, damping can be adapted to the road surface and driving mode in preset characteristic curves while the vehicle is moving. Since 2012, semi-active suspension technology known from the automobile industry has also been used in motorcycles.

By means of constant sensory measurement of the driving conditions and adapting damping to the situation, semi-active suspensions (SAS) should reconcile the conflict between sportiness and road safety (measured, e.g., by improved road contact due to reduced variations in wheel load) and comfort (measured, e.g., by lower vertical accelerations of the sprung mass).

So-called Continuous Damping Control (CDC) from ZF/Sachs, which uses electrically controlled proportional valves for variable damping, is currently the most widely used system. Despite their common technical basis, the design of the systems from the various manufacturers differs in terms of the type and number of sensors used, amongst other things, and consequently the control strategy also. Alongside spring travel sensors (e.g., BMW, Aprilia), or acceleration sensors on wheel carriers and superstructure (Ducati), pressure sensors are also used on the front fork (Aprilia) to detect chassis movements. Additional information about driving conditions (accelerations, braking, cornering, etc.) is available through networking with other control systems. Whether they go under the name of Dynamic Damping Control (DDC) or Dynamic ESA at BMW, Ducati Skyhook Suspension (DSS), Aprilia Dynamic Damping (ADD), or something else, what these systems all have in common is the trend towards increasing system integration – in the sense of Global Chassis Control (GCC) or Integrated Chassis Management (ICM). This means that the engine control unit, traction control, braking control, and semi-active suspension do not just coexist, operating as individual systems, but rather that their control activities are increasingly being coordinated to suit the driving situation.

For example, it was demonstrated in a road test with a car (BMW X5) that stopping distance could be shortened by 1.2 % by coordinating CDC and ABS (Reul 2011). A simulation study with a sport touring motorcycle even found a potential reduction in stopping distance of 2–4 % (Wunram et al. 2011).

Furthermore, SAS can also have a positive influence on the course of highsider accidents by stiffening the damping of the superstructure and discharging the preload energy stored in the springs, thereby also diminishing the typical "catapult effect." With their comparatively high damping forces – even at low damper speeds and system response times of below 15 ms over the entire adjustment range – electro-rheological dampers offer the best preconditions for this (Funke et al. 2010).

While semi-active suspensions can only influence the damping forces against the direction of motion of the wheel suspension, fully active systems allow adjustment of forces in both directions. However, from a physical perspective, even a highly dynamic fully active suspension such as that offered by BOSE (BOSE 2004) can hardly assist the rider in terms of stabilization (Seiniger 2009). Nevertheless, it is reasonable to assume that, by facilitating improved handling, SAS supports the rider's ability to stabilize the vehicle. It is also likely that the proven gain in comfort (Wunram et al. 2011) will have a beneficial effect, in that the rider will not tire so easily and will have greater confidence in the abilities of the machine. Ultimately, also a more relaxed rider is an important contributor to the active safety of the whole human-machine-environment system.

8 Future Vehicle Dynamics Control Systems

By taking account of the driving conditions (especially the roll angle) in braking and traction control systems, possibly even combined with a semi-active suspension, the driver assistance systems currently available on the market already cover a large number of driving situations. In order to assess the feasibility of more advanced vehicle dynamics control systems, first of all the question of relevant accident categories must be examined. Apart from cornering accidents in general (cf. Schneider 2013; Kühn 2009), a detailed analysis of the accident database of the German Insurance Association (GDV) – and expert consultations – indicated that unbraked cornering accidents constitute by far the largest group of accidents that is potentially still possible to influence (Seiniger et al. 2008).

As already mentioned in Sect. 4, the use of predictive systems – such as "predictive brake assist" or even "autonomous emergency braking" (AEB) – bears considerable potential to prevent accidents where there was no braking or the brakes were applied too little or too late, or at least to reduce their severity (cf. DEKRA 2010; Roll et al. 2009; Roll and Hoffmann 2010).

These systems, based on suitable environment sensors, do not assist with stabilization on their own but require special measures to keep the vehicle on a stable course in the event of an autonomous intervention. Thus the investigation of rider coupling and its dynamic interaction with the vehicle and control systems should play an important role in the future.

8.1 Potential Ways to Influence Braked Cornering Accidents

In addition to the curve-adaptive braking system recently introduced into series manufacture along with MSC by Bosch and KTM (Sect. 5.3.1), the so-called brake steer torque avoidance mechanism (BSTAM, in German: Bremslenkmomentverhinderer, BLMV) from Weidele (cf. Weidele 1994; Schröter et al. 2010, 2012, 2013) is a further option for beneficially influencing cornering accidents with braking-related righting behavior (Sect. 3).

In simplified terms, the operating principle of the brake steer torque avoidance mechanism is that the kinematic steering axis is shifted respectively inclined sideways so that its projection always runs through the front tire contact patch in the frontal view of the vehicle (see Fig. 13b, c as well as Fig. 14). A brake force acting there consequently no longer has a lever arm to the steering axis, does not generate any disruptive brake steer torque (BST), and therefore has no righting of the vehicle either.

However, such a system also interferes with the chassis geometry, which, especially in the case of modern sports bikes, is designed for a practically steering torque neutral behavior during free cornering. It is of crucial importance that the normal and lateral forces acting at the front tire contact patch also have lever arms

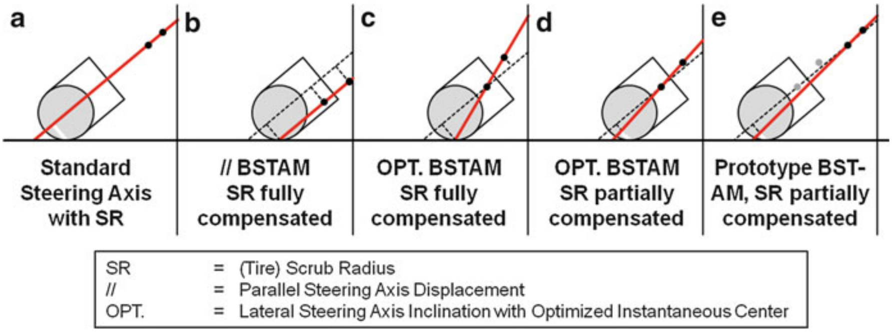


Fig. 13 Steering bearing positions, kinematic steering axis, and (partial) compensation of the (tire) scrub radius (SR) for standard steering and various configurations of a brake steer torque avoidance mechanism (BSTAM)

to the steering axis: Lateral forces turn the steering outwards via the trail, and normal forces turn it inwards, while both turn it outwards via the tire scrub radius.

If the tire scrub radius is eliminated by a brake steer torque avoidance mechanism with a parallel offset steering axis (in its original plane tilted around the steering head angle τ, Fig. 13b), this outward steering torque component is lost, and the rider has to apply a much higher steering torque during free cornering. Although this can theoretically be controlled by a greatly increased steering head angle (for the test motorcycle, e.g., approximately 50° instead of $23°55'$) in combination with similarly increased fork yoke offset (e.g., 140 mm instead of 30 mm) and only partial compensation of the tire scrub radius, this would seriously compromise handling characteristics. The more elegant option, which retains the basic geometry, consists of lateral steering axis inclination: This gives the outward effect of the side force a higher weighting than the inward-turning normal force, so that, despite complete compensation of the tire scrub radius, the original balance is restored (Fig. 13c). The geometry of the basic chassis defined by the steering head angle, fork yoke offset, and tire dimension establishes the kinematically optimum instantaneous center of steering axis inclination at the intersection of the standard steering axis with the vertical running through the front wheel hub (Fig. 14) in upright vehicle position. Regardless of the lateral inclination angle of the steering axis, this allows free cornering with the same steering torque requirement as for the standard setup. However, during corner braking with complete compensation of the tire scrub radius (Fig. 13c), this setup produces a steering torque requirement that decreases with increasing deceleration – whereas a rider who is used to conventional vehicles would intuitively expect the opposite, that is to say an increasing steering torque requirement. This familiar feedback can also be restored by reducing the steering axis inclination angle and therefore only partially compensating the scrub radius (Fig. 13d).

However, the practical implementation of such a system presents two major challenges: First of all, there is the previously described optimized instantaneous

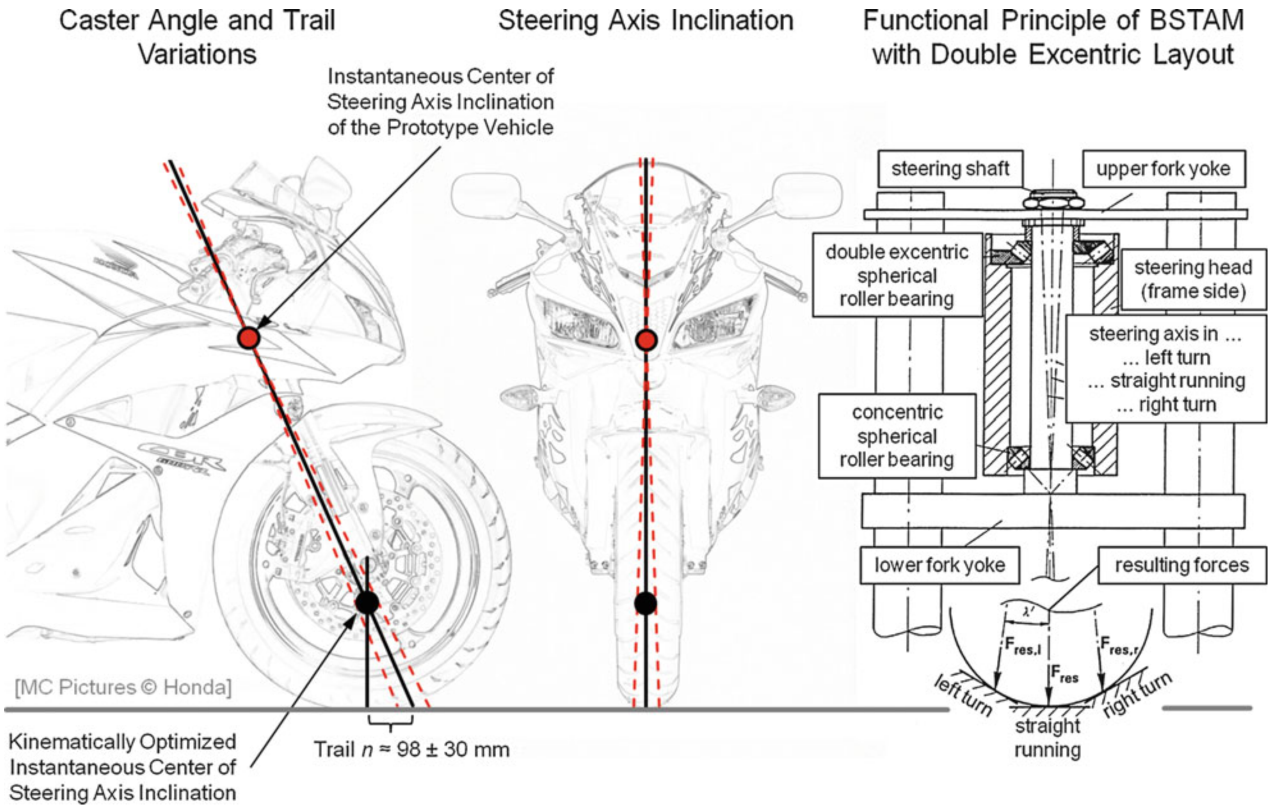


Fig. 14 Changes in chassis geometry and operating principle of brake steer torque avoidance mechanism according to Weidele (1994) using the example of the Honda CBR 600 RR test vehicle

center of rotation for standard chassis parameters and tire dimensions below the wheel hub (approx. 74 mm in the case of the test motorcycle with a front tire of the typical dimension 120/70ZR17; see Fig. 14). The obvious design of the brake steer torque avoidance mechanism based on hub-center or king-pin steering therefore necessarily brings the disadvantage of greater tire-sprung mass. Using other suspension/steering systems theoretically allows to avoid this by lateral adjustment of both steering bearings outside of the wheel's circumference. Practically, this necessitates large adjustment ranges, bringing downsides in construction space, design, and surplus mass; however, this time of superstructure sprung mass. Secondly, the wheel inertia must always initially be decelerated at the start of braking, before brake slip and braking forces can be generated. The required steering axis inclination angles of up to 14° give rise to outward steering torque components. With an estimated value of 10 Nm, these theoretically constitute a serious conflict to the aim of reducing brake steer torque. However, since they only occur in the first approx. 0.1–0.2 s of braking (cf. Fig. 15b), it is reasonable to assume that, in practice, they can be controlled by an appropriate buildup in brake pressure.

In order to investigate driving behavior in real riding experiments, a Honda CBR 600 RR supersport motorcycle (2010 model, including C-ABS, Sect. 5.2.3) was fitted with a brake steer torque avoidance mechanism (Fig. 14). The steering head bearings are kinematically designed as spherical joints (using spherical roller bearings), whereby the upper steering head bearing is electromechanically adjustable via a double excentric construction (Weidele 1994; Schröter et al. 2010). Since the construction space is severely restricted by the legs of the telescopic fork, excentricity is just 8 mm. Essential changes in steering head angle and trail

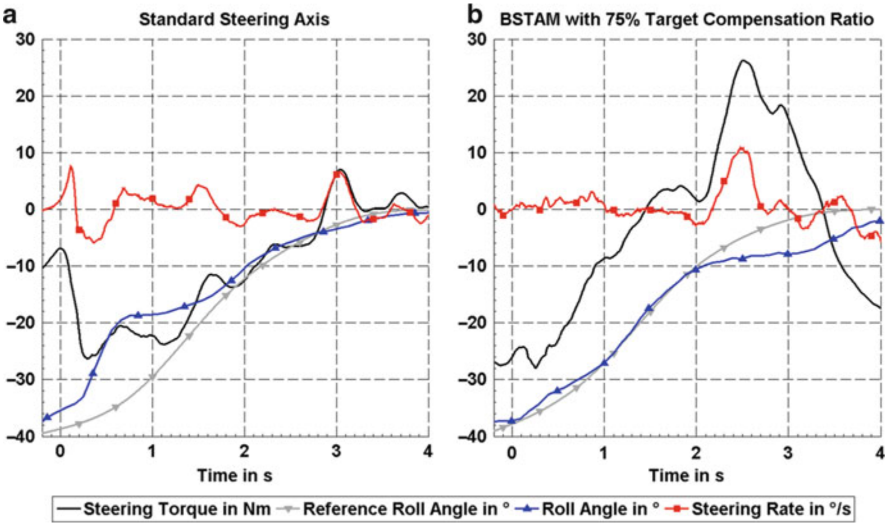


Fig. 15 Course over time of characteristic measurements during corner braking with and without brake steer torque avoidance mechanism ($R = 50$ m, $v_0 \approx 18$ m/s, $a_{y0} \approx 6$ m/s^2, $a_x \approx 5$ m/s^2)

therefore remain as small as the tilt angle of the steering shaft of around 2° (Figs. 13e and 14). The overall steering torque requirement of the real system is therefore similar to that of a parallel brake steer torque avoidance mechanism (Fig. 13b) with correspondingly reduced compensation rate. This gives rise to a significantly increased steady-state steering torque, a reduced steering torque jump at the start of braking – with only slight disruption due to slowing of wheel inertia – and consequently also only slight steering, rolling, and course deviations.

By way of an example, Fig. 15 shows the courses over time of some characteristic measurement variables during braking under typical rural road conditions with an average deceleration of around 5 $\mathrm{m/s^2}$ from a speed of approximately 18 m/s (initial lateral acceleration $a_{y0} \approx 6\ \mathrm{m/s^2}$, initial roll angle $\lambda_0 \approx 35°$) in a left-hand turn with a radius of $R = 50$ m for the standard suspension and active brake steer torque avoidance mechanism (in both cases without the standard HESD steer damper, Sect. 2).

When traveling with conventional steering geometry and centered steering axis (Fig. 15a), after the obligatory disengagement of the clutch, the rider applies a constant steering torque of 7–9 Nm towards the outside of the curve (negative in value). At the start of braking ($t = 0$ s), a jump in steering torque of around 19–20 Nm can be observed ($t \approx 0.2$ s). The proportion of the suddenly increased overall steering torque demand which is not immediately covered by the rider's steering effort accelerates the steering system rotationally towards the inside of the curve (see positive steering rate), and the subsequent righting motion allows the roll angle (negative in left-hand curves) to drop significantly below the ideal value for the speed and curve radius. From this point on, the overall steering torque demand is already significantly reduced by a superimposed gyroscopic steering torque component on the more upright vehicle position and is once again in balance with the rider's effort. The oscillations in the measured steering torque resulting from coupling of steering and rolling dynamics decay with the further declining steering torque requirements until the end of braking. In order to support the motorcycle at standstill, towards the end of braking (from $t \approx 2.5$ s), the rider takes both legs off the footrests, one after the other, and makes compensating steering and upper body movements to maintain balance. This causes disturbances in the measurements of steering angle and torque as the rider steadies himself on the handlebars.

Compared with the standard steering, corner braking with brake steer torque avoidance mechanism (Fig. 15b) starts with a much higher constant steering torque in the order of 24–25 Nm (as opposed to 7–9 Nm at $t = 0$ s). Apart from an expected small disruption due to inertia of around 1–2 Nm at the start of braking, there is a steering torque jump of only 3 Nm (compared with 18–20 Nm at $t \approx 0.2$–0.3 s), so that brake steer torque only causes a negligible steering angle disturbance and the roll angle initially follows the reference very closely. Because the simple control algorithm that has been chosen with constant compensation ratios that ignore additional effects upon steering torque, the brake steer torque is overcompensated from the middle of braking onwards and the steering torque demand changes sign (from $t \approx 1.4$ s). The roll angle remains greater than the reference value and the bike decreases its turning radius ($t \approx 2$ s); that would be a significant advantage in a narrowing radius turn. However, in the illustrated example, the curve radius is constant so that the rider has to steer harder into the curve ($t \approx 2$–2.5 s), in order to follow it. Although a glance to check the mechanics tilts the rider's upper body further inwards, the roll angle no longer fully approximates to the ideal value because of imminent standstill.

In conclusion it can be stated that the prototype brake steer torque avoidance mechanism effectively reduces the steering torque jump and the associated steering and righting movement at the start of braking. In contrast to MSC (Sect. 5.3.1), the almost complete prevention of brake steer torque offers advantages for course corrections "on the brake" and in respect of future systems such as predictive brake assist or autonomous emergency braking.

While it would be relatively simple to allow for the previously illustrated change of sign of steering torque via a variable compensation ratio (e.g., as a function of the roll angle, deceleration, or brake pressures) and the disruptive effects of slowing wheel inertia seem to be controllable via limited brake pressure gradients, there are obvious disadvantages to the mechanical complexity and weight of the system. Moreover, it is to be expected that optimization of high-speed stability and handling characteristics, which have currently only been rudimentarily investigated, would require a high development effort. From a current perspective, it therefore seems a better solution to predict the brake steer torque on the conventional chassis based on sensor data and counteract it, for instance, via an electrical actuator or targeted control of a semi-active steering damper. With their comprehensive sensor setups and control, modern brake systems such as C-ABS (Sect. 5.2.3) and MSC (Sect. 5.3.1) offer the best preconditions for this.

8.2 Potential Ways to Influence Unbraked Cornering Accidents

The largest group of accidents that can potentially still be influenced is unbraked cornering accidents (Seiniger et al 2008). Typically they occur as a result of a sudden drop in road surface friction (e.g., due to leaves, slippery asphalt, sand, or ice) or exceeding the maximum possible lateral acceleration. In both classes of accident, the lateral acceleration drops and no longer "suits" the roll angle as required to maintain stability through roll equilibrium; a fall is the inevitable result. In order to influence these two accident categories with a technical system, it must firstly be possible to detect them by sensors and secondly to influence them by technical measures.

Experiments and simulations have shown that the sideslip speed of the vehicle (speed of the change in vehicle sideslip angle) is a reliable criterion for detecting critical driving situations. In normal driving situations, the slip of the motorcycle tires is usually low; even the sideslip angle is small. Sideslip speed is therefore limited. However, in critical driving situations – when both wheels skid – the sideslip of the vehicle is unstable. In order to prove the suitability of sideslip speed as a criterion, unbraked cornering accidents were simulated on a low friction surface with a specially equipped motorcycle. The sideslip speed of a vehicle is

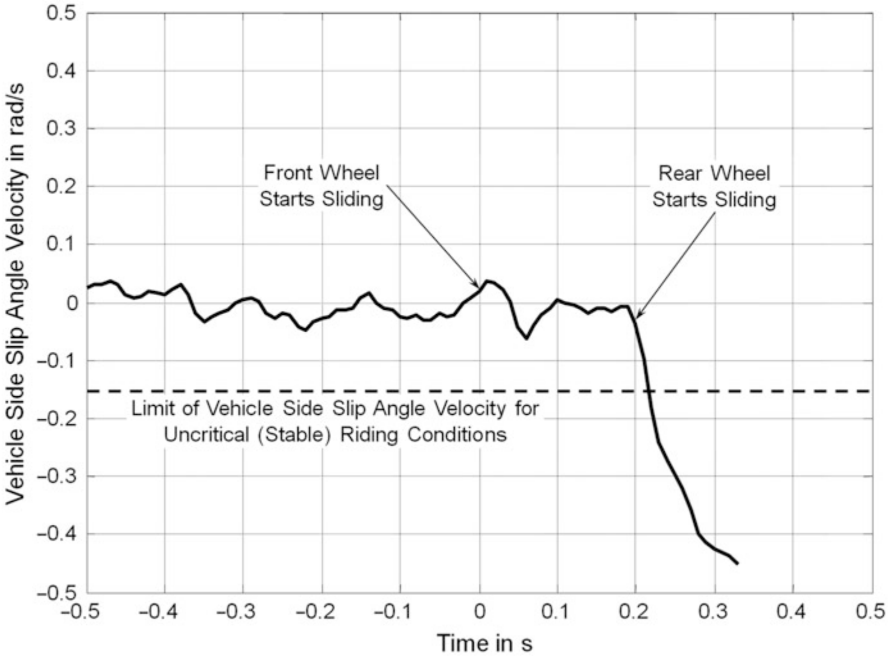


Fig. 16 Course over time of sideslip speed during a friction jump. Contacting the slide surface at a time $t = 0$ s

$$\dot{\beta} = \frac{\dot{\psi} + a_{\mathrm{y}}}{v} \tag{5}$$

with the road surface-related (leveled) variables, yaw rate $\dot{\psi}$, lateral acceleration a_{y}, and vehicle speed v. Direct measurement of road surface variables is not possible for motorcycles for two reasons:

- The sensors mounted on the vehicle also tilt into the curve.
- Roll speed and acceleration cause additional inertial forces acting in the sensor so that correction is necessary.

Fixed sensors for yaw rate, roll rate, lateral acceleration, vertical acceleration, and roll angle are required on motorcycles to determine sideslip speed. Figure 16 shows the course of sideslip speed during a typical accident of the “friction coefficient step” category. The front wheel of the motorcycle contacts the low friction surface at time $t = 0$ s: There is a small deflection in sideslip speed, which is, however, obviously corrected – because the rear wheel is not slipping, the vehicle is initially still stable. At time $t = 0.2$ s, the rear wheel is also on the slippery surface and the vehicle is noticeably building up sideslip speed. And so the “friction coefficient step” class of accident progresses in two phases: Each phase is characterized by incipient sliding of a wheel. In the "exceeding maximum lateral acceleration" class of accident, both wheels start to slide at approximately the same time. The expected maximum sideslip angle of the vehicle for a large roll angle is around 2°, and the expected maximum sideslip speed of the motorcycle is around 0.15 rad/s for stable driving situations. In each driving test evaluated, the sideslip speed limit was exceeded during the falling over phase: Under no circumstances was the limit exceeded during noncritical driving situations. Critical driving situations can therefore be detected from the sideslip speed. It is conceivable that in future Kalman filters, similar to those used in the ESC system for cars, can be used to determine the sideslip angle and improve detection of critical driving situations.

However, the aim of a vehicle dynamics control system is not only to detect but also to stabilize the critical driving situation. Roll instability obviously very quickly results in the vehicle falling over by exceeding the geometrical roll angle limits. In contrast to a yawing motion, which does not restrict the duration of the critical driving situation, as long as there is sufficient road space available, the rolling motion limits the time available to stabilize the vehicle. The primary aim of a vehicle dynamics control system must therefore be to stabilize the roll angle (Seiniger 2009; De Filippi et al. 2011b). When it comes to a fall for the case in point, the yawing motion is initially what is required – a vehicle turning into the curve, already skidding on the road and turning away from the skidding rider. A vehicle skidding on the road surface and turning out of the curve would push the rider ahead of it. Because the skidding bike has a much lower friction coefficient than the rider's usual protective clothing, this would increase the skid distance and therefore also the risk of injury to the rider. In the case of irreversible destabilization, sideslip speed turning into the curve is far preferable (over-steering). Yaw instability is critical, if the still skidding vehicle enters into an area of high friction with one or both wheels: The wheels then contact a gripping surface with a greatly increased slip angle, thereby producing far too great side forces, which usually causes the vehicle to tip over on the outside of the curve (highsider accident). This often happens so quickly that the rider is unable to stabilize the vehicle. Moments of a three-figure Nm range around the steering system are expected for front wheel transition from a low friction to a high friction surface: Depending upon the coupling of the rider and the elasticity in the steering system, these moments can result in kickback; furthermore, turning of the handlebars towards the outside of the curve can result in a reduction in the side force due to the negative sideslip angle at the front wheel. For negative sideslip values, the side force is less than that required to counteract the tipping moment. Hence, a vehicle dynamics control system must aim to limit the sideslip angle at the front wheel in the phase of transition from low to high friction to values of 0° in order to prevent kickback. In a conventional bike, too much side force at the rear wheel cannot be adequately dissipated by steering movements and results in the previously mentioned highsider accidents. Large sideslip angles at the rear wheel must therefore be prevented when transitioning from low friction to high friction surfaces. As a potential measure to influence the vehicle dynamics, the tire forces can principally be varied.

The yawing rotation of the vehicle can be influenced by targeted brake interventions and/or the use of highly dynamic active suspensions: This means that a control system to avoid highsider accidents with high to low to high friction variations is feasible.

Apart from using multitrack tilting wheel technology, e.g., in the Piaggio MP3 (Roll and Hoffmann 2010), it is not possible to make a vehicle dynamics control system for single-track vehicles similar to the electronic stability control (ESC) system from the automobile sector with the abovementioned measures, because of the rolling instability of the system (Seiniger 2009) or – even if there is adequate traction available – only within very limited boundaries (De Filippi et al. 2011b).

8.2.1 Roll Stabilization by Double Gyroscope

However, a possible means of roll stabilization is a double arrangement of gimbaled gyroscopes – invented for single-track trains already at the beginning of the twentieth century. The setup has been further developed by the American company LIT Motors over the last few years and has been patented for use in an electrically driven cabin motorcycle (Kim et al. 2013). The current “C-1” prototypes use two gyroscopes rotating in opposite directions around the vertical axis, and their mountings can be independently turned around the transverse axis relative to the vehicle body. While the gyroscopic reaction moments eliminate one another during normal travel, a considerable rolling moment (in the example, up to 2.3 kNm) can be imposed as a function of gyroscopic inertia (e.g., 0.07 kgm^{2}of each gyroscope) and rotation and swivel speed (approx. 15,000 min^{-1} and 100 min^{-1}). For improved protection of the occupants in a typical side-on collision at an inner-city junction, this should be sufficient to prevent the bike falling over sideways, but instead allow it to be pushed away, like a car. Influencing the roll equilibrium also makes it possible to ride through curves at a different rolling angle from the usual value or even completely upright – which makes it easier, for example, to filter through traffic jams. Furthermore, the stabilizing gyroscopes also help to increase the efficiency of electrically powered vehicles, as they can also be used as energy accumulators. At standstill and at lower speeds, they rotate very quickly to cover the stabilization requirement; at higher speeds gyroscopic stabilization is increasingly provided by the wheels. The rotation speed of the additional gyroscopes can therefore be reduced and the energy that is released can be used to accelerate the vehicle. This is reversed on deceleration and braking energy is recovered by renewed acceleration of the stabilizing gyroscopes.

However, it remains to be seen to what extent this technology – which is also promising in the sense of an emergency brake function – proves itself in the cabin motorcycle and can be transferred to conventional motorcycles with their limited construction space.

Acknowledgments The authors would like to thank Jörg Reissing, Alfred Eckert, and Jürgen Bachmann, who, as co-authors of the first two editions, made a huge contribution towards the writing of this chapter. Special thanks also go to all our contacts at BMW Motorrad, KTM, Honda, Bosch, and Continental for providing us with technical information and images.

References

ACEM (European Association of Motorcycle Manufacturers) (2009) MAIDS: motorcycle accidents in depth study – in-depth investigations of accidents involving powered two wheelers. www.maids-study.eu, Final report 2.0, Published Apr 2009

ADAC (2010) Viel zu wenige Motorräder haben ABS (Far too few motorcycles have ABS). ADAC Press Release, Published 1 Apr 2010

Bayer B (1986) Das Pendeln und Flattern von Krafträdern (Weave and wobble of motorcycles). Institut für Zweiradsicherheit (Institute for Motorcycle Safety), Bochum (meantime moved to Essen)

Böhringer U (2013) Adaptives Dämpfungssystem – Fahrwerks-Finessen beim Motorrad (Adaptive damping control – finesse for motorcycle chassis). www.faz.net. Published 27 Sept 2013

Bosch (2013a) Motorrad-Stabilitätskontrolle von Bosch geht in Serie (Bosch motorcycle stability control goes into series production). www.bosch-presse.de, PI 8314. Published 23 Sept 2013

Bosch (2013b) Eine Million Motorräder fahren mit ABS von Bosch (One million motorcycles fitted with Bosch ABS). www.bosch-presse.de, PI 8362. Published 5 Nov 2013

BOSE Corporation (2004) Bose suspension system. www.bose.com. Published 2004

DAT (Deutsche Automobil Treuhand) (2014) DAT report 2014. www.dat.de/report. Accessed 22 Dec 2014

De Filippi P, Tanelli M, Corno M, Savaresi S, Fabbri L (2011a) Semi-active steering damper control in two-wheeled vehicles. IEEE Transactions Control Systems Technology 19(5):1003–1020

De Filippi P, Tanelli M, Corno M, Savaresi S (2011b) Enhancing active safety of two-wheeled vehicles via electronic stability control. In: 18th IFAC world congress, Milan

DEKRA (2010) DEKRA Verkehrssicherheitsreport Motorrad 2010 (DEKRA motorcycle traffic safety report 2010). www.dekra.com. Published in April 2010

DESTATIS (2013) Verkehrsunfälle Zeitreihen 2012 (Time series of road accident data in Germany 2012), Sections 5.1.2 (2) and 7.4, last figures for 2010. www.destatis.de.Published. 10 July 2013

EU (European Union) (2013) Regulation (EU) No 168/2013 of the European Parliament and the Council of 15 January 2013 on the approval and market surveillance of two-, three-, and four-wheeled vehicles. http://eur-lex.europa.eu/. Accessed 22 Dec 2014

Funke J (2007) Belastung und Beanspruchung von Motorradfahrern bei der Bremsung mit verschiedenen Bremssystemen (Stress and strain of motorcyclists when braking with different brake systems). VDI-Verlag, Düsseldorf

Funke J, Savaresi S, Spelta C (2010) Electrorheological dampers as a basis for semi-active motorcycle suspensions. In: Proceedings of the 8th international motorcycle conference, Cologne, pp 211–224

Gwehenberger J, Schwaben I, Sporner A, Kubitzki J (2006) Schwerstunfälle mit Motorrädern (Most serious accidents with motorcycles). In: VKU Verkehrsunfall und Fahrzeugtechnik 01/2006. Springer Fachmedien München GmbH, Munich, pp 11–18

Hurt H, Ouellet J, Thom D (1981) Motorcycle accident cause factors and identification of countermeasures (Final report). U.S. Department of Transportation, NHTSA, Washington, DC

IIHS and HLDI (Insurance Institute for Highway Safety and Highway Loss Data Institute) (2013) Status report, vol 48, no 4, 30 May 2013

IRTAD (International Traffic Safety Data and Analysis Group) (2013) Road safety annual r2013, OECD International Transport Forum, Paris

Kim D, Bretney K, Shao A, Tsang A (2013) Electronic control system for gyroscopic stabilized vehicle, Patent US 8532915-B2

Kühn M (2009) An analysis of motorcycle accidents, no. 5. http://udv.de/en/publications/compact-accident-research. Accessed 22 Dec 2014

Landerl C, Deissinger F, Wagner H-A, Jahreiss H-J (2010) Erweiterte Fahrerassistenz durch die Verknüpfung der Motor- und Fahrwerksregelsysteme der BMW S 1000 RR (Enhanced rider assistance via connection of the engine- and suspension control systems of the BMW S 1000 RR). In: Proceedings of the 8th international motorcycle conference, Cologne, pp 362–377

Nishikawa Y, Nanri T, Takenouchi K, Takayanagi S, Tani K, Fukaya S (2008) Experimental study of pitching control of large motorcycles with short wheelbases using a brake-by-wire system. In: Proceedings of the 7th international motorcycle conference, Cologne, pp 430–446

Reissing J, Wagner H-A, Jahreiss H-J, Bachmann J, Müller P (2006) Integral ABS und ASC – Die neuen Fahrdynamik Regelungssysteme von BMW Motorrad (Integral-ABS and ASC – the new driving dynamics control systems by BMW Motorrad). In: Proceedings of Brake.tech conference, Munich

Reul M (2011) Bremswegverkürzungspotential bei Informationsaustausch und Koordination zwischen semiaktiver Dämpfung und ABS (Braking distance reduction potential through information exchangeand coordination of semi-active damping with ABS control). VDI-Verlag, Düsseldorf

Roll G, Hoffmann O (2010) Evaluation of the safety benefits of electronic brake-controlsystems in single-track vehicles. In: Proceedings of the 8th international motorcycle conference 2010, pp 423–468

Roll G, Hoffmann O, König J (2009) Effectiveness evaluation of anti lock brake systems (ABS) for motorcycles in real-world accident scenarios. In: Proceedings of the 21st international technical conference on the enhanced safety of vehicles (ESV), Stuttgart. www.nhtsa.gov/ESV. Accessed 22 Dec 2014

Schneider R (2013) Vollbremsung in Schräglage (Slamming the brakes while cornering). In: Motorrad 23/2013. Motor Presse Stuttgart, Stuttgart, pp 52–55

Schneider R (2014) Schrecklage und Schrägbremsen (Panic Situation and Braking while Cornering). In: Motorrad 04/2014, Motor Presse Stuttgart, Stuttgart, pp. 38-41

Schröter K, Bunthoff J, Fernandes F, Schröder T, Winner H, Seiniger P, Tani K, Fuchs O (2010) Brake steer torque optimized corner braking of motorcycles. In: Proceedings of the 8th international motorcycle conference, Cologne, pp 396–422

Schröter K, Wallisch M, Vasylyev O, Schleiffer J-E, Pleß R, Winner H, Tani K, Fuchs O (2012) Update on brake steer torque optimized corner braking of motorcycles. In: Proceedings of the 9th international motorcycle conference, Cologne, pp 2–46

Schröter K, Wallisch M, Weidele A, Winner H (2013) Bremslenkmomentoptimierte Kurvenbremsung von Motorrädern (Brake steer torque optimized corner braking of motorcycles). In: ATZ 05/2013. Springer Vieweg, Wiesbaden, pp 436–443

Seiniger P (2009) Erkennbarkeit und Vermeidbarkeit von ungebremsten Motorrad-Kurvenunfällen (Technical potential to detect and avoid un-braked motorcycle accidents in curves). VDI-Verlag, Düsseldorf

Seiniger P, Winner H, Schröter K, Kolb F, Eckert A, Hoffmann O (2006) Entwicklung einer Rollwinkelsensorik für zukünftige Bremssysteme (Development of a roll angle sensor technology for future brake systems). In: Proceedings of the 6th international motorcycle conference, Cologne, pp 369–388

Seiniger P, Winner H, Gail J (2008) Future vehicle stability control systems for motorcycles with focus on accident prevention. In: 9th biennial ASME conference on engineering systems design and analysis, Haifa

Sporner A, Kramlich T (2000) Zusammenspiel von aktiver und passiver Sicherheit bei Motorradkollisionen (Combination of primary and secondary safety aspects within motorcycle collisions). In: Proceedings of the 3rd international motorcycle conference, pp 55–80

Stoffregen J (2010) Motorradtechnik (Motorcycle technology). Vieweg + Teubner, Wiesbaden

Tani K, Inoue N, Kato M, Hikichi T, Thiem M (1993) Research on Traction Control System for Motorcycle. In:VDI Berichte Nr. 1025, VDI-Verlag, Düsseldorf

Tani K, Toda M, Takenouchi K, Fukaya S (2010) Research on brake-by-wire system for superbike race motorcycle. In: Proceedings of the 8th international motorcycle conference, Cologne, pp 378–395

Wakabayashi T, Sakai K (2004) Development of electronically controlled hydraulic rotary steering damper for motorcycles. In: Proceedings of the 5th international motorcycle conference, pp 489–509

Weidele A (1994) Untersuchungen zum Bremsverhalten von Motorrädern unter besonderer Berücksichtigung der ABS-geregelten Kurvenbremsung (Investigations on the braking behavior of motorcycles with focus on ABS-controlled corner braking). VDI-Verlag, Düsseldorf

Willig R, Lemejda M (2012) A new inertial sensor unit for dynamic stabilizing systems of powered two wheelers. In: Proceedings of the 9th international motorcycle conference, Cologne, pp 66–84

Wunram K, Eckstein L, Rettweiler P (2011) Potential aktiver Fahrwerke zur Erhöhung der Fahrsicherheit von Motorrädern (Potential to enhance motorcycle safety through active suspensions). Report F81 of the German Federal Highway Research Institute (BASt), Verlag Neue Wissenschaft, Bremerhaven

Yildirim F, Mörbe M (2013) Modern brake control systems and sensor systems for power two wheeler (PTW). In: Proceedings of EuroBrake conference 2013, Dresden

Vehicle Dynamics Control with Braking and Steering Intervention

41

Thomas Raste

Contents

Abstract

From a driver's perspective, controlling a vehicle means controlling the speed and the path curvature. In exceptional circumstances, e.g., in emergency evading situations, also the orientation of the vehicle has to be controlled. In a narrower sense, vehicle handling refers to vehicle dynamics like cornering and swerving and includes the vehicle stability. The advances in global chassis control technology have been used to further improve the vehicle safety and handling qualities. The effects of active systems are well understood in the context of how they contribute to the overall vehicle performance.

Altering the path curvature can easily be achieved by increasing the yaw gain such that the driver steering input is small. This strategy is only applicable up to a medium speed. The yaw rate's normal driving range decreases significantly with vehicle speed, because the available tire–road friction is quickly saturated at high speed when the steering wheel angle input is too high. The strategy at high speed therefore must be to decrease steady-state yaw gain.

T. Raste (✉)
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: thomas.raste@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_41

At the limit of friction, where safety becomes relevant, the handling controller determines how the vehicle remains stable. All available actuators are incorporated and coordinated to reach this goal. The active chassis gives the driver optimal support for avoiding accidents. In the region beyond the limit of friction, the main task of the control system is to prevent the car from skidding heavily so that the car remains on track.

During normal driving car drivers usually expect a linear yaw response of the vehicle with small phase lag. Most drivers have no experience of loss of linearity caused by saturation of tire forces. If saturation happens at the rear axle, the sideslip angle will increase quickly and therefore causes a hazardous driving problem for many drivers. The primary task of the control system should be to keep the vehicle sideslip angle small. An average driver feels uncomfortable when the magnitude of the sideslip angle exceeds a few degrees. State-of-the-art electronic stability control (ESC) systems limit the sideslip angle indirectly. ESC uses a reference yaw rate limited by the actual acceleration to account for the tire saturation. Additionally, the rate of change of sideslip angle is calculated and also limited.

Global chassis control delivers significant benefits in normal driving and particularly in emergency situations. The configuration and coordinated interaction of the active systems are the key success factors for enhancing the vehicle performance. International standards like ISO 26262 ensure quality and safety of the overall control system at the highest level.

1 Introduction

The value of modern braking systems with electronic stability control (ESC) lies in their ability to make a car behave more predictably, keep it stable over a broad range of conditions, and make it easy to control in borderline situations. Stability means that a car reacts to a maneuver in the manner expected. Handling is stable if it remains unchanged in the face of unchanged input and changes slightly only if the input changes slightly. Stability covers the normal range of driving in which a driver perceives comfort and enjoyment through the chassis setup. On the other hand, should a slight amount of input from a driver lead to major changes in handling (say in the case of a minor steering correction leading to a skid), handling will then be instable. A car will now be borderline, and car and driver will now be in a closed loop as depicted in Fig. 1. Drivers may steer, accelerate, or brake, but their commands will increasingly not translate directly into action. Instead, active systems will “filter” their commands to ensure the best and safest handling possible.

Active steering systems fall into the following categories:

- Systems that superimpose torque permit influence on steering independently of driver input. The system can give a driver a steering cue through the wheel in a critical situation.

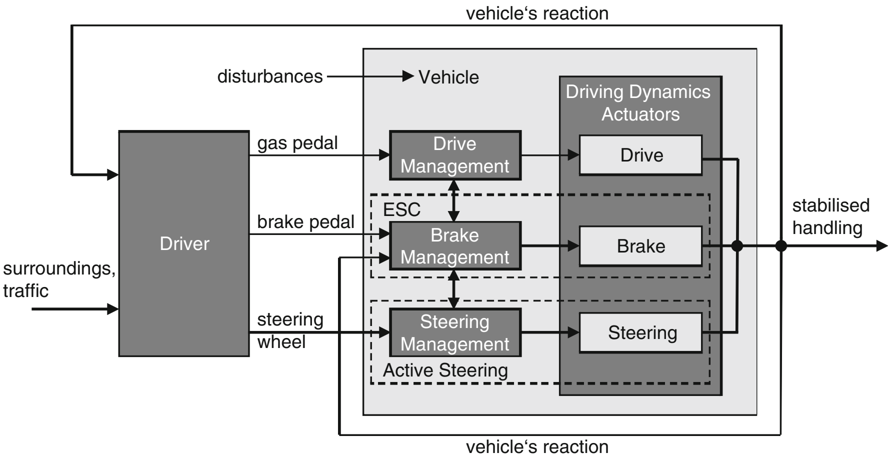


Fig. 1 Driver, car, and surroundings with ESC and active steering system form a closed loop

– Systems that superimpose steering angles make it possible to modify a steering angle that a driver has chosen for the front wheels or to modify the angle of the rear wheels dictated by the front wheels.
– Systems that superimpose both torque and steering angles combine the advantages of the above two systems. The actuators here can be concentrated locally in a single housing, thus saving space, or they can be placed separately at various places along the steering gear.
– Steer-by-wire systems are paving the way for completely new kinds of man–machine interfaces such as side-stick steering instead of a conventional steering wheel.

Active steering systems not only offer great potential for networking functions at the stability level but also for driver-assistance functions associated with keeping in lane. Figure 2 shows several functions that are currently available on a mass basis or will become available shortly.

2 Requirements for the Additional Function of Stabilization with Braking and Steering

The system context shown in Fig. 3 defines the functional units of regulating handling through steering intervention. It also defines interface requirements for working with other car systems. It is the responsibility of carmakers to decide what hardware to use and to which control devices to assign the software. One common variant is to center stabilizing functions in the ESC control device. An expanded ESC with built-in lateral-motion control employs the steering system as an actuator

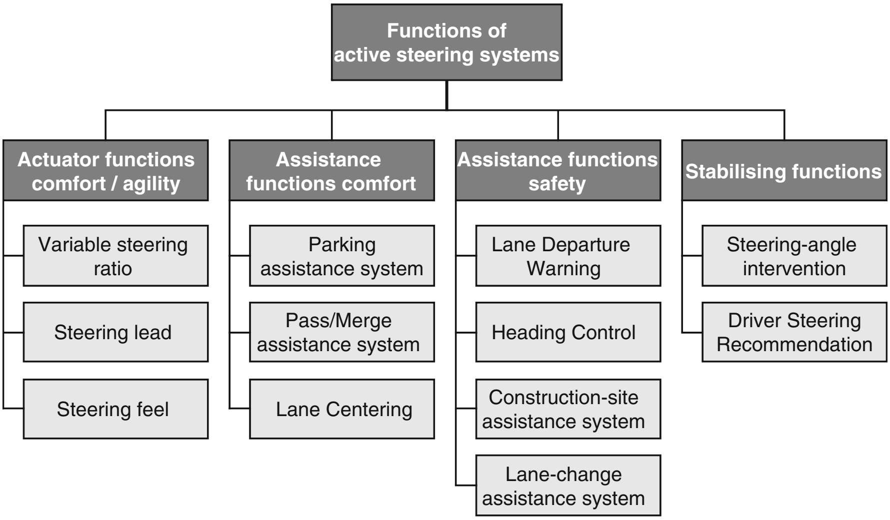


Fig. 2 Functions of active steering systems

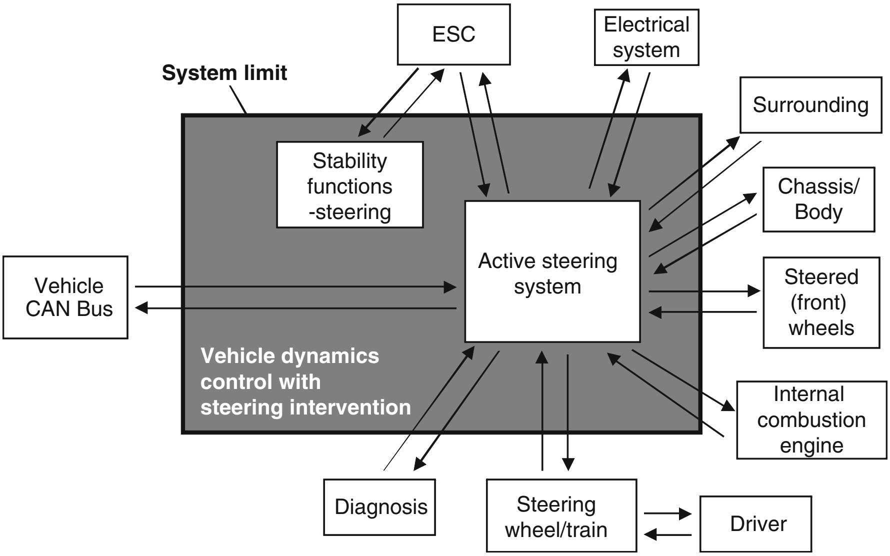


Fig. 3 System context and interfaces of handling under steering intervention

for stabilizing activities. The following catalogue of requirements for combined brake and steering interventions results for users:

– Improved lane-holding and directional stability under such conditions as changing loads, panic stops, partial braking in curves, or slaloms.

- Greater stability in extreme steering maneuvers such as emergencies or quick lane changes so as to reduce the danger of skidding.
- Less steering effort and better use of the potential for force closure when braking and accelerating, particularly on nonhomogeneous road surfaces. This will lead to improved braking distance and better traction with constant, or even better, stability.

Electronic stability control (ESC) with active steering systems opens up entirely new ways to stabilize a vehicle. A combination of braking and steering can counteract undesirable yaw reactions quickly and conveniently. Stabilizing functions come to bear principally in the following situations:

- Braking on split-μ
- Acceleration on split-μ
- Oversteering
- Understeering
- Risk of rollover
- Trailer stability

3 Concept and Principle of Braking and Steering Control

Combined braking and steering regulation is based on a graduated, cascading concept of regulation (Fig. 4). Sensors on the brakes, steering, and gas pedal capture a driver's input and compare the car's actual movements. The regulator corrects deviations by inputting ideal values, which leads to a modification of the car's

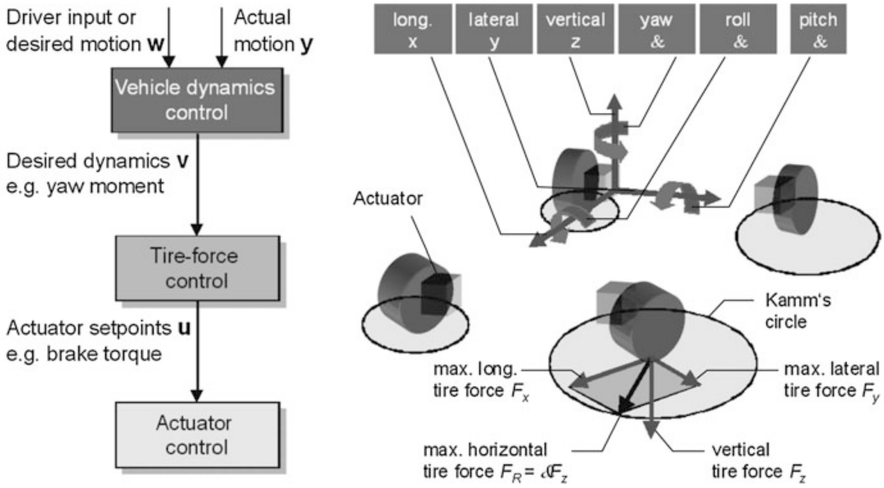


Fig. 4 Hierarchically arranged cascading control concept (tire forces are shown at road)

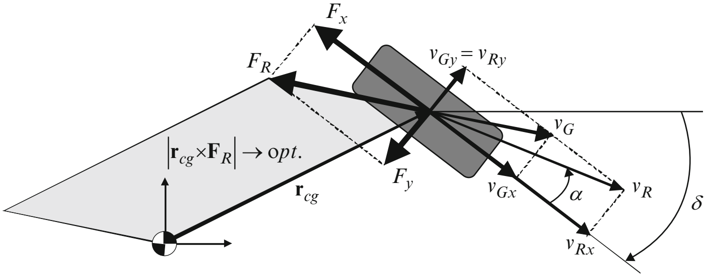


Fig. 5 Tire forces, speed, and yaw moment of a wheel

motion. The tire forces at the point of contact (tires–road) are responsible for changing direction; actuators readjust them. The frictional coefficient μ and the contact force F_z, both parameters of the Kamm circle, define the limits for the maximum horizontal tire forces at each wheel.

Figure 5 shows how a single wheel contributes to yaw motion (Salfeld et al. 2007). The horizontal forces are a function of slip, which can be derived from the sliding speed v_G and the absolute speed v_R of the wheel. The resultant F_R of the horizontal forces and the sliding speed v_G are located at opposite ends of the same line of application.

The amount of yaw moment that each wheel generates reaches its maximum when the vector product from the position vector r_{cg} from the car's center of gravity to the center of a wheel and the vector F_R of the resultant tire force are at their maximum. The amount of force F_R increases through the steering angle δ at the wheel, and simultaneously, the force and position vector can be approximated in orthogonal fashion via braking or driving. However, minimal yaw moment produced at a wheel is necessary in certain situations. Turning the steering wheel will likewise have this effect; however, the vector product, i.e., the area spanned by F_R and r_{cg}, should become as small as possible. Figure 6 depicts some typical applications for the conjunction of brake and steering systems in stabilizing a car.

In Case A, braking occurs on a road surface presenting unequal traction (split-μ). A great deal of yaw moment accumulates within a very short time. Without a regulating system, a driver will have difficulty handling this situation. Today's stability control systems such as ESC weaken the destabilizing influence of yaw moment by slightly delaying the accumulation of braking force on the front axle. Moreover, the program prevents generation of yaw moment on the rear axle by using the lowest coefficient of friction to determine braking force on both wheels, something known as the "select-low" strategy. However, the strategies thus described cause a conflict of interest between maximum stability and minimum braking distance. One can reduce this conflict of interest greatly by coordinating braking and steering. Pointing the front wheels in the direction of a lower friction coefficient reduces the amount of yaw moment acting on the car. It is therefore not necessary to delay accumulation of braking forces on the front axle, nor is it

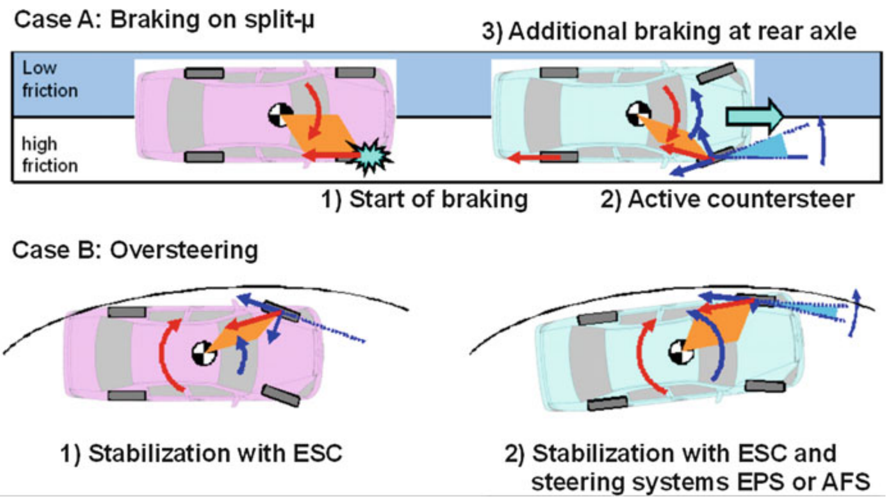


Fig. 6 The conjunction of brake and steering systems stabilizes a car in critical situations

necessary to follow the select-low strategy on the rear axle. This greatly reduces braking distance while providing good stability in a straight line.

Case B shows oversteer in a curve. Oversteering results in a reduction of stability because the vehicle yaws in the direction of the curve's outer diameter. There is then danger of skidding. In such a case, the ESC brakes the front wheel at the outside of the curve to produce a stabilizing yaw or to weaken the destabilizing yaw. The steering system can add greatly to stabilizing yaw as well. This occurs by steering back, i.e., by diminishing the steering angle at the front axle. This results in a larger angle between the position vector to the center of gravity and the resultant force, thus stabilizing the yaw tendency of the left front wheel.

Figure 7 depicts the potential for generating yaw moment that ESC with the brakes, the AFS (active front steering) with angle overlap on the front axle, and the ARK (active rear axle kinematics) with angle overlap at the rear axle have (Schiebahn et al. 2007). ESC has the greatest potential to stabilize an oversteering car in borderline friction limit situations. Steering systems allow one to diminish lateral forces very effectively in borderline situations. With AFS, this leads to a high degree of yaw moment that turns outward, while with ARK, it leads to an even higher degree of yaw moment that turns inward.

4 Function Modules for Steering-Angle Interventions

Figure 8 depicts the typical function modules for steering-angle intervention. The steering angle of the wheels δ is a product of the steering angle desired by a driver δ_{FW} and the overlapping angles δ_{FB} from the yaw controller and δ_{FF} from yaw torque compensation. The system utilizes driver input, consisting of the steering

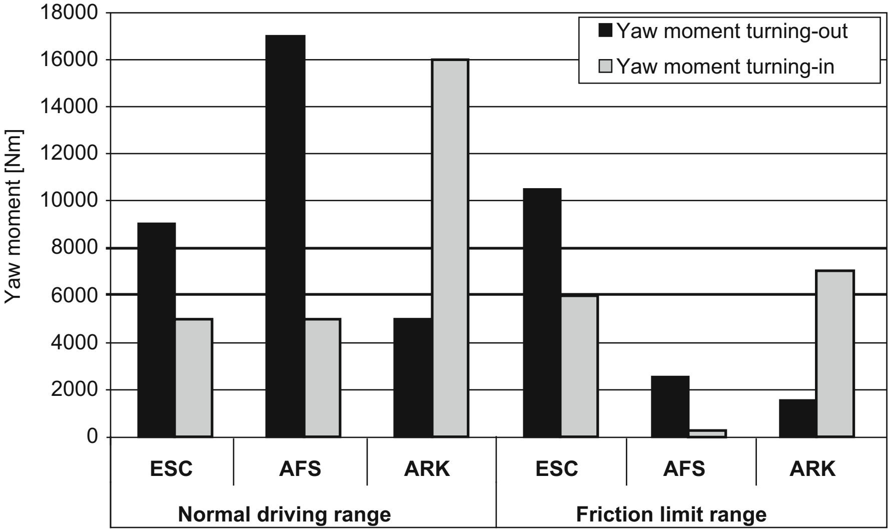


Fig. 7 Potential of braking and steering systems to generate additional yaw moment to force a vehicle in or out a curve when driving in steady state with constant radius. The friction limit range is defined by maximum lateral acceleration

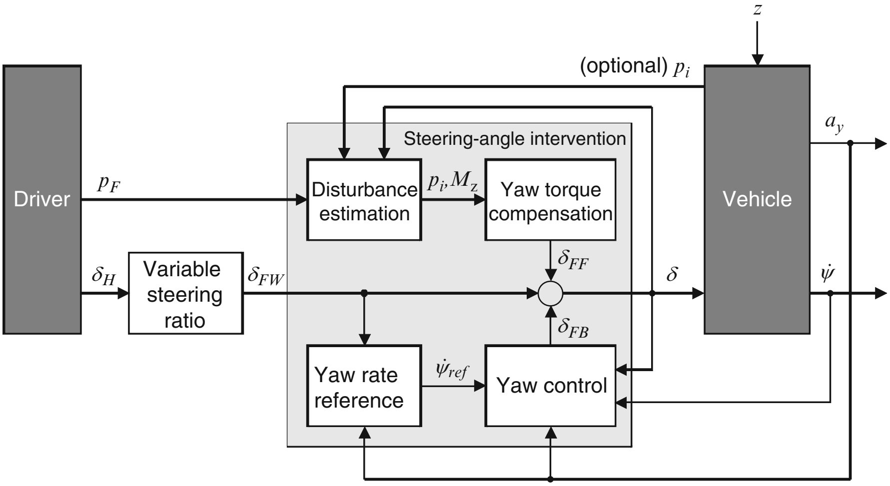


Fig. 8 Steering-angle intervention with yaw control and disturbance compensation

wheel angle δ_H and the pressure that the driver applies to the brakes p_F. The car measures its yaw rate ˙and lateral acceleration a_y and conveys this information to the controller. Not shown is the use of the car's speed as determined by the wheel-speed indicators. The reference yaw rate takes into account both the stationary and the dynamic behavior of the car and must be limited by a physically relevant measure, determined by the maximum frictional coefficient. The yaw control follows up on the yaw rate so as to support a driver and limit the sideslip angle and/or the sideslip rate in order to improve a car's stability. Yaw torque compensation is a disturbance feed-forward that compensates for the negative effects of disturbance variables z on handling during braking or acceleration. The estimated disturbance yaw moment M_z necessary for compensating for yaw torque is estimated from braking pressure or from the braking forces at each wheel. The function improves considerably if brake pressure p is measured at each wheel.

The added safety becomes evident primarily when braking on a road surface presenting uneven traction (split-μ) (Fig. 9). Since tires are able to transfer more braking force on road surfaces with more traction than on slippery surfaces, a car will want to turn in the direction of the portion of the road presenting the higher traction. ESC, which has been enhanced to include steering-angle intervention, counters this tendency with automatically dosed corrections in the opposite direction, thus freeing drivers of the necessity of stabilizing a car themselves. At the same time, ESC can set exactly the highest braking pressure at each wheel so that breaking distance shrinks with significantly better stability. All a driver needs to do in a stressful situation like this is to steer in the direction that he wants.

Yaw control improves a car's handling by intervening in steering in curves. For a brief time, it steers the front wheels somewhat faster than the movement of the steering wheel would indicate. The control causes a car to react quickly in emergency situations with better stability and less effort steering (Fig. 10). Countersteering occurs automatically and can begin quite early without a driver noticing it. As the sideslip angle increases, braking interventions increase in intensity.

5 Functional Modules for Driver Steering Recommendations (DSR)

If the steering system is designed to superimpose torque, intervention will be in the form of a Driver Steering Recommendation (DSR). If there is danger that a car will stray from the course set by a driver, the steering wheel will give off an unmistakable impulse indicating in which direction to turn the wheel. The functional modules are the same as for the angle overlap in Fig. 8. Only one module to transfer the ideal steering angle to the superimposed torque M_{DSR} needs to be added. The driver is now in a closed loop with the regulating system. Electric power steering can serve as an example for the chain of effects (Fig. 11): the driver applies steering torque M_F, the wheels apply self-aligning torque M_R, and the power steering applies assistance torque M_A. The torsion bar measures the reaction as hand torque M_H, which the power steering magnifies together with the superimposed torque. This

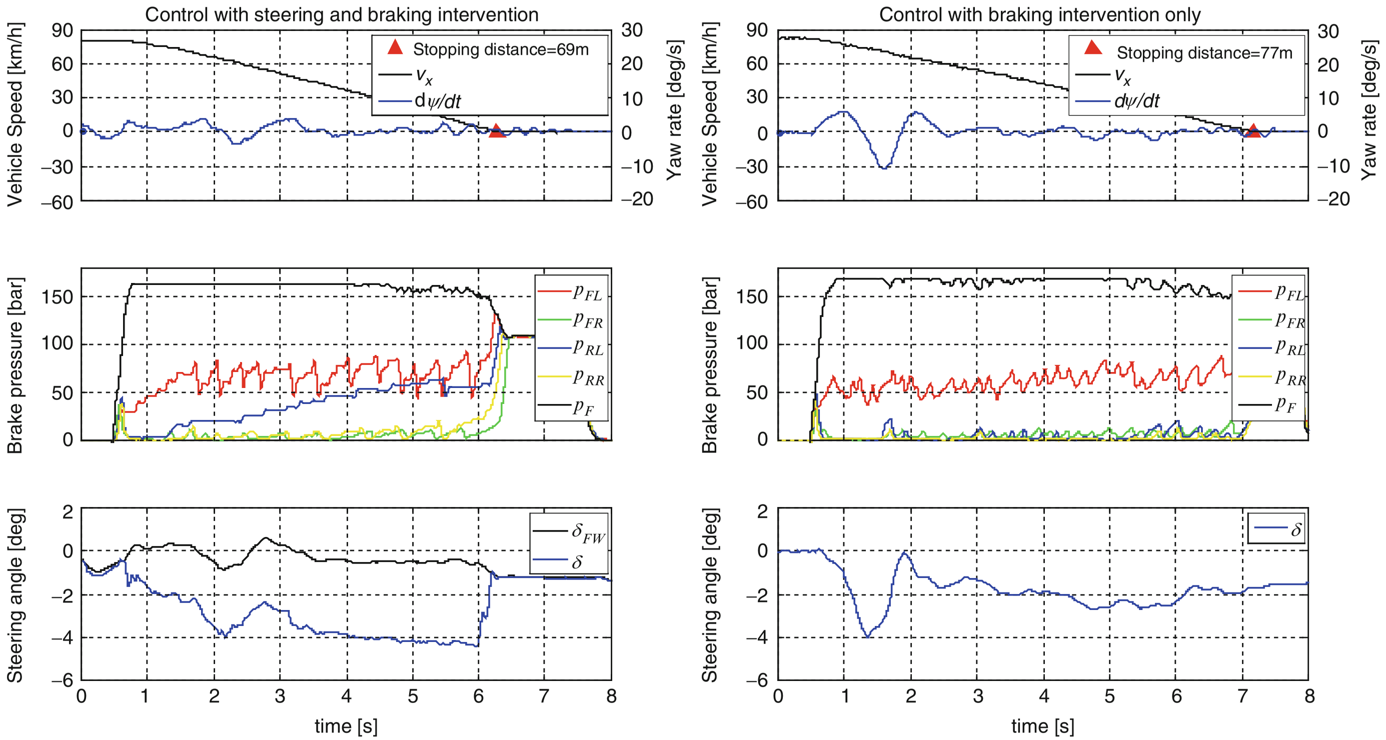


Fig. 9 Braking on split-μ with steering-angle intervention to compensate yaw torque and combined braking intervention at the rear axle compared to ESC without steering intervention

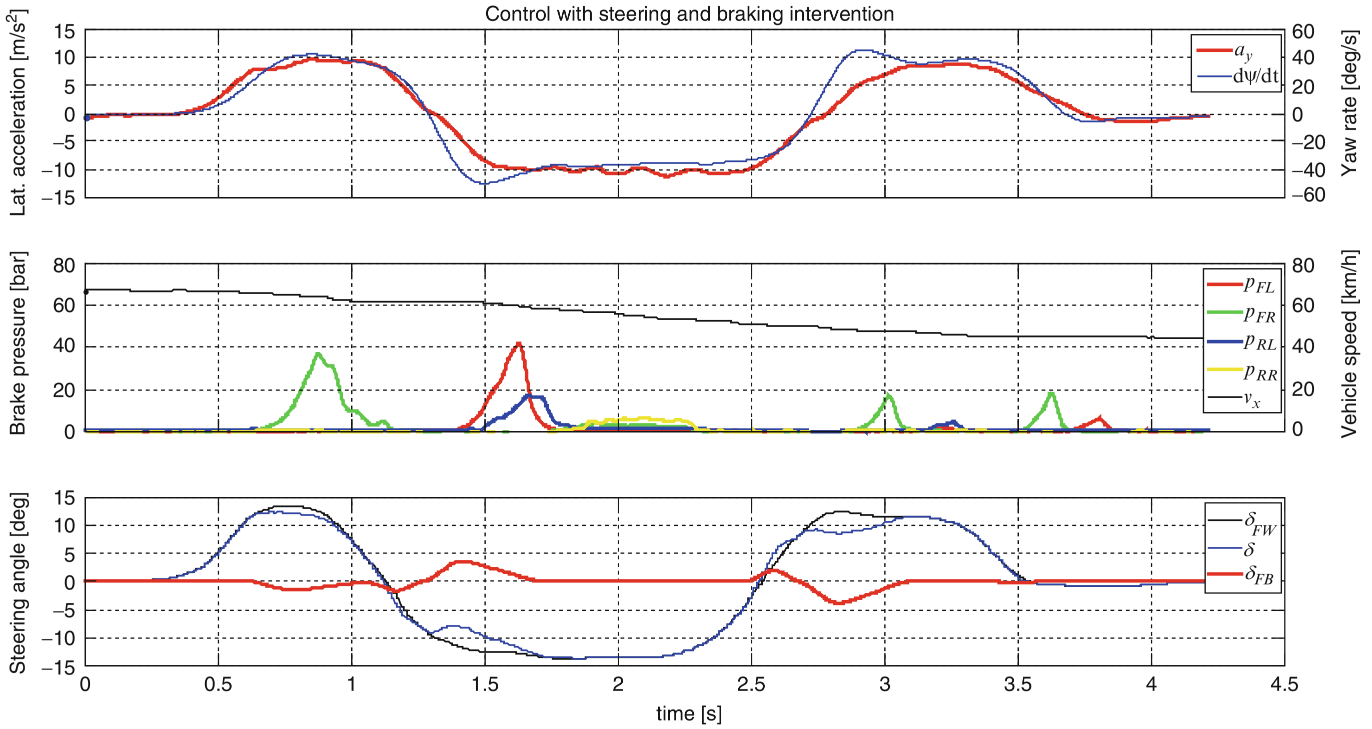


Fig. 10 VDA lane change with combined steering-angle and braking intervention to regulate yaw motion

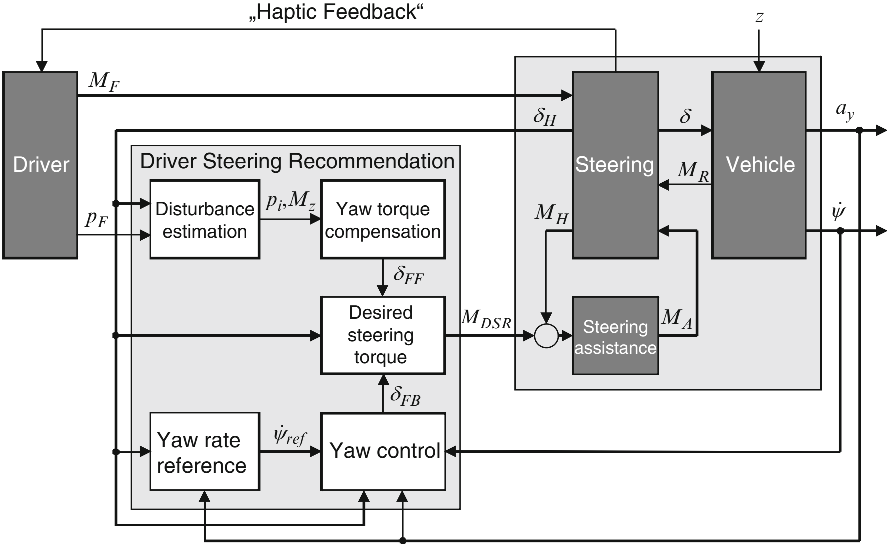


Fig. 11 Driver Steering Recommendation with torque overlap

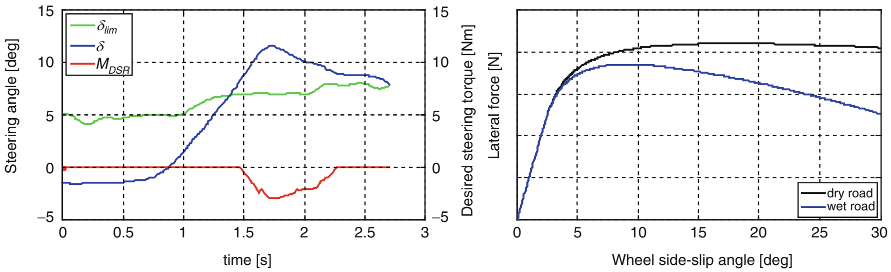


Fig. 12 Understeer situation with Driver Steering Recommendation and tire lateral force at different coefficients of friction and at higher speeds

causes the impulse in the steering system which helps drivers to react quickly and correctly in critical situations.

In oversteer and split-μ situations, superimposition of torque ensures that a driver can countersteer in stable fashion. In understeer situations in which a car tends to the outside via the front axle, a driver should not be able to steer beyond maximum lateral force so quickly. Most drivers react automatically to this situation by steering still further. Driver Steering Recommendation motivates a driver to stop turning the wheel and dial it back instead. If a steering-angle limit δ_{lim} is exceeded, it triggers superimposition of torque M_{DSR}. It relents only when a driver has restored steering wheel angle δ. This provides maximum lateral grip at the front axle for a given coefficient of friction (Fig. 12).

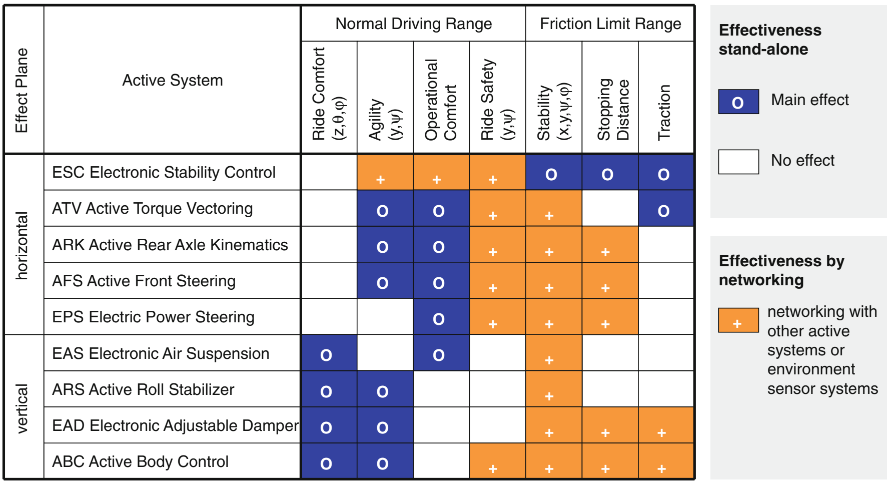
6 Specific Developmental Challenges and the Way Forward

Carmakers and suppliers agree that the systems that control vehicle dynamics will become increasingly linked. Such concepts as global chassis control (GCC) are opening new dimensions in driving dynamics, stability, and comfort by integrating active chassis systems (Fig. 13). The goal is to optimize the potential of each individual system and integrate it into an intelligent overall system (Schuster et al. 2008; Lunkeit and Weichert 2013). AUTOSAR hard- and software will support functional integration (see ▶ Chap. 7, “AUTOSAR and Driver Assistance Systems”).

Linking systems that control vehicle dynamics is an ongoing project. Intensive effort is proceeding on the following challenges (Hartmann et al. 2009; Raste et al. 2010; Raste and Rieth 2014):

- Indentifying areas where it is possible and desirable to determine a car’s characteristics by control systems
- Assembling the best possible portfolio of active systems for a given car or family of cars
- Designing chassis control functions for a given electronic architecture with the need to reign in complexity

The goal of a comprehensive coordinating concept for vehicle dynamics control that is common to all carmakers is still a long way off. Nevertheless, there is unanimity regarding the target. Chassis control should provide maximum comfort and enjoyment under normal circumstances. Carmakers have all the freedom they



Effect Plane	Active System	Normal Driving Range				Friction Limit Range		
		Ride Comfort (z,θ,φ)	Agility (y,ψ)	Operational Comfort	Ride Safety (y,ψ)	Stability (x,y,ψ,φ)	Stopping Distance	Traction
horizontal	ESC Electronic Stability Control		+	+	+	O	O	O
	ATV Active Torque Vectoring		O	O	+	+		O
	ARK Active Rear Axle Kinematics		O	O	+	+	+	
	AFS Active Front Steering		O	O	+	+	+	
	EPS Electric Power Steering			O	+	+	+	
vertical	EAS Electronic Air Suspension	O		O		+		
	ARS Active Roll Stabilizer	O	O			+		
	EAD Electronic Adjustable Damper	O	O			+	+	+
	ABC Active Body Control	O	O		+	+	+	+

Effectiveness stand-alone

O Main effect

☐ No effect

Effectiveness by networking

\+ networking with other active systems or environment sensor systems

Fig. 13 Potential of active chassis systems. Linking them increases their effect want to create an individual car's character. In borderline situations at the friction limit, every available actuator will go into action. An active chassis will help a driver to avoid an accident.

References

Hartmann B, Eckert A, Rieth PE (2009) Emergency steer assist – assistenzsystem für Ausweichmanöver in notsituationen. In: Proceedings of VDI-Tagung Reifen-Fahrwerk-Fahrbahn, VDI-Berichte No 2086, pp 131–148

Lunkeit D, Weichert J (2013) Performance-oriented realization of a rear wheel steering system for the Porsche 911 Turbo. In Proceedings of TÜV-Congress chassis tech, Munich, 13–14 June

Raste T, Kretschmann M, Lauer P, Eckert A, Rieth PE, Fiedler J, Kranz T (2010) Sideslip angle based vehicle dynamics control system to improve active safety. In: Proceedings of FISITA world automotive congress, Budapest

Raste T, Rieth PE (2014) The potential for handling improvements by global chassis control. In: Crolla D (ed) Encyclopedia of automotive engineering. Online, Wiley

Salfeld M, Stabrey S, Trächtler A (2007) Analysis of the vehicle dynamics and yaw moment maximization in skid maneuvers. In: Proceedings of TÜV-congress chassis tech, Munich, 1–2 Mar

Schiebahn M, Zegelaar PWA, Hofmann O (2007) Yaw torque control for vehicle dynamics systems. Theoretical generation of additional yaw torque. In: Proceedings of VDI-Tagung Reifen-Fahrwerk-Fahrbahn, VDI-Berichte No 2014, pp 101–119

Schuster M, Herold D, Wallbrecher M, Thalhammer T (2008) The new steering system of bmw – the integral active steering. synthesis of agility and sovereignty. In: Proceedings of FISITA World automotive congress, Munich

Brake-Based Stability Assistance Functions for Commercial Vehicles

42

Falk Hecker

Contents

Abstract

Vehicle stability functions like ABS, TCS, or ESP were firstly introduced on passenger vehicles. As the specific properties of heavy commercial vehicles lead to a very different behavior, the corresponding functions for commercial vehicles require at least significant modifications or in several cases completely new approaches to achieve the requested functionality.

F. Hecker (✉)
Knorr-Bremse AG, Schwieberdingen, Germany
e-mail: falk.hecker@knorr-bremse.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_43

With the stabilizing functions for commercial vehicles described in this chapter, at least a similar improvement of the vehicle safety could be achieved as for passenger cars. Due to the significantly higher inertia of heavy vehicles and the related more severe consequences in accidents, the lawmaker in Europe (and successively also in other parts of the world) decided to mandate the antilock braking (ABS) and the vehicle dynamics control (ESP).

This chapter focuses on the specifics of vehicle stability functions for commercial vehicles and their differences to similar functions for passenger vehicles.

Keywords

Electrically controlled steerings • Chassis • Steering • Heavy-duty vehicle design • Loaded-empty ratio • Trailer operation • Single-link trailers • Semi-trailers • Center-axle trailers • Multi-link trailers • Full-trailers • Fifth wheel coupling • ISO 11992 • Non-friction brakes • Retarders • Operating hours • SAE J1939 • Antilock braking system • ABS • Kamm circle • ABS control • Reference speed • μ-split • Yaw moment • Traction control system • TCS • TCS engine controller • TCS brake controller • Drag torque control • DTC • Conventional pneumatic service brake • ECU • Wheel speed sensors • Passive inductive wheel speed sensors • Tone wheel • Clamping sleeve • Air gap • Active wheel speed sensors • Pressure control valves • ABS valves • Electronically controlled service brake • Electronic braking system • EBS • Brake-control-by-wire • Brake CAN bus • Electronic brake-force distribution • EBD • Brake diagnostics • Differential lock management • Bus stop brake • Door brake • Steering brake • Off-road ABS • Trailer systems • ESP • Vehicle frame • Oversteering • Understeering • Jackknifing • Rollover • Yaw stabilization • Yaw-rate controller • Reference yaw rate • Effective steering ratio • Effective wheelbase • Self-steering gradient • Single track model • Measured yaw rate • Sideslip angle speed • Rollover stabilization • Rollover limit • Dynamic steering maneuvers • Step steer maneuver • Full- and multiple-trailer combinations • Full-trailer combination • Eurocombi • A-double combination • B-double combination • Road trains • System architecture • Steering wheel sensor • Yaw rate • Lateral acceleration • Optical sensors • Sensor CAN • Vehicle dynamics sensors • TRSP • Trailer roll stability program • All-wheel-driven vehicles • Specifics of ABS

1 Introduction

This chapter describes assistant functions for the stabilization of heavy commercial vehicles. The differentiation to the passenger vehicle is mainly made with regard to the braking system: In this chapter, commercial vehicles with pneumatically operated service brakes (power brakes) are discussed, as they are mainly used in medium- and heavy-duty vehicles (GVW above 6 t).

In the first part, wheel slip-based stabilization functions are explained, where the control loop is closed via the wheel speed information. The second part deals with the electronic stability program, which considers the entire vehicle dynamics in longitudinal and lateral dimensions by comparing the actual vehicle behavior with the driver's intended behavior. Finally, a short overview over further developments is given.

2 Specifics of ABS, TCS, and DTC for Commercial Vehicles in Comparison to Passenger Vehicles

2.1 Commercial Vehicle-Specific Features

With regard to the wheel slip-based stabilization functions, ABS (antilock braking system), TCS (antislip regulation), and DTC (engine drag torque control), the following distinctive differences with passenger vehicles apply (see also Hoepke and Breuer 2006):

- ***Chassis***: The typical heavy-duty vehicle chassis is based on a ladder-type frame construction with rigid axles. For cost reasons, a leaf spring construction is often used for the front axle. Besides the spring function, such a suspension also provides axle guidance both in longitudinal and lateral direction. On the other hand the tensioning of the leaf spring during braking maneuvers (s-shaped bending) creates some challenges for the wheel control functions.

  On the rear axles, normally an air suspension together with axle rods is used (improved suspension comfort and level adjustment after load change). The impact on the wheel control function here depends on the suspension kinematics and elasto-kinematics. Unfavorable kinematics (for instance, drawn axles) can, for example, lead to a "jumping" of the axle when braking is induced by the brake torque.

  Besides the standard chassis version (leaf spring on the front axle and air suspensions on the rear axles), there are further variants for other fields of application, e.g., construction vehicles for off-road use with leaf springs on all axles, special off-road vehicles with coil spring suspension, or vehicles with air suspension on all axles.
- ***Steering***: Commercial vehicles are mostly equipped with a servo-hydraulic recirculating-ball steering that transmits the driver demanded steering torque via a steering spindle amplified by a hydraulic piston and a related linkage to the wheels. Due to a normally positive steering roll radius, a noticeable feedback about the steering in case of different braking forces per side (e.g., during an ABS braking process) appears, which has a significant influence on the adjustment of the ABS system.

- ***Diversity of versions***: Heavy-duty vehicle design is based on a modular construction set that provides a huge flexibility to fulfill all customer needs. On top of the common modularity which is also standard in passenger vehicles (variations in gearboxes, engines, etc.), this also applies to the number and types of axles (two to five axles, optionally driven and/or steered and/or liftable), axle mounting, length of wheelbase (often selectable in 10 cm grid), size and stability of the ladder frame, steering gear type, etc. Furthermore, the OEM often delivers the "naked" vehicle to a body builder, who then completes it for the final purpose (e.g., dumper, platform body, loading crane, cement mixer, etc.). These modifications and attachments even increase the variety of physical driving properties.

  On certain markets (e.g., North America) the diversity of versions is even larger, as only the cabin and the frame are brand specific, and the customer can select the main technical components like motor, gearbox, and axles more or less freely from suppliers. For large fleets this has the advantage that they can use trucks from different manufacturers with nearly identical technology (i.e., motor, axles, and gearbox) which eases maintenance and repair for them.

  For wheel slip control functions like ABS or TCS, the diversity of versions means considerable requirements with regard to robustness, because during the development and application phase of the systems, only a very limited selection of vehicle versions can be tested. All other possible combinations must be covered by a robust system design.
- ***Wheel and axle loads, respectively***: Typical wheel and axle loads are substantially higher than those of passenger vehicles (up to factor of 15). This leads to much higher contact forces on the tire surface (for comparison, tire pressures in commercial vehicles are 6–8 bars, in passenger vehicles 1.5–3 bars). Together with a wear-optimized tire design, this effect leads to lower maximum friction values and thus to lower possible vehicle deceleration (maximum achievable deceleration in a commercial vehicle approx. 7–8 $\mathrm{m/s^2}$).
- ***Vehicle weight***: Commercial vehicles are designed for transporting passengers and goods in larger amounts and therefore for maximum load capacity. This leads to a much higher loaded-empty ratio k_{Load} than in the passenger vehicles:

$$k_{\mathrm{Load}} := \frac{m_{\mathrm{full}}}{m_{\mathrm{empty}}} \tag{1}$$

  For a mid-range passenger car with empty weights of 1,000–1,500 kg, k_{Load} values of 1.2–1.4 are usual. For load-dependent functions this means a variation of the weight of maximum ±16 %. For heavy-duty vehicles with empty weights of 6,500–9,000 kg and vehicle load capacities of 11,500–17,000 kg, the values are $k_{\mathrm{Load}} = 2.7$–2.9. This means a weight variation of up to ±50 %. With typical trailer operation or heavy haulage, these values increase further (up to $k_{\mathrm{Load}} = 15$).

  The high weight of commercial vehicles furthermore leads to higher inertia and thus to reduced vehicle dynamics. This also applies to the wheels that have a significantly higher inertia torque than passenger vehicles. Thus the required control frequency for a wheel slip control system is lower than in a passenger vehicle.

- ***Trailer operation***: Especially heavy commercial vehicles with a gross vehicle weight (GVW) of more than 11 t often pull one or more trailers. In this context we differentiate between the following types of trailers:
  - Single-link trailers (e.g., semi-trailers or center-axle trailers)
  - Multi-link trailers (e.g., full-trailers)

    Furthermore there are different types of trailer couplings:
  - Couplings via a trailer coupler (= ball joint) that can only transmit forces and no torques
  - Couplings via a fifth wheel coupling that (in a certain range) also transmits roll torques

  These properties lead to additional degrees of freedom, which have a major impact on the physical vehicle dynamics behavior. When it comes to wheel control systems, every vehicle segment is seen autonomously, which means that every vehicle segment is equipped with its own ABS. Only in case the towing vehicle and trailer are equipped with an electronic braking system (EBS), there is a communication interface between the vehicles (CAN bus according to ISO 11992) via which a limited amount of information can be exchanged. In context of the mandatory introduction of ABS (1991 in Europe), many tests have been made with regard to the dynamic behavior of different combinations (towing vehicle with/without ABS and trailer with/without ABS). The results of these tests show that ABS fundamentally improves vehicle stability, even if not all vehicle segments are equipped with it.
- ***Service brake***: The main feature is the pneumatically driven actuation of the service brake. Here the energy required for the brake actuation is provided in the form of compressed air and transformed into a clamping force on the brake pads of the disk or drum brake via brake cylinders using lever kinematics. The control of the braking pressure is either purely pneumatically (conventional braking system) or electronically (EBS, see Robert Bosch GmbH 2014).
- ***Non-friction brakes***: Besides the "normal" service and parking brakes, commercial vehicles often also have one or more non-friction brakes (so-called retarders) that work without any wear. These include the engine retarder that can be found in almost all commercial vehicles and in which the drag torque of the motor is increased by technical measures (exhaust cover flap, valve regulation, etc.). Furthermore, there are electrodynamic or hydrodynamic retarders. All retarders have in common that the braking force is initiated via the drivetrain and the driven wheels. We differentiate basically between primary retarders which are connected to the engine (before the clutch) and secondary retarders (connected after the clutch, often directly on the drive shaft). With low axle loads (empty vehicle) the activation of the retarder at low friction values leads to very high wheel slip values and thus to instability of the driven wheels, which has to be taken into account with regard to the ABS.
- ***Component requirements***: Normally commercial vehicles are designed for a service life of up to 1,500,000 km or up to 50,000 operating hours. This is

significantly higher than the service life of a passenger vehicle (factor of 3–5). Together with the much tougher environmental conditions, this leads to considerably higher requirements for commercial vehicle components.

- ***Vehicle communications architecture***: In most commercial vehicles an SAE J1939 normed CAN databus (see Hoepke and Breuer 2006) is used for the drivetrain communication. It connects the main power train control devices, like engine ECU, gearbox ECU, retarder ECU, and brake ECU. The communication is done via defined messages. In this way the integration of electronic systems is made significantly easier. Nevertheless there is a strong trend in the industry to more and more use a specific CAN communication at each OEM.

2.2 Control Targets and Priorities

2.2.1 Antilock Braking System ABS

Wheel Slip Control The ABS controls the wheel slip of the single wheels while braking. Wheel slip λ_w is defined as

$$\lambda_w = \frac{v_w - v_u}{\mathrm{MAX}(v_u, v_w)} \tag{2}$$

with

v_u as wheel rotation speed [m/s]
v_w as vehicle speed in the wheel contact point [m/s]

The wheel speeds are sent as measured actual values to the controller which compares them with the nominal value (reference speed of the vehicle) considering the target slip. Deviations are then corrected by means of brake pressure modulations. The target slip is automatically adapted during an ABS operation, with the aim of finding the best possible compromise between stability and deceleration/traction. Due to the tire characteristics, the achievable cornering force decreases drastically with increasing longitudinal slip (simplified, shown in the Kamm circle or in the traction-slip diagram in Fig. 1). Reasonable compromises for optimal target slips are in the range of $\lambda_w \approx 8$–$20\ \%$.

In commercial vehicle ABS, not only classic PID controllers but also matrix controllers are used which automatically adapt to different friction value curves. Figure 2 displays a typical ABS control cycle.

In addition to the intervention in the service brake, the available retarders are switched off by ABS when the related wheels start locking.

The vehicle reference speed that is necessary for the regulation is determined from the single wheel speeds. Special algorithms and plausibility checks must ensure under all circumstances that the vehicle reference speed is properly in line with the actual vehicle speed. Critical cases are, for example:

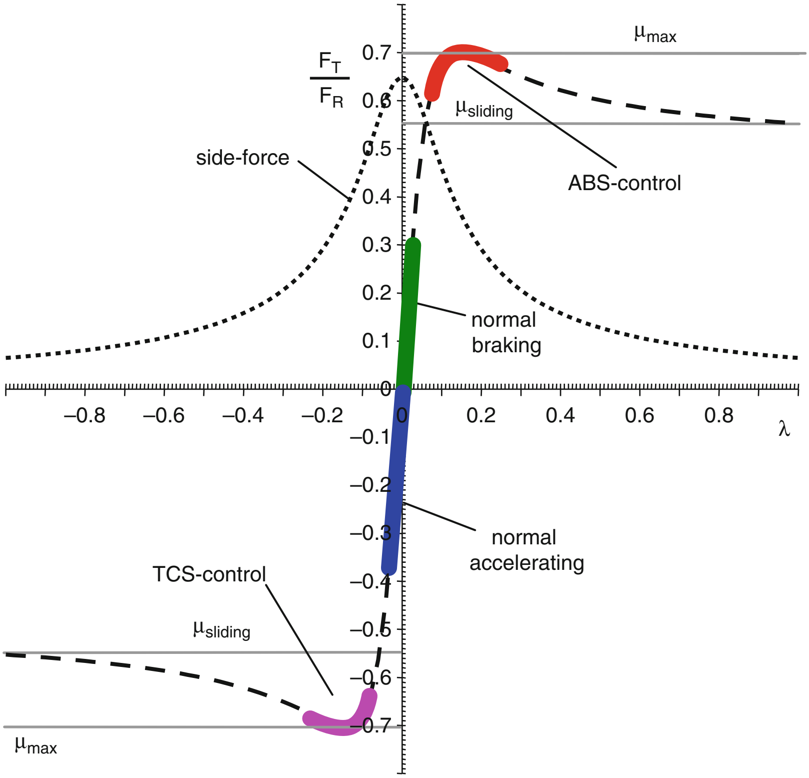
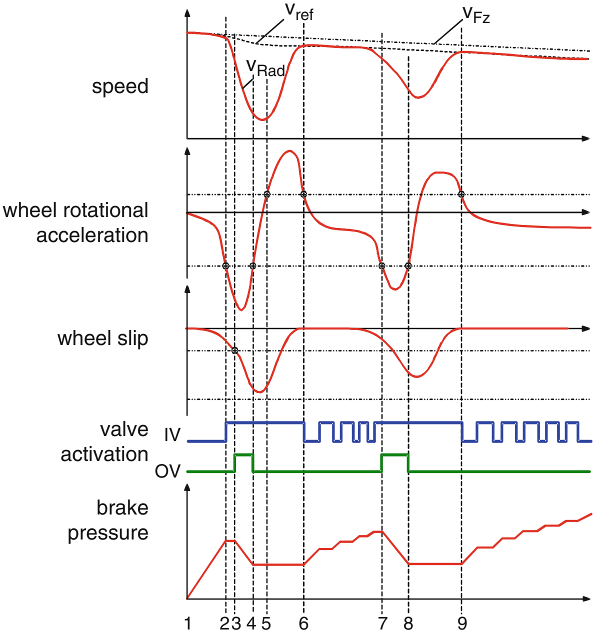


Fig. 1 Diagram of tire force versus wheel slip with the operating areas of the different stabilization functions

- The potential "drag-down" of the vehicle reference speed if all wheels reach a high slip at the same time. The consequence is overbraking of the wheels, which leads to a loss of steerability. As a countermeasure, the ABS algorithm underbrakes single wheels at certain points in time, which then accelerate to vehicle speed and thus support the determination of the reference speed. These time spans are so short that the impact on the brake path is negligible.
- The "speedup" of the vehicle reference speed if a wheel speed signal is disturbed, for example, by poor tone wheels or electromagnetic interference and thus fakes a higher wheel speed. As a consequence this high vehicle reference speed would lead to an under-braked vehicle. To avoid this, the ABS algorithm incorporates various plausibilization algorithms using all wheel speeds.

Strategies for Vehicle Stabilization Due to the fact that the ABS controller has no information about the actual vehicle dynamics (other than wheel speeds), it contains sophisticated strategies for securing vehicle stability (e.g., when braking on a one-sided slippery road, called μ-split). Due to this lack of information, the achieved degree of stabilization depends on the adjustment of ABS, which depends

Fig. 2 Typical operation in ABS control: *1* braking start, pressure buildup; *2* wheel starts locking, maintain pressure; *3* increased locking tendency (wheel unstable), pressure decrease; *4* wheel stabilizes, maintain pressure; *5* wheel accelerates, maintain pressure; *6* wheel stable, pressure buildup (pulsed); *7* wheel starts locking again, pressure decrease; *8* wheel stabilizes, maintain pressure; *9* wheel stable, pressure buildup (pulsed)



on the vehicle geometry and is performed during the application phase. Vehicles with a short wheelbase and low rear axle load react very sensitively when braking on µ-split (e.g., empty semi-trailer tractors). This will be explained in the following in a simplified way (rear axle unbraked). The braking force difference on the front axle induces the yaw moment:

$$M_{zB} = \Delta p_{fd} \cdot k_{FB} \cdot \frac{b_{fa}}{2} \tag{3}$$

with

Δp_{fd} as difference braking pressure on the front axle [bar]
k_{FB} as braking factor [N/bar]
b_{fa} as effective track width of the front axle [m]

The yaw moment, which is generated from the single wheels on the rear axle via drift angles, works against that.

$$M_{zra} = \left(F_{yrl} + F_{yrr}\right) \cdot l_H \tag{4}$$

with

$F_{yrl,r}$ as lateral force (on the left or right rear wheel) [N], which depends on the drift angle and the drift rigidity: $F_{yrl,r} = \alpha_{l,r} \cdot C_{l,r}$
l_H as distance between rear axle and point of gravity [m]

Due to the fact that both moments must be in balance $(M_{zra} = M_{zB})$ so that the vehicle does not spin, the following relation results between the forces and the geometric data:

$$\frac{F_{yrl} + F_{yrr}}{\Delta p_{fa} \cdot k_{FB}} = \frac{b_{fa}}{2 \cdot l_H} \tag{5}$$

If we assume the lateral forces $F_{yrl,r}$ and the track width b_{fa} as constant, we see that the permissible difference braking pressure is proportional to the distance between the center of gravity and rear axle and thus to the wheelbase. Simply spoken, the shorter the wheelbase, the smaller the permissible difference braking pressure. De facto the dependence is nonlinear due to the neglected effects and is especially amplified by the dynamic axis load transfer during braking which increases with shorter wheelbases.

2.2.2 Traction Control System (TCS)

The TCS is active during acceleration with two basic targets: increase driving stability and improve traction by utilization of the maximum possible friction values at all driven wheels. The so-called TCS engine controller serves to increase stability by limiting the engine torque in a way that a given target slip on the driven wheels is not exceeded.

The target slip is – similar to ABS – selected as the best possible compromise between traction and stability, while the focus is more on traction in commercial vehicles. For some systems the target slip is adapted dynamically depending on the accelerator pedal position or in cornering situations. Thus we achieve an optimized traction in straightforward driving with higher slip and at the same time maximum stability in cornering situations with reduced slip.

During acceleration, one-sided driven wheel spinning occurs especially with different friction values, due to the fact that the differential gear distributes the drive torque at a ratio of 50/50 on both sides, and thus the side with the lowest friction value limits the maximum transmittable drive torque. In this case the so-called TCS brake controller intervenes by limiting the wheel slip on the spinning wheel by active braking. The brake torque created by this is mirrored through the differential gear to the other side, increasing the overall drive torque. By this so-called electronic diff lock function, a similar traction force can be generated as with a mechanical diff lock (in an ideal case). The disadvantage is that a certain amount of engine power is consumed in the brake.

In commercial vehicles with two driven rear axles (e.g., three-axle vehicles with the wheel formula 6 × 4), a central differential gear distributes the drive torque onto both drive axles. In this case the TCS braking controller acts on all four drive wheels and thus controls the slip of all drive wheels with the target to maximize traction.

In contrast to the ABS, the generation of the reference speed is relatively easy in the TCS, because the non-driven wheels (front wheels) deliver a very good representation of the vehicle speed. A TCS for vehicles with driven front wheels requires additional sensors, such as a longitudinal acceleration sensor, for supporting the reference speed.

2.2.3 Drag Torque Control DTC

Especially empty commercial vehicles with low rear axle load can only exert very low forces on the drive wheels on slippery ground. Therefore in overrun operation (release of accelerator pedal), a large wheel slip occurs due to the relatively high drag torque. This reduces vehicle stability considerably. The effect is even bigger when gearing down, because the drag torque changes immediately due to the gear switch process. High friction forces in the power train, for example, at very low temperatures, further increase this effect.

The DTC detects an increased wheel slip due to the drag torque on the driven wheels and actively increases the engine torque with the aim of reducing the wheel slip and thus of stabilizing the vehicle. Basically, the same slip control cycle is used as in the TCS, the difference is that the target slip is positive this time.

2.3 System Setup

2.3.1 Conventional Pneumatic Service Brake

Figure 3 shows a conventional pneumatic braking system with ABS/TCS (see Breuer and Bill 2006).

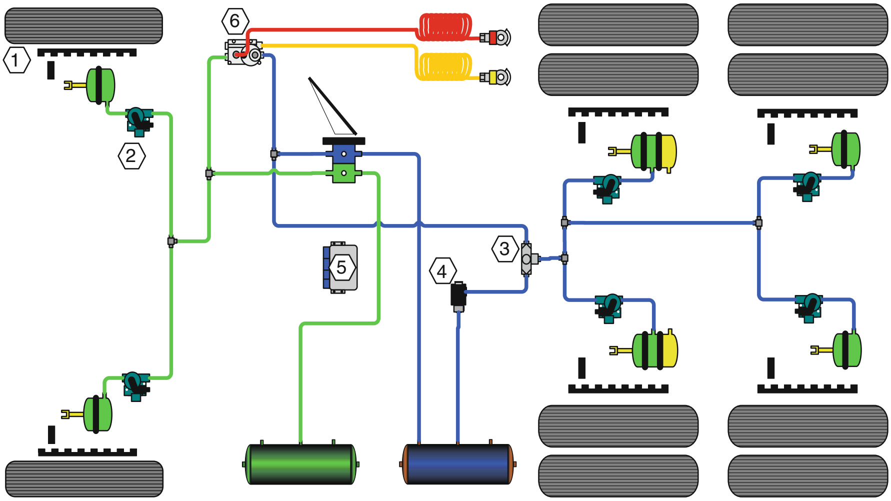


Fig. 3 System setup of a conventional pneumatic braking system with ABS and TCS: *1* wheel speed sensor, *2* ABS valve, *3* relay valve, *4* TCS valve, *5* electronic control unit, *6* trailer control valve

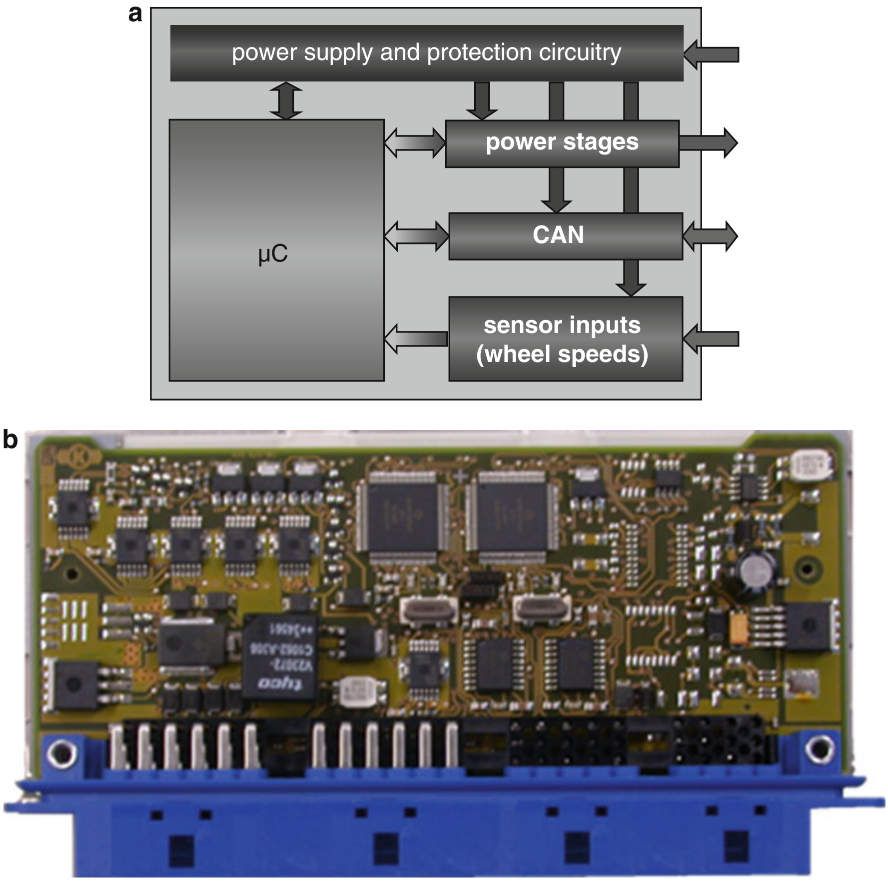


Fig. 4 (**a**) Block wiring diagram ABS control device (**b**) ABS control devices in open state (Source: Knorr-Bremse)

Electronic Control Unit (ECU) The ABS and TCS algorithms run in a microcontroller, which, together with the power stages for the control of the ABS and TCS valves, the power supply, and other peripheral components, is integrated in the electronic control device. The block diagram in Fig. 4a highlights the internal setup of the control device which is realized in conventional SMD technology. For µ-controllers, generally 16-bit controllers with clock frequencies of 20–40 MHz are used which are supplemented by a monitoring µC. The memory requirements are at 128–512 kB ROM (mostly flash technology) and 4–12 kB RAM. An EEPROM which is mostly integrated in the µC serves for the application parameters and for storing learned values and errors (see Robert Bosch GmbH 2014). For communication with external systems (engine and retarder), the ECU is connected to the vehicle data network (mostly CAN bus according to SAE J1939). Via this bus the control of the engine, or retarder torque, is communicated, for example. On the other hand, the databus provides information such as current engine and retarder torque, engine speed, accelerator pedal position, etc.

Wheel Speed Sensors Almost only passive inductive wheel speed sensors are used in heavy commercial vehicles. By turning the tone wheel (60–120 teeth), a current is induced in the sensor with a frequency that is proportional to the wheel speed and which is processed by the ABS ECU. The sensors are inserted in a holder by means of a clamping sleeve (force fit attachment). By doing so it is ensured that the sensors are "pushed back" by an asymmetric tone wheel: The air gap is set automatically. This generally robust construction can lead to a widening of the air gap in case of strong oscillations or contamination so that the wheel speed information is no longer sufficient: the so-called air gap limit, meaning the speed from which an evaluable AC current is induced, increases. By simple "pushing in" of the sensor, this problem can be solved.

Active wheel speed sensors, which are common in the passenger vehicles, are not yet established in commercial vehicles. This is due to the high number of axle versions in commercial vehicles in connection with the decisively longer and unsynchronised development cycles and the low technical and commercial advantages of active sensors.

Actuator In a commercial vehicle ABS, so-called pressure control valves (ABS valves) are used as actuators. These are functionally designed as 3/3 valve with two magnets and serve the purpose of maintaining the braking pressure in case of an increased wheel slip (e.g., to prevent further pressure buildup) or to reduce the braking pressure. In Fig. 5 the internal setup of the ABS valve is displayed.

Due to the fact that a direct control of large valve cross sections requires very large magnets to provide the large forces, the actual valve function is carried out by two elastomer diaphragms which are pre-controlled by relatively compact magnetic valves. This is displayed in Fig. 6 for different operating conditions (unbraked, braked, and during ABS control).

Thus the braking pressure can be lowered or limited by means of the ABS valves. For the TCS brake controller, a braking pressure must be built up actively in order to brake single wheels without the intervention of the driver. For this a separate ASR valve (3/2 valve) supplies the reservoir pressure for the ABS valves which then perform the pressure modulation for the TCS brake control function (see Fig. 6).

2.3.2 Electronically Controlled Service Brake (Electronic Braking System (EBS))

In the case of the electronically controlled service brake (electronic braking system (EBS)), the ECU measures the driver brake demand via a brake pedal stroke sensor and determines the axle individual braking pressures considering the actual conditions (driving situation, loading condition, etc.). These brake pressure demands are then electronically controlled in electro-pneumatic modulators (EPM) individually for every wheel (see Robert Bosch GmbH 2014). Because it is a basic principle

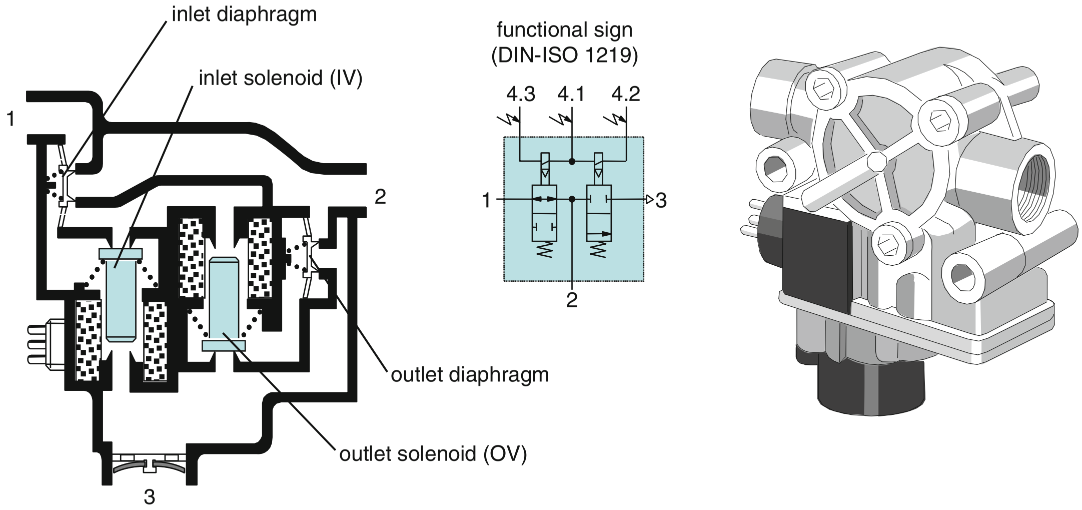


Fig. 5 ABS pressure control valve (fittings: *1* foot brake valve, *2* brake cylinder, *3* exhaust)

Operating condition

Unbraked
Connections 1 and 2 are depressurized. Inlet and outlet diaphragm are closed. The two magnets (I, II) are not activated.

Braking without ABS intervention
The active braking pressure on Port 1 opens the inlet diaphragm. Via the upper valve seat of II, brake-pressure gets into the compartment (b). Thus the outlet remains closed and Port 2 is pressurized.

ABS intervention Maintaining braking pressure
By activating solenoid I, the lower valve seat closes and at the same time the upper one opens. Thus compartment (a) is aerated and the inlet diaphragm closes. The outlet also remains closed due to the pressure in compartment (b). Thus the pressure on Port 2 remains constant.

ABS intervention Lowering braking pressure
Solenoid II closes the upper valve seat and at the same time opens the lower one. Compartment (b) is vented. Due to the brake-cylinder pressure, the outlet diaphragm opens by which the braking pressure is locked using air vent 3.

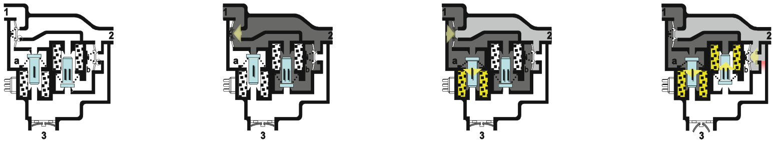


Fig. 6 Function of the pressure control valve

"brake-control-by-wire," all technical preconditions for the autonomous braking pressure modulation are already given. The ABS, TCS, and DTC algorithms are integrated in the EBS central ECU, as all other brake control-related functions, and send the calculated braking pressure demand via the brake CAN bus to the EPMs.

The wheel speed sensors are identical to those described in Sect. 2.3.1.

2.4 Special Functions in Commercial Vehicles

2.4.1 Towing Vehicle Systems

On top of ABS and TCS, a number of add-on functions (value-added functions) exist in the infrastructure of the ABS/TCS or EBS systems. These include:

- ***Electronic brake-force distribution (EBD)***: The braking pressure on the rear axle is reduced by using the ABS valves depending on the wheel slip difference between the front and rear axle in order to adapt the brake force of the rear axle to the loading condition. Thereby the ALB valve (load-sensing valve) that was necessary previously is replaced functionally.
- ***Brake diagnostics***: By means of long-term comparisons of the single wheel slip values during braking, the system detects malfunctions of single brakes or varying braking behavior.
- ***Differential lock management***: The function supports the driver when engaging the differential locks to protect the mechanics. This is realized by active synchronization of the wheel speeds via the brakes and controlling the engagement of the diff locks.

- ***Bus stop brake (door brake)***: Automatic activation of the service brake in buses when the bus stops and the doors are opened and automatic deactivation when closing the doors and drive off. Thus the bus driver does not have to engage the parking brake at a bus stop.
- ***Steering brake***: Especially vehicles with the axle formula 6 × 4 (three axles with two driven rear axles) tend to understeer on slippery ground (e.g., in roadwork operation) in sharp curves. In this case the function brakes the rear wheels single-sidedly depending on the steering wheel angle, in order to support the cornering willingness. This results in a reduced turning circle by the additional yaw moment.
- ***Off-road ABS***: Especially for military vehicles and other vehicles that are primarily operated on non-asphalted surfaces, there are modified ABS algorithms that increase the wheel slip drastically especially at low speeds. Thus the braking force is increased via the created brake wedge on loose ground (e.g., gravel or loose snow).

2.4.2 Trailer Systems

There is also a number of additional functions with regard to the trailer that use the infrastructure of the EBS (or ABS) of the trailer. Among these are functions that control various things at the trailer depending on speed and/or load (e.g., resetting the level into driving position, controlling a lift axle, etc.).

3 Specifics of ESP (Electronic Stability Program) for Commercial Vehicles

3.1 Commercial Vehicle-Specific Characteristics

Basically the vehicle dynamics control ESP builds upon the ABS/TCS or EBS in a modular way and uses the already available infrastructure with regard to components as well as to functions.

On top of the features and characteristics described in Sect. 2.1, certain other characteristics become important when it comes to vehicle dynamics control:

- ***Height of the center of gravity***: Commercial vehicles have a total height of up to 4 m (in some countries up to 4.5 m) which lead to center of gravity heights of 1.2–2.5 m in combination with the payload. Therefore heavy commercial vehicles tend to roll over much earlier than passenger vehicles – often already in quasi-stationary maneuvers. Typical lateral acceleration values that lead to rollover are in the range of 4–6 m/s^2.
- ***Flexible vehicle frame***: Commercial vehicle frames are very flexible due to their design (open U profile). The behavior in curves is thus very complex and cannot be described by a simple rigid body model. The frame stores the roll energy partly due to its twisting in curve driving and releases this energy, for example, in S-shaped curves with the consequence that the rollover tendency increases

further. In case of flexible load, this behavior gets even more complex (e.g., liquid tank).
- ***Degrees of freedom***: As already described in Sect. 2.1, the number of degrees of freedom increases due to trailer operation. Especially with regard to the vehicle dynamics control, this has a decisive impact on the control strategy to be selected.
- ***Uncertainties in the steering system***: The steering angle sensor, which is necessary for the vehicle dynamics control, is found in the steering column as in passenger vehicles. Due to the large adjustment range of a commercial vehicle steering column, which is realized using a cardan joint, relatively large irregularities in the measured steering wheel angle signal occur, which have to be tolerated by a robust system design.

3.2 Control Targets and Priorities

Heavy commercial vehicles can additionally to the well-known spinning and sliding in passenger vehicles (oversteering or understeering) assume other unstable states. These include:

- Jackknifing of multi-segment vehicle combinations, for example, triggered by pushing trailers
- Rollover due to high lateral acceleration

Thus, the vehicle dynamics control in commercial vehicles has to address the jackknifing and rolling-over in addition to the stabilization functions known from the passenger vehicles.

The vehicle dynamics control systems that are nowadays available are designed for the use in almost all single vehicles and in almost all various vehicle combinations with one or more articulation joints (e.g., articulated vehicles, articulated trains, or Eurocombis).

3.2.1 Yaw Stabilization

The yaw stabilization is based on a yaw-rate controller which compares the measured vehicle yaw rate with the one desired by the driver (reference yaw rate) and which compensates deviations using braking and engine torque interventions.

The reference yaw rate is defined by the system using a simple physical model that was deducted from the plane vehicle dynamics equations for vehicle combinations with articulation joints (single track model, see Hecker et al. 1997):

$$\dot{\psi}_Z = \frac{\delta_h}{i_L} \cdot \frac{v_{\mathrm{cog}}}{l + EG \cdot v_{\mathrm{cog}}^2} \tag{6}$$

with

$\dot{\psi}_Z$ as reference yaw rate [rad/s]
δ_h as steering wheel angle [rad]
i_L as effective steering ratio [-]
v_w as vehicle speed in the wheel contact point [m/s]
l as effective wheelbase [m]
EG as self-steering gradient [s^2/m], which describes the self-steering behavior of the vehicle combination.

The parameters occurring in the models are either programmed at the end of the line (e.g.,, wheelbase) or adjusted online using special adaptation algorithms (parametric rating) to the specific behavior of the vehicle (e.g., self-steering gradient).

Although the model was deducted for a vehicle combination with an articulation joint, it corresponds in its structure to the single track model for single vehicles (see Zomotor 1991). The impact of the attached trailer is included in the self-steering gradient. The model structure also applies to vehicles with more than two axles. In this case the adaptation is made by the effective wheelbase which includes the effects, for example, of a double axle unit (see Winkler 1996).

A massive deviation of the measured yaw rate from the reference yaw rate leads to a control error, which is transformed by the actual controller to a corrective nominal yaw moment with due regard to the physical limits. The physical limits represent the yaw values currently possible under the given friction conditions and are determined via a friction value estimation. Due to the fact that only the utilized friction value can be estimated and thus a certain safety margin becomes necessary, this leads as a consequence to a limitation of the sideslip angle speed to an extent manageable by the driver.

The level of the nominal yaw moment depends not only on the control error but also on the prevailing vehicle configuration (wheelbase, number of axles, operation with or without trailer, etc.) and the loading conditions (weight, position of center of gravity in longitudinal direction, inertia torque around the vertical axle, etc.). Due to the fact that these parameters are variable, they have to be calculated by the ESP continuously. This happens, for example, in the case of the loading conditions by an estimation algorithm which permanently identifies the current vehicle weight using signals from the engine control (engine speed and torque) and vehicle longitudinal movement (based on wheel speeds).

In order to transform the nominal yaw moment into a stabilization intervention, the ESP does a rough classification of the driving situation in “oversteering” and “understeering”:

- Oversteering describes situations in which the rear of the vehicle moves sidewards, i.e., the vehicles turn faster than necessary for the desired curve radius. This situation can lead to jackknifing in articulated vehicles and is difficult to manage by the driver.
- In the case of understeering, the vehicle pushes the front wheels to the outside of the curve rim (comparable to a front-driven passenger vehicle on slippery

ground), which especially applies to vehicles with two rear axles (double axle unit).

Additionally the system includes the estimated articulation angle into the evaluation of the driving situation.

Depending on the evaluated driving situation and the calculated yaw moment, the brake interventions are implemented on selected wheels appropriately. Wheels are prioritized in which the braking force buildup and the thereby induced side force loss create an aligned yaw moment (see Fig. 7). The stabilization effect is supported by targeted modification of the ABS target slip values which especially come into effect in the braked driving status.

In addition to the wheel-individual braking interventions on the motor vehicle, the trailer is also braked in certain situations. Here individual wheelbraking intervention is not possible due to technical reasons, i.e., the trailer is only braked as a whole.

In Fig. 7, the stabilization interventions for definite over- and understeering are displayed as an example. In addition to these clear situations, there are further critical vehicle conditions in which also other wheels or combinations of wheels are braked according to the nominal yaw moment.

3.2.2 Rollover Stabilization

Due to the often high position of the center of gravity of a commercial vehicle, the sliding and jackknifing occurs mainly with low and medium friction values. In the case of high friction values, the rollover tendency is more distinctive. The rollover limit does not depend on the height of the center of gravity, but on the chassis (axle mounting, anti-roll bars, spring basis, roll center, etc.) and the kind of loading (solid or flexible loading). An approximative calculation of the rollover limit is displayed in Robert Bosch GmbH (2014).

When observing the actual rollover process in quasi-stationary circle driving, the basic reason for rollover is a lateral acceleration that is too high which is caused by too high vehicle speed at the given curve radius.

The ESP uses the physical coherences to reduce the rollover risk: As soon the vehicle gets closer to the rollover limit, it is decelerated first by a reduction of the engine torque and if necessary by additional braking (see Fig. 8). The rollover limit is calculated in the ESP depending on the loading status of the vehicle and the load distribution, whereby the loading status of the vehicle is continuously identified.

Dynamic steering maneuvers often lead to stronger roll movements and thus strengthen the tendency to rollover. Examples are the overshooting in a step steer maneuver or the transmission of roll energy in S-shaped curves (roundabouts and evasion maneuvers). Therefore the determined rollover limit is modified depending on the specific driving situation. This ensures, for example, a reduction of the rollover limit with the target of early intervention in fast and dynamic driving situations (evasion maneuvers).

In contrast to that, the rollover tendency is smaller in very slow driving maneuvers (e.g., narrow serpentine curves uphill) which is why the system

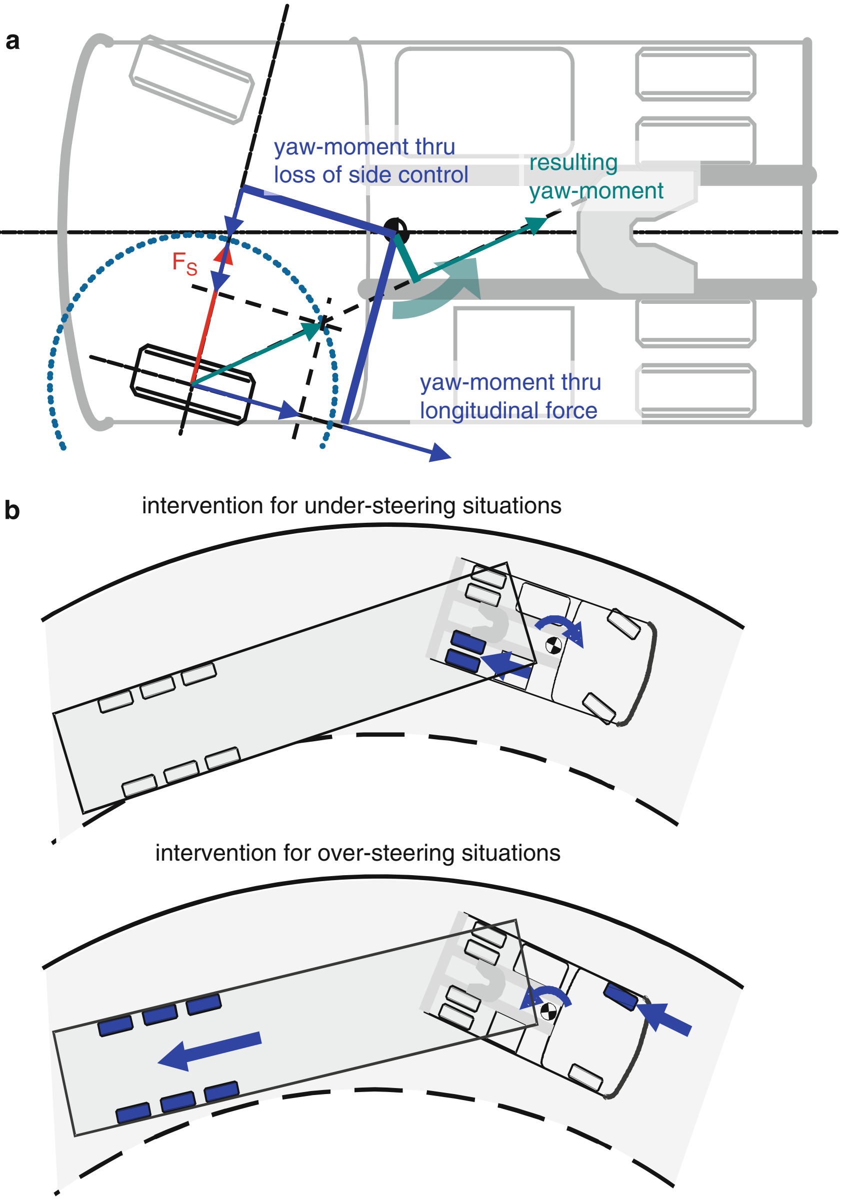


Fig. 7 Impact of a braking intervention on a wheel on the yaw moment (*left*) and intervention strategy of the vehicle dynamics control (*right*)

increases the rollover limit to prevent unnecessary and disturbing braking interventions.

The bases for the determined rollover limit are certain assumptions with regard to the height of the center of gravity and the driving behavior of the vehicle

Fig. 8 Cornering at 60 km/h without/with ESP at high friction value with fully loaded semi-trailer combination (Source: Knorr-Bremse)

combination at known axle load distribution. Thus the ESP covers the major part of existing vehicle combinations. In order to guaranty stabilization also in cases of large deviations of these exceptions (e.g., extremely high center of gravity positions), the system additionally detects the lifting of curve-inside wheels. Thereby these are monitored with regard to non-plausible wheel speed behavior. If necessary, the whole vehicle combination is strongly decelerated by a suitable braking intervention.

The lifting of curve-inside wheels on the trailer is detected by means of the trailer EBS. Therefore a slight test braking at the trailer is made at certain lateral acceleration, which, together with a strongly relieved wheel, leads to blocking and thus leads to an activation of the trailer ABS. This is communicated to the towing vehicle via CAN communication lines (SAE J 11992). For combinations with conventionally braked trailers (just equipped with ABS), the recognition of the wheel lifting on curve-inside wheels is limited to the motor vehicle.

3.3 ESP for Full- and Multiple-Trailer Combinations

The terms full- and multiple-trailer combinations represent all vehicle combinations that feature additional joints in comparison to articulated vehicles. These include the following combinations among others:

- **Standard full-trailer combination**: Heavy-duty vehicles with full-trailers, whereby the full-trailer normally has two or three axles, in northern countries up to four or five axles
- ***Eurocombi***: Heavy-duty vehicle with dolly (very short, mostly two-axle trailer vehicle with a fifth wheel, coupled to the vehicle using a draw bar) and semi-trailer
- ***Eurocombi***: Tractor-trailer and in addition coupled center-axle trailer
- ***A-double combination***: Tractor-trailer and coupled full-trailers (as an alternative to a full-trailer also dolly with semi-trailer)
- ***B-double combination***: Tractor-trailer with two articulated trailers (the first one is designed with a so-called dolly link with fifth wheel attached to the second semi-trailer)

The first combination is mainly used in Central and Northern Europe, while the other combinations are licensed, for example, in Scandinavia, Australia, and North America. Additionally there are so-called road trains existing in Australia and other countries, i.e., vehicle combinations with more than two trailers (partly up to 50 m in length and 150 t in weight).

Due to the additional joints, we receive degrees of freedom that lead to a significantly more complex driving behavior. The vehicle dynamics control takes this into account and introduces stabilization interventions much earlier, but also applied much more carefully. The reason is that a too powerful stabilization intervention could destabilize the vehicle combination, which has to be prevented under all circumstances.

In order to be able to evaluate the driving situation correctly, extended reference models have to take the additional degrees of freedom into account. This is aggravated by the fact that the number and type of trailers are rarely known to the system and no additional sensor information is recorded. For a robust illustration of the trailer behavior, the jackknifing estimation was extended by additional estimation factors, for example, the trailer lateral acceleration.

3.4 System Architecture

3.4.1 Conventional Pneumatic Service Brake

In the case of a conventional pneumatic braking system, the ESP is added to the ABS/TCS architecture (see Sect. 2.3.1). Thereby there is at least the opportunity to brake single wheels at the driven (rear) axles independently of the driver. The vehicle dynamics control needs additional autonomous braking interventions at the

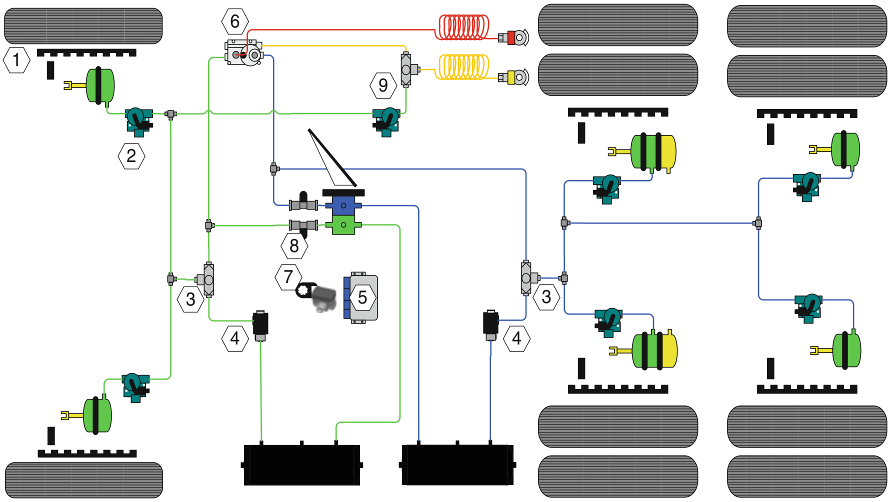


Fig. 9 System setup of a conventional pneumatic service braking system with vehicle dynamics control: *1* wheel speed sensor, *2* ABS valve, *3* relay valve, *4* TCS valve, *5* ECU, *6* trailer control valve, *7* ESP sensors, *8* pressure sensors, *9* trailer control

front axle and trailer. The system setup of a vehicle dynamics control based on a conventional pneumatic braking unit is displayed in Fig. 9.

Sensors In addition to the steering wheel sensor, commercial vehicle ESP requires sensors for the yaw rate and the lateral acceleration, which are similar to the passenger vehicle dynamics control (position 7 in Fig. 9).

Usually, the steering wheel angle is measured directly under the steering wheel in the steering column. Here, on the one hand, multi-turn capable magnetic field sensors are used, which can sense several revolutions with the help of mechanic gears. On the other hand optical sensors are used, which can only measure one revolution and thus realize the measurement of several revolutions using software functions. The sensors include a microcontroller and communicate with the central control unit via a CAN bus. It is either the general vehicle CAN (e.g., according to SAE J1939) or a separate sensor CAN bus.

For measuring the vehicle movement (yaw rate and lateral acceleration), modified vehicle dynamics sensors from the passenger vehicle sector are used. The mounting is done near to the center of gravity in the vehicle frame. Therefore the sensors must especially be adapted to the tough operating conditions in commercial vehicles (environmental impacts, vibrations, etc.).

In addition to the actual vehicle dynamics sensor, there are pressure sensors needed for sensing the driver demanded braking pressure, because of the fact that this is decoupled from the brake cylinders in case of a stabilization intervention and thus must be electronically controlled by means of the vehicle dynamics system (position 8 in Fig. 9).

Actuators In order to realize the extended intervention options of the vehicle dynamics control on the front axle and for the trailer, an additional TCS valve for the front axle braking circuit is deployed (position 4 in Fig. 9). From this braking circuit, the trailer is actuated with the help of another ABS valve (position 9 in Fig. 9).

3.4.2 Electronically Controlled Service Brake (EBS)

Due to the fact that the EBS already features technical preconditions to autonomously brake single wheels, the vehicle dynamics control needs only the vehicle dynamics sensors described in the previous paragraph. These communicate also via a CAN databus with the EBS central control unit, in which the algorithm of the vehicle dynamics control is implemented. The calculated nominal braking pressure is sent to the EPMs via the brake CAN bus, respectively via the trailer CAN bus to the trailer.

3.5 Special Functions in Commercial Vehicles

3.5.1 Basic Systems for Rollover Stabilization

Besides the vehicle dynamics control, there are basic systems which only address the rollover of the vehicle. These built up the ABS/TCS architecture and use the integrated lateral acceleration sensor for detecting the rollover tendency. If the vehicle tends to roll over, there is an active braking process on the rear axle and thus the vehicle speed is reduced. This process is realized with the help of the TCS valve and the ABS valves, as in the TCS braking controller. By means of another TCS valve, which is installed in the braking circuit of the trailer, the trailer can also be braked. As the TCS valve is a mere switching valve, the actuation of the trailer brakes is modulated, so that the mean effective braking pressure in the trailer remains limited due to the inertia of the braking system.

Due to the fact that no other sensors are present for the determination of yaw stability and furthermore only the rear axle and the trailer are braked actively, the achievable system performance in comparison to the full ESP is already basically limited. Furthermore, the braking interventions must take place carefully to prevent system-induced instabilities.

Expansion stages of this system use steering wheel-angle sensors additionally, in order to shape the stabilization interventions in dynamic maneuvers more efficiently.

The basic systems described here are mainly used outside of Europe up to now.

3.5.2 Trailer Systems for Rollover Stabilization

In addition to the vehicle dynamics control installed in the motor vehicle, which stabilizes the entire combination from the motor vehicle, there is an autonomously acting stabilization system to avoid rollover inside the trailer. This system called TRSP (Trailer Roll Stability Program) brakes the trailer autonomously in case of a rollover risk. As a principle, the TRSP functions similar as described in Sect. 3.2.2, but as measurement values there are only the load information and the lateral acceleration in addition to wheel speeds available. Due to the braking interventions limited to the trailer and the limited sensor information (only lateral acceleration), the performance with regard to the ESP is limited, which is nevertheless partly compensated using the locally available wheel speed values and extended options to detect rollover.

4 Outlook

The vehicle dynamics controls for commercial vehicles that are on the marked today are available for the following vehicle configurations:

- Heavy-duty vehicles with wheel formulas 4 × 2, 6 × 2, 6 × 4, and 8 × 4 in solo operation and with one or more trailers (full- and multi-trailer combinations, Eurocombi).
- Semi-trailer combinations with wheel formula 4 × 2, 6 × 2, and 6 × 4.

The stabilization interventions include engine interventions as well as active braking of single wheels on the motor vehicle and braking of the trailer.

In Europe, the ESP is mandated since 2009 for commercial vehicles with a maximum of three axles (stepwise, beginning with tractor semi-trailer). Therefore at the moment intensive further developments are taking place with the target of applying the vehicle dynamics control also for additional vehicle types, for example, for all-wheel-driven vehicles.

Due to the increasing distribution of the ESP for commercial vehicles, the algorithms must be increasingly robust and auto-adapt the vehicle behavior as far as possible. In addition to the adaptations to vehicle weight and self-steering behavior as described above, therefore, also the height of the vehicle center of gravity should be taken into account in the system.

Furthermore additional actuators will be integrated in the control in the future (e.g., steering units similar to passenger vehicles) as soon as these are available in commercial vehicles.

4.1 ESP for All-Wheel-Driven Vehicles

The challenges for a vehicle dynamics control for all-wheel-driven vehicles are mainly:

- The determination of the vehicle reference speed which must be supported on all wheels using additional sensors (e.g., for the longitudinal acceleration) on slippery ground due to the drive torques
- The off-road operation, which requires especially an adaptation of the ESP internal vehicle dynamics models and monitoring functions due to its complex driving situation

Furthermore, most of the available differential locks must be integrated in an appropriate way in the vehicle dynamics interventions.

4.2 Further Adaptation Algorithms in the ESP

As described in Sect. 3.2.2, current ESP systems assume medium center of gravity heights that lead to a very good compromise between drivability and safety in the system design with regard to the weight of the vehicle. In case of a known high center of gravity position, an increase in the ESP performance using early stage interventions is possible.

The roll and pitch behavior of the vehicle for known vehicle weights gives an indication of the vertical center of gravity position. Therefore the sensors of the level control system which are normally mounted at the rear axle on the left and right side are evaluated in an identification algorithm.

4.3 Use of Additional Actuators

In heavy commercial vehicles, there is no merely electric servo steering on the front axle existing, in contrast to passenger vehicles. Nevertheless, there are hydraulic servo steerings with electric torque overlay technology on the market since 2012 (e.g., Mercedes and Volvo). The electric actuator is similar to the electric servo steering of passenger vehicles. Thus the steering forces are further reduced and, e. g., at Mercedes the second hydraulic circuit for vehicles with double steered front axles can be skipped. Furthermore, electrically controlled steerings on additional axles (pusher or tag axle) have been in place for quite some time. Hereby the actual steering angle of the front axle is measured and converted into a nominal steering angle for the additional axle. With the help of a servo-hydraulic steering actuator, the steering angle is then adjusted.

Future vehicle dynamics control systems in commercial vehicles will include the possibility of such additional actuators in the control strategy in order to constitute the best possible stabilization function. The target here is to overcome the disadvantages of the braking interventions, which cannot be implemented continuously and in certain situations lead to traction losses, with the help of a continuous actuator. The steering interventions are therefore prefixed with regard to the braking interventions.

References

Breuer B, Bill KH (2006) Bremsenhandbuch. Grundlagen, Komponenten, Systeme, Fahrdynamik (Brake-Handbook. Basics, components, systems, vehicle dynamics). Vieweg + Teubner Verlag, Wiesbaden

Hecker F, Hummel S, Jundt O, Leimbach KD, Faye I, Schramm H (1997) Vehicle dynamics control for commercial vehicles. SAE-Paper 973284

Hoepke E, Breuer S (2006) Nutzfahrzeugtechnik. Grundlagen, Systeme, Komponenten (Commercial vehicle technology. Basics, systems, components). Vieweg + Teubner Verlag, Wiesbaden

Robert Bosch GmbH (2014) Kraftfahrtechnisches Taschenbuch (Bosch automotive handbook). Vieweg + Teubner Verlag, Wiesbaden

Winkler CB (1996) Simplified analysis of the steady state turning of complex vehicles. Int J Vehicle Mech Mobility 29(3):141–180

Zomotor A (1991) Fahrwerktechnik: Fahrverhalten (Chassis technology: road performance). Vogel Buchverlag, Würzburg

Part IX

DAS on Road- and Navigation Level

Visibility Improvement Systems for Passenger Cars

43

Tran Quoc Khanh

Contents

Abstract

Traffic accidents at night time have big economic consequences hence the characterization of accident events is very important. At night, the velocity of processing relevant visual information is low in the human subject and the risk of an accident is increased compared to daytime. For lighting design for night-time traffic, the visual process is not only influenced by the visual targets to be detected but also the observers (i.e., the drivers). For good object detection and recognition performance, a certain minimal luminance level must be provided on the road and its surrounding field by the vehicle's lighting system. The low luminance on the road during dark hours, the small contrast between the objects and their surroundings and the low conspicuity of the objects in traffic space originate from the limited range of illuminated roadside with current low

---

T.Q. Khanh (✉)
Fachgebiet Lichttechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: khanh@lichttechnik.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_44

beam front lighting illumination systems. The improvement of visibility and a substantial reduction of traffic accidents are only possible by an increased usage of optimized high beam. To develop high-quality front lighting systems, advanced light source technologies, adaptive light distributions and so-called "light-based" lighting functions shall be used. The primary aim is the achievement of long visibility distances in all traffic situations.

1 Frequency of Traffic Accidents at Night-Time and Under Adverse Weather Conditions

Traffic accidents at night time have big economic consequences. According to the study of (Rumar 2001), the estimated cost of traffic accidents in 1999 was more than 160 billion Euros which equals about twice the budget of the EU countries in the regarded time period.

For the analysis in this chapter, relevant accident research data of the institute for vehicle safety in Munich was used (Langwieder and Bäumler 1997). According to this data source, accidents with participation of pedestrians show different characteristics depending on the type of site where they occur. About a third of the 43,789 accidents with injured pedestrians in 1995 took place at night. About 60 % of the 1336 accidents with killed pedestrians in 1995 occurred in the darkness. About 84 % of the pedestrians participating in the accidents inside a town wore clothes in rather dark colour tones. Regarding the data in Langwieder and Bäumler (1997), in 70 % of the analysed accidents, street lighting systems were operating and could be evaluated subjectively as good.

The evaluation of night-time traffic accidents was based on the data published by the Federal Highway Research Institute of Germany in 2005 (Lerner et al. 2005) which collected the official data about traffic accidents in the time period from 1991 until 2002. These data can be regarded as valid for the traffic situations in this period. In this data pool, there are many aspects which are useful for the analysis and characterization of traffic accidents. The data include:

- Distribution of accidents according to federal states and sites (inside towns, federal streets and highways);
- Time distribution of traffic accidents;
- Accident type and accident conditions;
- Accident participants (age, gender); and
- Main causers of the accident depending on the type of the participants (pedestrians, cars, trucks, bicycles, motorbikes).

From the point of view of the lighting engineer, the characterization of accident events according to their time distribution is very important. The percentage of night-time accidents depending on month was calculated from 1999 until 2002. It can be concluded that the accident rate had a maximum in the autumn-winter months from October to February and a minimum in the spring-summer months

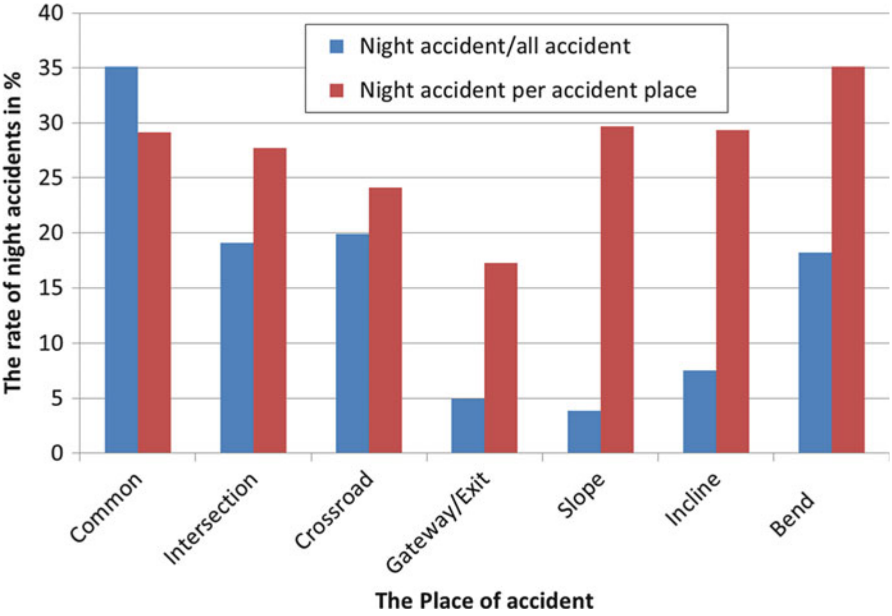


Fig. 1 Accident ratios at night time listed according to the accident sites in the years 1991, 2001 and 2002 (Lerner et al. 2005)

from May to July. The ratio of night-time accidents in the months from November until January turned out to be three times as high as the values in the months from May until July. The reason for this increase can be attributed to the deterioration of sight conditions during the darkness periods which are longer in the winter than in the summer. Other reasons include worse weather and road conditions in the winter months.

If night time accidents are analysed from the point of view of accident type according to the conflict situation and the conditions before the accident, the following problems can be pointed out (see Fig. 1):

- The number of accidents relative to the total number of all accidents (sum of daytime and night-time accidents) is especially high at crossing areas, intersections and bends with the percentage values between 15 % and 20 %;
- The number of night-time accidents relative to the total number of all accidents for a given accident site has a maximum at bends, with 35 %. However, this ratio is also high at upward and downward slopes, at crossing areas as well as at intersections, with values of more than 20 %.

This result implies that the visibility of objects and obstacles on the road and on both roadsides during the night – depending on the type of the vehicle's light source (Xenon HID lamp or halogen tungsten lamp), the light distribution on the road and the correct aiming of the front lighting system – is not sufficient. This deficit can

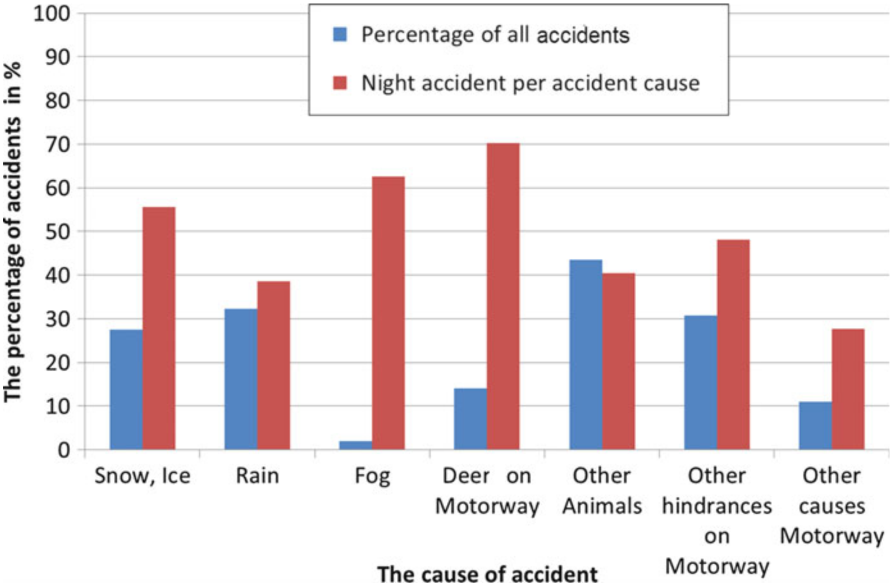


Fig. 2 Night time accidents grouped according to different accident causes in 2002 (Lerner et al. 2005)

have serious consequences at crossing areas and intersections. On special road tracks with upward and downward slopes sections, visibility distances can be reduced strongly hence detection ability is reduced. Additionally, even in case of a correctly aimed low beam, it is often very difficult to detect an object and explore the structure of a bend simultaneously.

If the causes of night-time accidents are analysed in the year 2002 in detail, the following aspects can be seen (see Fig. 2):

– The ratio of night time accidents with snow, ice and rain relative to the total number of accidents is comparatively high with a value of 27 %;
– For every cause of an accident, the ratio of night-time accidents is remarkably high. This ratio was found to be more than 55 % for the conditions with snow, ice, fog and wild animals. Additionally, the high ratio of night-time accidents with the cause of "other animals on the road" or "other obstacles on the road" implies the late detection of the objects on the road during night-time periods.

Generally, the relatively low luminance on the road during the dark hours, the small contrast between the objects and their surroundings and the related low conspicuity of the objects in traffic space originate from short visibility distance and the limited range of illuminated roadside with current low beam front lighting illumination systems. This will be the subject of analysis in the next pages. Manually operated high beam can ensure a much longer visibility distance but it

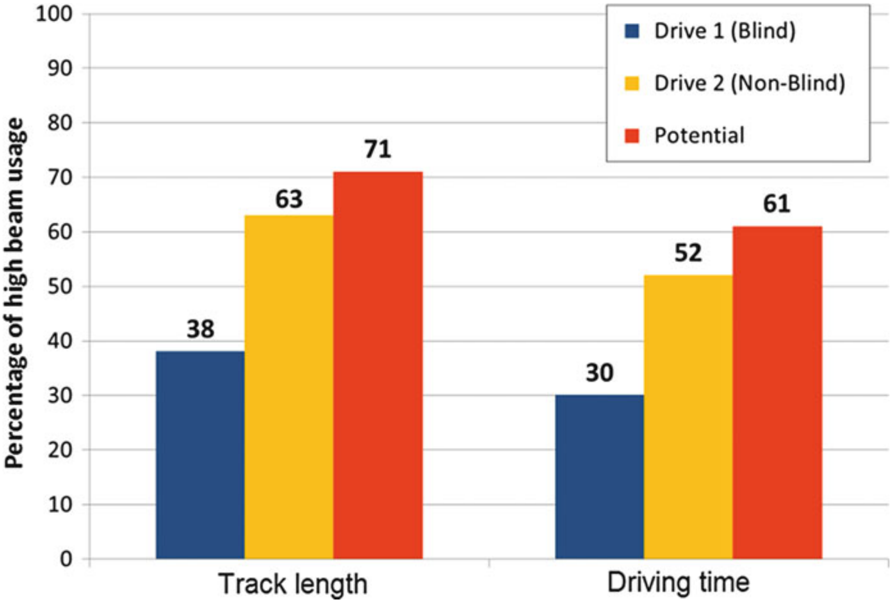


Fig. 3 Percentage of high beam usage during "blind" and "non-blind" tests (see text) on German roads (Sprute 2012)

may also produce more glare effect for the other traffic participants (oncoming and preceding traffic, pedestrians). Therefore, this manual handling of high beam is not widely used.

Sprute (2012) studied high beam application behaviour in his Ph.D. thesis. His vehicle was equipped with a camera system to recognize the objects and to record whether a situation allowed the use of high beam or not. The test track was driven along twice. First, test drivers did not know the aim of the experiment (i.e., high beam usage behaviour, this is why it was a "blind" test). Second, test drivers were asked to use high beam as often as they could ("non-blind" test). In Fig. 3, the percentage of high beam usage rate, calculated according to the length of the test track and to driving time, and the best possible way of usage are shown. Using the criterion of test track length, high beam usage rate increased from 38 % (blind test) to 63 % (non-blind test).

This knowledge is important for automobile lighting technology. It implies that the improvement of visibility and a substantial reduction of traffic accidents are only possible by an increased usage of high beam systems. This can be achieved either by a camera-controlled and timely activated high beam function depending on the traffic situation (high beam assistance, see Fig. 4), or by a permanent high beam function system with an automatic object identification and localization unit so that other traffic participants can be detected in the actual traffic space, and the illumination levels in the corresponding space

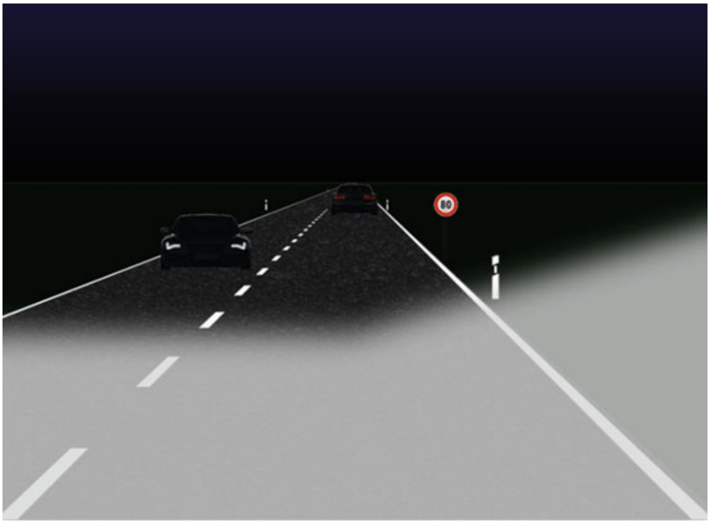


Fig. 4 High beam assistance (Totzauer 2013)

segments can be reduced down to a value under the just-glaring limit (glare-free high beam, pixel light beam). These lighting functions shall be the subject of the next sections.

2 Consequences for Lighting and Vehicle Design for the Systems That Improve Visual Conditions

It is a well-known fact that about 90 % of the information from our environment is registered by the aid of the visual path. The amount of information flow during the daytime is very high and it cannot be processed within a short time. At night, the velocity of processing the relevant visual information is relatively low and therefore the risk of an accident is substantially increased compared to daytime. The luminance range in traffic space on the road and at the roadsides is generally between 0.01 and 10 cd/m^2 so that mesopic vision can be assumed in which rods and cones act parallel.

According to literature (Eckert 1993), the visual process can be divided into three subsequent steps: seeing (with the traffic relevant task of visual search), detection and identification. The optical radiation in the visible range (the light) that enters the eye apparatus passes through the cornea, the eye lens with the iris and reaches the retina containing the photoreceptor structure including cones, rods and ganglion cells. All components of this eye pathway influence visual capacity in a specific way. After having absorbed the photons hence detecting visual information, the resulting neural signals are transferred to the brain's processing units so that the seeing step can take place. If the visual process of the cognitive stage delivers relevant information (i.e., if a certain contrast can be registered), the detection and identification process will begin. After that, a comparison of the identified object and object position with previous visual experience will lead to an action (e.g., the onset of braking).

From the point of view of lighting design for daytime traffic and especially for night-time traffic, the visual process can be influenced by two main groups of factors. These two groups are:

- Visual targets and objects to be detected; and
- Drivers (observers).

The object aspects can be characterized by the following components:

- Optical characteristics of the object:
  - Form, size, colour (hue, saturation);
  - Reflectance;
  - Presentation time and position in the viewing field.
- Optical characteristics of the object's surrounding field:
  - Contrast between the object and its direct surrounding field;
  - Illumination from street lighting luminaires and vehicle front lighting systems;
  - Disturbance of the seeing (incl. visual search) and detection process from glaring light sources and advertisement panels in a city at night.

The driver aspects can be characterized by the following components:

- *Bright adaptation process and dark adaptation process:*
  At the transition phase from a bright to a dark or from a dark to a bright environment, the human visual system must adapt to the current luminous intensity level by means of several human eye physiological processes. This procedure can take place, e.g., when entering a tunnel or when leaving a tunnel.
- *Aging:*
  Visual performance parameters like visual acuity, contrast detection performance and reaction capability becomes worse with increasing age.
- *Glare:*
  Glare is caused by high luminance or by a highly inhomogeneous luminance distribution in the viewing field. The human eye-physiological cause for the experience of glare is stray light in the eye apparatus which originates from the glaring light sources (e.g., from the front lighting systems of an oncoming vehicle). The stray process takes place when light passes through the ocular media (eye lens, cornea, glass body). It also results from light reflections from the retinal surface. This stray light is superimposed onto the image of the object to be detected and its surrounding field hence leading to the reduction of contrast.

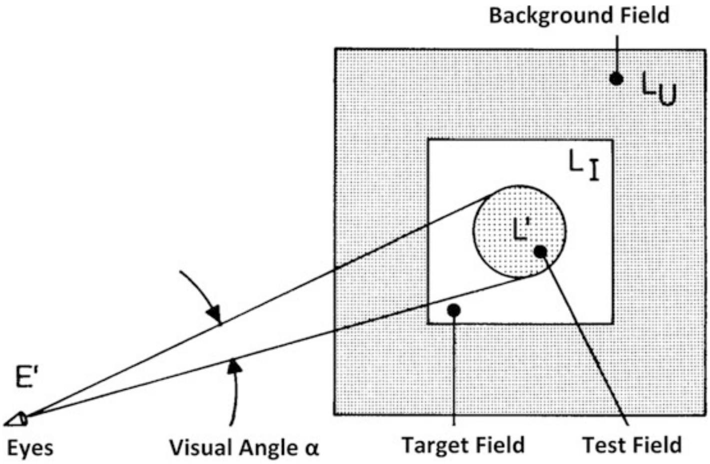


Fig. 5 A typical viewing situation in which contrasts are perceived (Khanh 2013)

In lighting science and engineering, contrast (C) is defined as shown in Eq. 1.

$$C = \left(L' - L_U\right)/L_U = \Delta L/L_U \tag{1}$$

With:

L': luminance of the target (object) in cd/m^2

L_U: luminance of the surrounding field of the target in cd/m^2

The inverse value of the contrast C is called contrast sensitivity (U_{E}).

$$U_E = 1/C = L_U/\Delta L \tag{2}$$

Figure 5 illustrates a typical viewing situation in which contrasts are perceived.

There is a well-defined relationship between contrast sensitivity and adaptation luminance which can be computed, e.g., as the average luminance on the road in front of the vehicle. The higher the adaptation luminance, the higher is contrast sensitivity. In other words, the just-noticeable contrast between the target (e.g., an object on the road or an animal beside the road or traffic signs) and its direct surrounding field is reduced. As a consequence, for sufficient object detection (and recognition) performance, a certain minimal luminance level must be provided on the road and its surrounding field by vehicle lighting systems.

Sprute (2012) studied the relationship between glare illuminance measured at the eye and the minimal just-detectable luminance difference of a target to its surrounding area experimentally, for four distances between the glaring light sources and the observer. If a car driver is glared by the front lighting system of an oncoming vehicle then the illuminance measured at the eye depends on the distance between the glaring vehicle and the observed vehicle. As a consequence, the minimal just-detectable luminance difference between the target (object) and its direct surrounding area also changes with this glare illuminance, see Fig. 6.

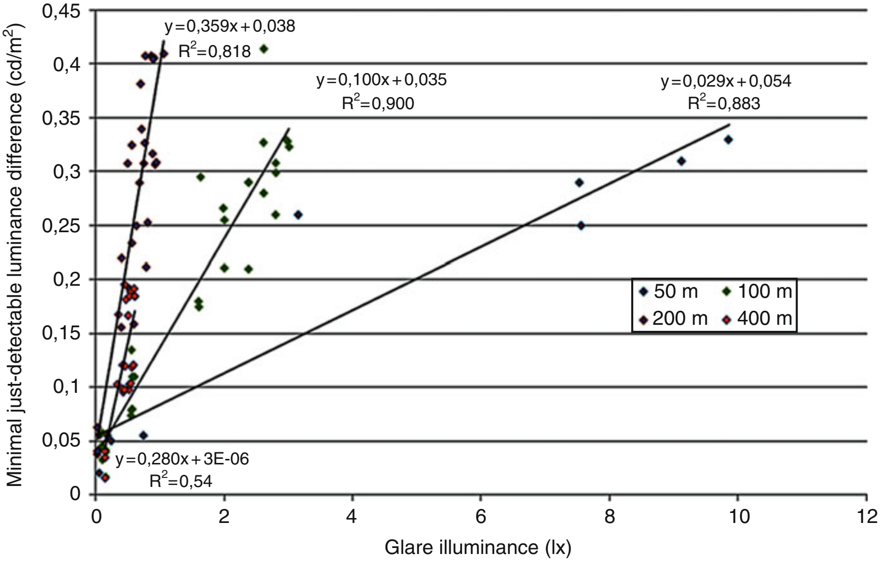


Fig. 6 Relationship between glare illuminance and minimal just-detectable luminance for four object-observer-distances: 50, 100, 200 and 400 m (Sprute 2012)

From the above described aspects, the following requirements of lighting and vehicle technology can be derived in order to improve the visual conditions of vehicle drivers:

- *Requirement 1*:
  Realization of a bright and homogenous lighting distribution on the road with the front lighting system in order to achieve the maximally possible visibility distance. This means a wide lateral illumination of the areas on the road and at the two roadsides for the safe and fast identification of traffic shields, guide posts and other objects alongside the road (e.g., trees). This broad road illumination improves the safety feeling of car drivers during the journey. Also, a remarkable portion of luminous flux emitted by front lighting systems must be imaged along the road axis towards long distances in front of the car in order to improve contrast and enhance visibility distance. Headlamps with their optical imaging systems use light sources with a high amount of luminous flux and optimize the optical components of the front lighting systems.
- *Requirement 2:*
  Minimization or elimination of glare for the oncoming and preceding car drivers. The luminous intensity distribution of front lighting systems and the whole operating system of the vehicle (e.g., the dynamic levelling unit) must be designed so that the illuminance at the eye of the oncoming and preceding car drivers should not exceed the maximally permitted values defined in the international regulations.

At the beginning of the automotive era, high beam was permanently used due to the fact that the luminous intensity in that time was not very high. According to the two requirements described above, low beam has been introduced and steadily improved by the optimization of the mechanical, lighting and electrical components. Depending on their configurations, current low beam systems can have a visibility distance between 50–60 m and 88 m. Since 1995 until now, lighting, mechanical and electrical systems have been optimized continuously in order to achieve substantially longer visibility distances and a better glare reduction. These systems will be described in Sect. 3.

3 Current and Future Front Lighting Systems to Improve Visual Performance

The development of current and future front lighting systems for the improvement of visual conditions, visual performance and visibility can be characterized by three different development concepts:

- Advancement of light source technologies;
- Advancement of adaptive light distributions;
- Further development of so-called “light-based” lighting functions.

3.1 Systems to Improve Visual Performance Based on Light Source Technology

Current vehicle front lighting systems have been using tungsten halogen lamps (from 1959 on) or xenon discharge lamps (Xenon-HID, from 1990 on), light-emitting diodes (LEDs, since 2007) as well as laser diodes (since 2014). From current data, the market share of front lighting systems with different light source technologies are shown in Table 1.

Besides the technological considerations for the light sources, environmental aspects will also play an important role in the lighting concepts of car makers. For the low beam with tungsten halogen lamps, an electrical power of 62 W coming from 55 W for the lamp and 7 W for the ballast unit is needed so that the optimization of tungsten halogen lamp technology or its substitution by a new, more energy efficient light source does also mean a substantial additional value in terms of environmental protection. Since 2007, LED technology has also been used in front lighting systems because this semiconductor light source has the following significant advantages:

- LED components with high power and with a current of about 1 A generally have a life time of more than 10,000 h which is longer than the life time of the whole vehicle. Failure rates are very low so that a change of the light source unit

Table 1 Market share of front lighting system lamp technologies (Fratty et al. 2014)

Year/ type	Halogen lamp (%)	Xenon (%)	LED (%)	Laser (%)	Hybrid LED/laser (%)
2013	84	15	1	0	0
2020	60	10	28	0	2
2025	30–50	0–5	30–50	1–5	10–20

during the car's entire lifetime can be avoided. It is a big advantage in comparison to the other light sources like tungsten halogen lamps or Xenon-HID lamps;

– LED devices can be dimmed or switched (on/off) very fast (of the order of microseconds) without changing their light colour and constitute the precondition for the design of intelligent and adaptive front lighting and signalling systems in order to vary light distributions in traffic space depending on traffic situation and driver characteristics (age, eye, viewing behaviour, fatigue as well as their individual needs concerning light intensity);

– LED devices are relatively small, compact and stable against vibrations hence suitable for the design of various lighting functions (high beam and low beam, bending light, cornering light, rear lamps and day time running light) in which the functionality and design orientation can be combined for higher user acceptance and value of the car brand.

The current dynamic development of LED technology in automotive lighting driven by the technological, environment-protective and design-orientated aspects enables the development of complete front lighting systems with a high degree of functionality in the vehicles of higher class and these development experiences can be transferred to the front lighting systems in medium class cars. In total, LED based lighting systems will substitute Xenon-HID lamp technology in the next years. In the next decade, two development tendencies shall be considered:

(a) *Development of energy efficient LED low beam*: current halogen tungsten lamps have a luminous flux of about 1500 lm at their nominal voltage and at 55 W electrical power and are able to deliver an effective luminous flux of 450 lm onto the road by means of the optical system with 30 % optical efficiency. This effective luminous flux can be achieved with a well-designed LED low beam with a typical optical efficiency of 50 % and with a luminous efficacy of current automotive-qualified white LEDs of 83 lm/W and with an electrical power of the LEDs of 11 W. If the efficiency of the power supply electronics is assumed to be 90 %, a total electrical power in the order of 12 W can be estimated. This LED low beam with 450 lm on the road should be the substitution for the tungsten halogen lamp low beam in the lower class car types.

(b) Development of a full-LED front lighting system with complete lighting functions (high and low beam, day time running light, position light, cornering and bending light, marking light): The current LED low beam system achieves the capacity of the best Xenon low beam system with a luminous flux on the

Table 2 Lighting characteristics of the light sources for vehicle front lighting systems (Khanh 2013)

Lamp type	Luminous flux	Max. luminance	Luminous efficacy	Colour temperature
Halogen tungsten lamp (H7)	~1500 lm	~30 Mcd/m^2	25 lm/W	3200 K
Xenon lamp (D2S)	~3200 lm	~90 Mcd/m^2	90 lm/W	4000–4200 K
LED	~150–1500 lm	~40 Mcd/m^2	83 lm/W	4000–5500 K

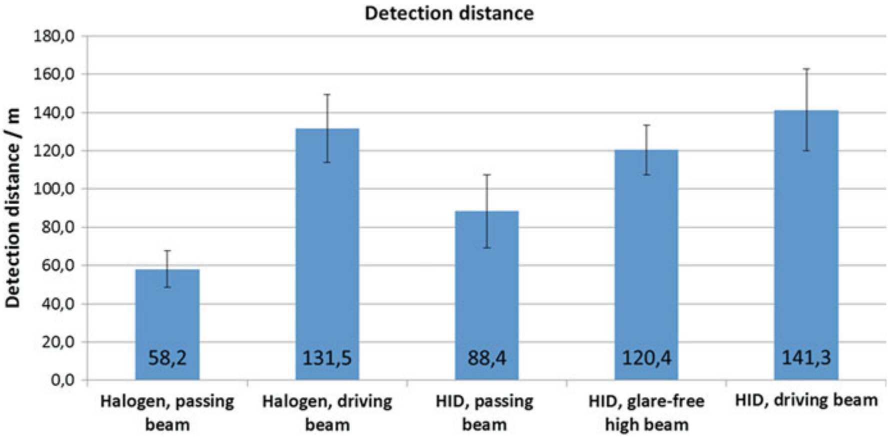


Fig. 7 Mean detection distance for different beam types (Zydek et al. 2013)

road of about 950–1050 lm. This tendency and concept are used in the high and medium class vehicles of premium car makers. A current Xenon low beam with 1000 lm on the road has an electrical power of 42 W (35 W for the lamp and 7 W for the electrical unit). An LED low beam with 1000 lm in the full front headlamp will only need about 27 W.

Table 2 summarizes the most important characteristics of the three relevant light sources for automotive front lighting systems (Khanh 2013):

In the last years, field tests with tungsten halogen lamp low (passing) beams and high (driving) beams with reflection optics, with Xenon lamp low and high beam with projection optics and with glare-free high beam with Xenon lamp have been reported (Zydek et al. 2013). Figure 7 shows visibility distances for the above front lighting systems.

The improvement of detection distance in comparison to the respective low (passing) beam is shown in Fig. 8. The study (Zydek et al. 2013) together with the result of former field tests (Schiller and Khanh 2007, 2008; Schiller 2007) also indicate that low beam front lighting systems with halogen tungsten lamp and

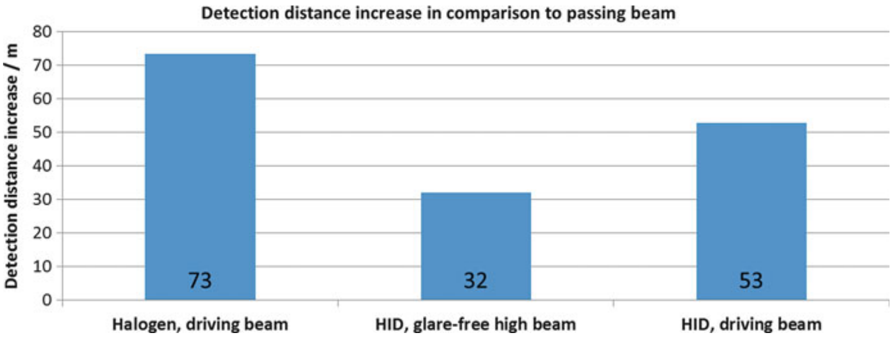


Fig. 8 Increase of detection distance (Zydek et al. 2013) relative to low (passing) beam

Xenon lamp do not exhibit a difference concerning the subjective glare evaluation by the test drivers if these low beams are correctly aimed.

3.2 Systems to Improve Visual Performance Based on Adaptive Light Distributions

In Sect. 3.1, visibility distances of current low beam (passing beam) systems were analysed (in Fig. 7). It was shown that this distance of the Xenon lamp low beam can only be of the order of 88 m. Generally, the visual process leading to braking activity after the identification of hazardous objects or dangerous situations on the road can be divided into three subsequent steps:

- A seeing (search) process with a successive fixation procedure in order to bring the image of the object into the foveal area (the site with the best visual acuity);
- A basic reaction time interval in which the object situation is being evaluated and decisions have to be made about the reaction to this situation;
- A braking procedure with several steps. The foot must be brought to the pedal which is then routed downwards until the braking activity can be activated. After this activity the vehicle begins to be decelerated with an average rate of -5.8 m/s^2.

From the calculations for this seeing and braking process, the relationship between braking distance and driving velocity can be determined. This is shown in Fig. 9. According to this diagram, a maximal visibility distance of 88 m can only allow a maximal speed of 90 km/h during the night. With modern low beam front lighting systems with tungsten halogen lamps with a visibility distance of 65 m only a maximum speed of 75 km/h can guarantee safe driving.

From the above discussion it is clear that an improvement of the visual conditions for night-time driving cannot only be achieved based on light source technology. The analysis of these complex problems and the accident analysis in Sect. 1

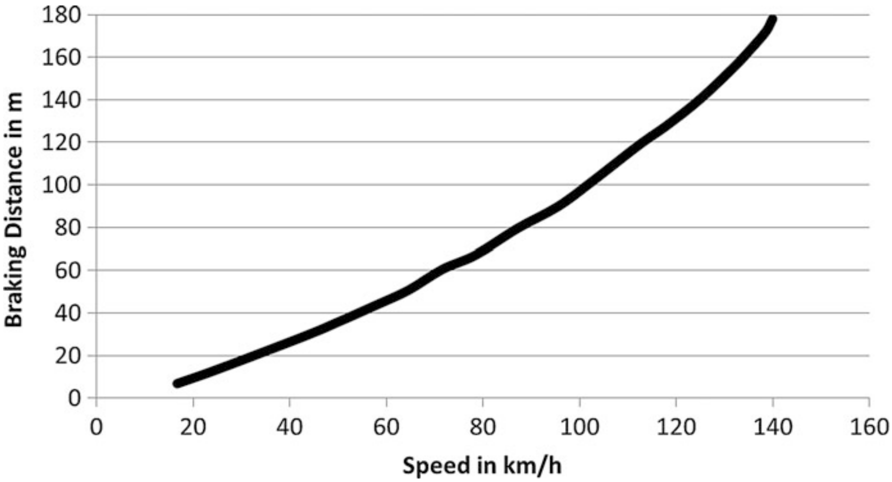


Fig. 9 Relationship between braking distance and driving velocity (Rosenhahn and Hamm 2003)

lead to the conclusion that intelligent adaptive front lighting systems have to be developed. These systems should have the following features:

- A maximal visibility distance which is much longer than the order of the current low beam functions. The latter should be controlled adaptively depending on road topology (i.e., raising and falling areas);
- A light distribution which can be adjusted for different traffic situations adaptively (driving speed, relative position to oncoming and preceding traffic participants, weather conditions like fog or rain). This light distribution has to offer the best possible viewing conditions along the road axis and lateral to the road axis with maximal visibility distance and minimal visual discomfort;
- A light distribution to be adjusted to the structure of traffic space in an adaptive manner (bends, intersections, city space) with variable distribution width for each case.

From 1995 on, many research studies have been conducted in order to formulate the concepts and steps for the realization of adaptive front lighting systems (AFS) in the vehicle industry. All these efforts led to the approval of the (UN/ECE standard 123 2007). AFS lighting systems (***A***daptive ***F***ront lighting ***S***ystem) include lighting functions like city light, country light, adverse weather light, motorway light, high beam and bending light with dynamic and static operations which will be described below. Figure 10 shows a selection of four different light distributions.

3.2.1 Low Beam/Country Light

Country light has been designed based on the principle of current low beam with an asymmetric light distribution with more light intensity falling on one’s own driving

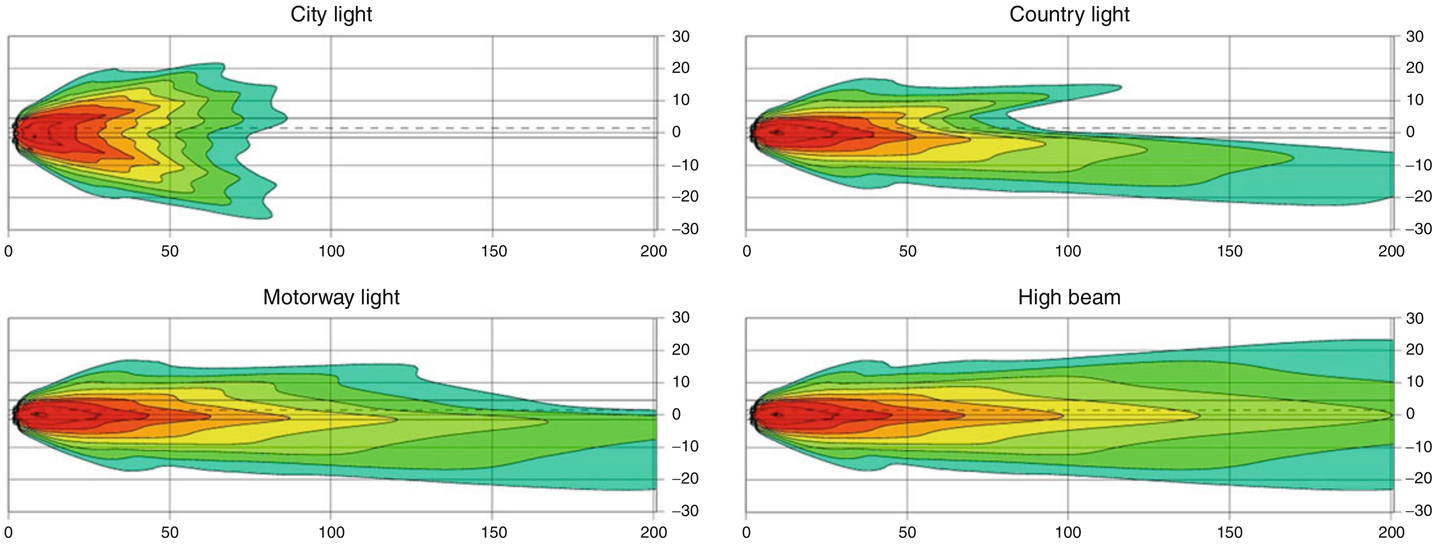


Fig. 10 Principal light distribution functions on the road from the bird's eye view (Kalze and Schmidt 2007)

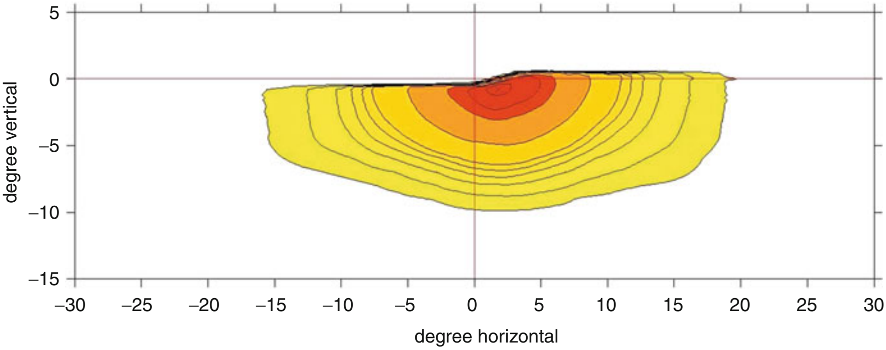


Fig. 11 Light distribution of country light constructed with high-power white LEDs on the 25 m ECE measuring screen (Rosenhahn 2007)

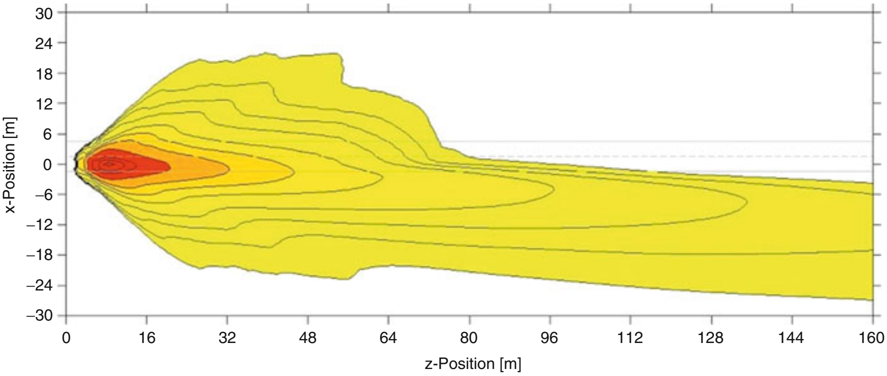


Fig. 12 Light distribution of country light on the road from the bird's eye view (Rosenhahn 2007)

lane, see Fig. 12. Figure 11 shows the light distribution of country light constructed with high-power white LEDs on the 25 m ECE measuring screen (Rosenhahn 2007) on which a horizontal light distribution with a width of about 36° and a well-defined cut-off-line with a focused hot-spot light zone (in red) below the crossing point of the horizontal and vertical axis (H-V-point) can be seen.

3.2.2 City Light

The light distribution of city light in Fig. 10 is broad on both sides and symmetrical in order to enhance object detection at the side areas of the road and at the intersections if driving speed is lower than 50 km/h. The visibility distance along the road axis is reduced.

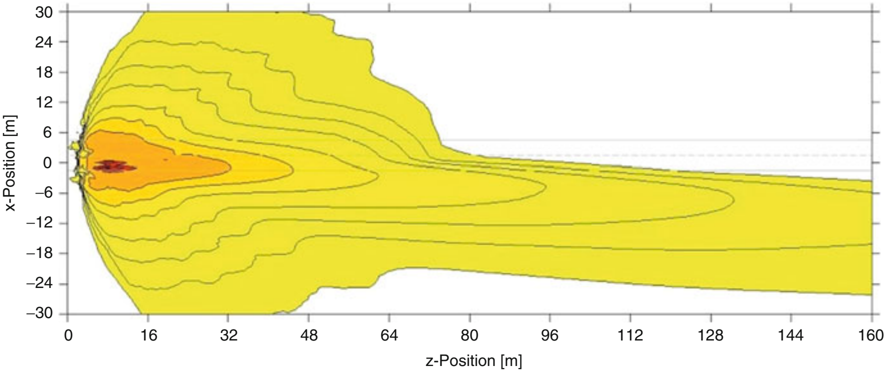


Fig. 13 Light distribution of adverse weather light on the road from the bird's eye view (Rosenhahn 2007)

3.2.3 Adverse Weather Light

According to the ECE regulation 123, the construction of adverse weather light follows two primary aims:

- The reduction of luminous intensity in the area in front of the vehicle. A strong luminous intensity which is strayed back by rain and fog particles increases glare potential and reduces contrast in front of the car.
- The increase of luminous intensity at both sides of the road enhances visual orientation under adverse weather conditions.

Figure 13 shows an example for adverse weather light distribution based on the principle of LED technology according to Rosenhahn (2007). This light distribution has a broader side luminous intensity in the range until 48 m in front of the vehicle in comparison to the distribution in Fig. 12.

3.2.4 Bending Light

The development and introduction of bending light in 2003 was the second important milestone after the introduction of Xenon-HID lamps in modern automotive lighting technology. The task of bending light is the enhancement of visibility distance in a bend. Dynamic bending light is realized in most cases by the rotation of the whole headlamp around the vertical axis until ±18° following steering angle or the radius of the bend (see Fig. 14).

By the use of LED technology, headlamp rotation is not necessary and bending light with LEDs can be constructed in two configurations:

- If the low beam consists of several LED units then only the LED unit for the focused hot-spot lighting area below the H-V-point must be rotated, see Fig. 11. This type of LED unit has been proven and tested with high power LEDs (Rosenhahn 2007).

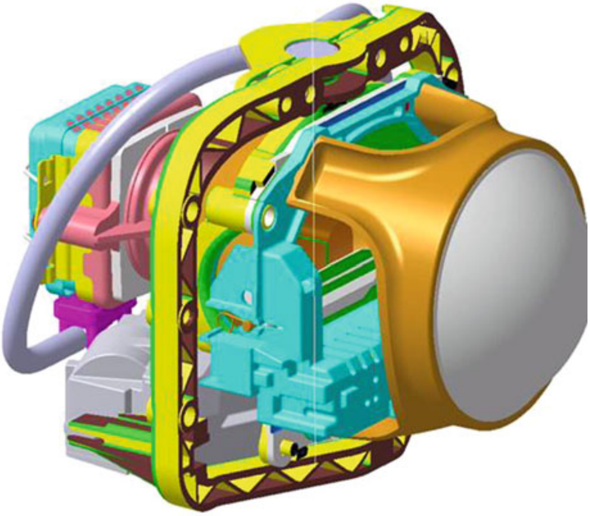


Fig. 14 Bending light projector on the basis of Xenon-HID lamps (Source: Company Valeo/France)

- If the bending light consists of a low beam with a certain light source (halogen tungsten lamp, Xenon lamp or LEDs) then, in the bend, a virtual subsequent illumination shift is realized with additional LED light source units which are activated depending on the actual angular position. This principle (according to Grimm and Casenave (2007)) is visualized in Fig. 15.

3.2.5 Motorway Light

By the use of motorway light, the visibility distance on the highway can be increased from about 88 m with conventional low beam front lighting systems up to about 110–120 m. From the technological point of view, there are three possibilities to realize this lighting function (Rosenhahn 2007):

- Raising the cut-off-line in vertical direction from $\beta = -0.57°$ to $\beta = -0.23°$;
- For the LED low beam consisting of several LED light source units (see Fig. 11 of ▶ Chap. 7, "AUTOSAR and Driver Assistance Systems"), the electrical current of the LEDs in the LED group for the focused lighting area below the H-V-point can be increased for a greater luminous intensity in the hot spot area. The effort for the electrical circuit and for the thermal management is relatively high;
- Activation of an additional lighting unit for the focused lighting area below the H-V-point.

In Fig. 16, the light distribution of motorway light based on LED technology (according to Rosenhahn (2007)) is illustrated.

Adaptive front lighting systems (AFS) with tungsten halogen lamps and Xenon HID lamps have been introduced into the market since 2006. In Fig. 17, its technical implementation is illustrated (Kalze and Schmidt 2007).

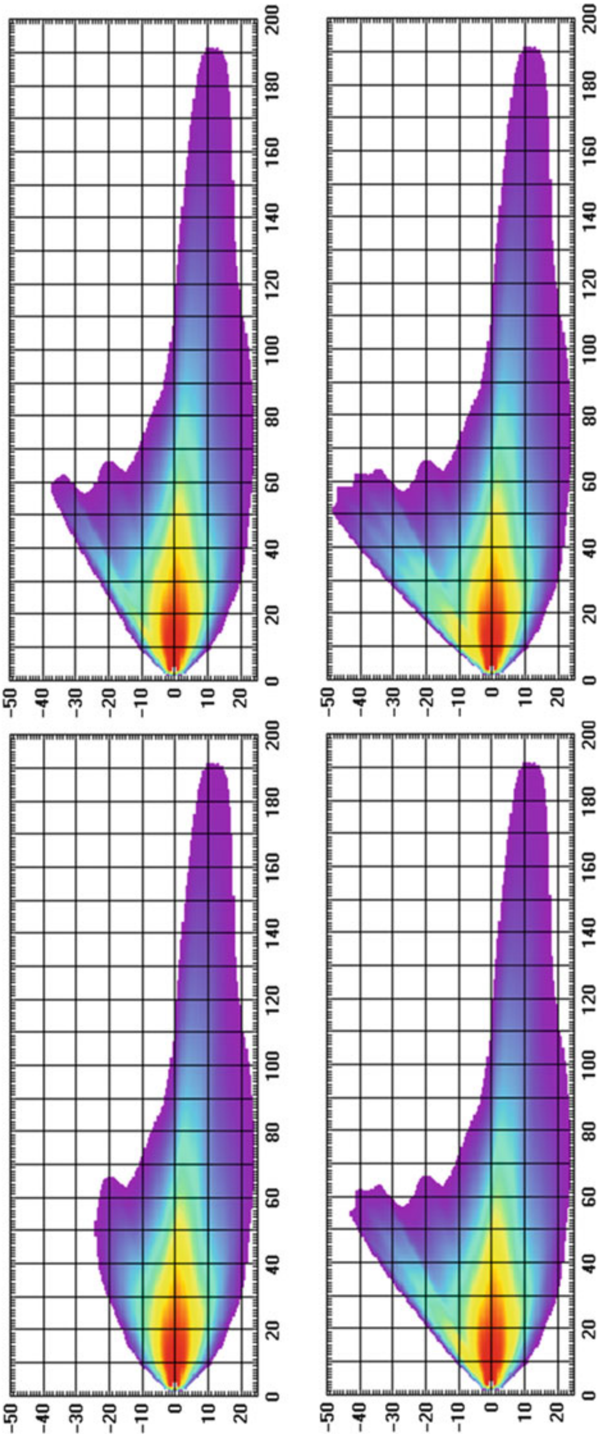


Fig. 15 Light distributions resulting from the sequential activation of three LED light source units in the bend (Grimm and Casenave 2007)

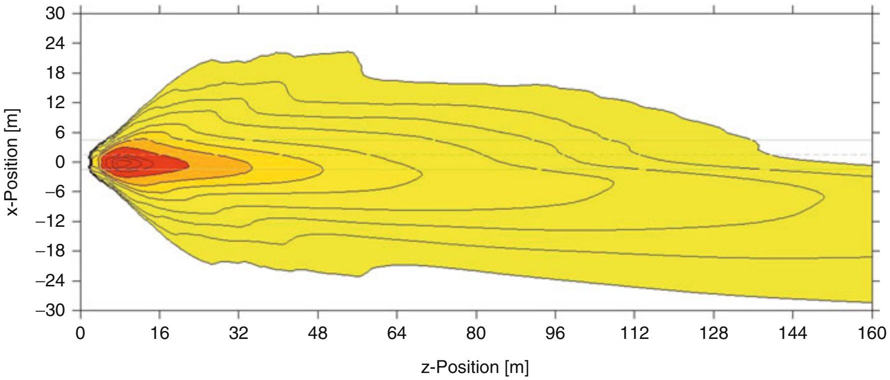


Fig. 16 Light distribution of motorway light based on LED technology (Rosenhahn 2007)

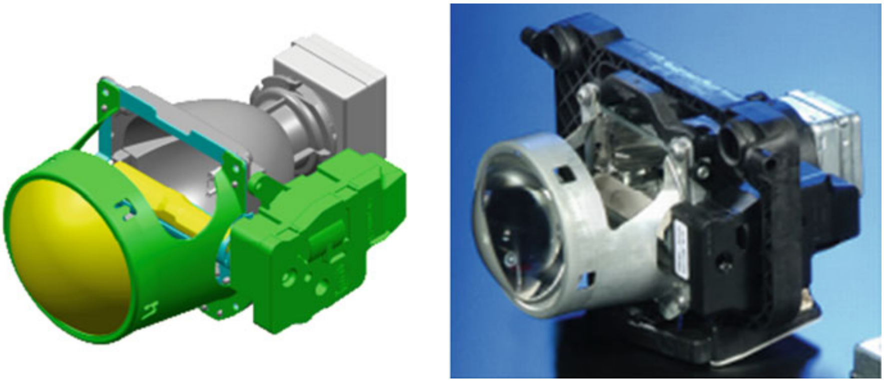


Fig. 17 The VarioX front lighting system with AFS function

The light source (tungsten halogen lamp or xenon lamp) is positioned into the 1st focal point of an ellipsoid mirror reflector so that the coil of the tungsten halogen lamp or the plasma arc of the xenon lamp can be imaged into the 2nd focal point of the reflector (see Fig. 17, left picture). Next to this 2nd point, a free-form cylinder is placed which can be rotated in defined steps by means of a precise motor. The cylinder has on its own cylinder barrel different contours with different forms for the realization of different AFS light distributions. Depending on the traffic situation, the suitable contour can be rotated into the optical path of the reflector.

The activation of AFS lighting functions is based on the evaluation of the different signals continuously recorded from traffic space by means of sensor systems like imaging cameras, LIDAR, RADAR and GPS. Additional signals from steering wheel or rain sensors can also be evaluated simultaneously. After the signal processing process, instructions are being generated for a suitable AFS lighting function. The structure of AFS systems is described in Fig. 18 (according to Sprute and Khanh (2007)).

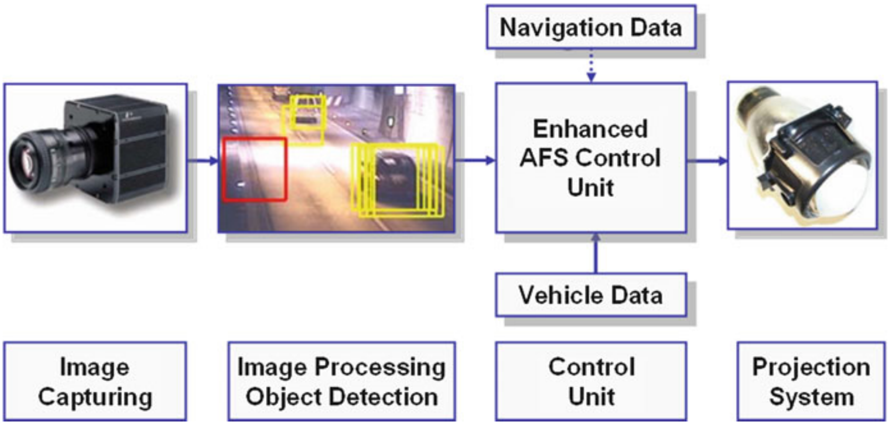


Fig. 18 Structure of AFS systems (Sprute and Khanh 2007)

3.3 Systems to Improve Visual Performance With Light-Based Adaptive Light Distributions

All AFS lighting functions described in Sect. 3.2 arose from a long-term extensive research and technological development and can be regarded as a substantial innovation step and progress in comparison to the current low beam headlamp. However, they have only been designed for traffic situations with general characteristics like bends, streets in the city or highways. For actual driving situations with dynamic changes (traffic density, occupation of the left and right lanes on the highway, changing distances between vehicles in the city and on the country road, etc.) intelligent front lighting systems are needed for optimal illumination (i.e., to ensure the minimal glare and maximal visibility condition) in every traffic situation. For this purpose, two technical requirements must be taken into account:

- Realization of a network of sensors which
  - Record, control and image traffic space around the actual car, with sufficient temporal and spatial resolution;
  - Detect and classify the objects in traffic space in a safe manner;
  - Determine the angular positions of the objects (other vehicles, traffic signs, pedestrians and animals on the road and the roadside, obstacles) in the vertical and horizontal direction and the distance between these objects and one’s own vehicle. The safe classification of the objects plays an important role in order to differentiate street lights from traffic reflectors and the front lighting headlamps of other cars from traffic shields. The signals of the different sensors have to be combined (signal fusion) and weighted depending on traffic situation in order to localize and identify substantial hazards well in advance;

- Realization of new concepts for front lighting systems which can generate spatially and temporally adaptive and variable light distributions in a dynamic process.

If the above two preconditions are fulfilled then it will become possible to develop intelligent headlamps with light distributions which

- Draw the attention of drivers on direct and indirect hazards during driving (e.g., an animal on the road). This is the principle of marking light;
- Vary the cut-off-line vertically depending on the distance between one's own car and other oncoming and preceding vehicles on the road. In this way, the visibility distance for the driver can be increased and the glare load for other traffic participants can be minimized. This is the technical principle of "*variable cut-off line*";
- Operate as a permanent high beam whose luminous intensity is reduced or switched off in those light cones in which other preceding and oncoming vehicles and traffic participants have been detected. This is the principle of the headlamp with "*glare-free high beam*".

These three lighting functions can be described in more detail as follows.

3.3.1 Marking Lights

Marking light consists of a pixel-resolved camera system (near or far infrared camera) which records and delivers optical information on the objects and their angular positions in horizontal and vertical directions and transfers this information to the controlling electronics in the car. After signal processing, a light projector with a small light cone is switched on and directed to the object without glaring it. The attention of the vehicle driver can be drawn in a short time towards this object effectively in order to initiate a suitable manoeuvre and avoid collision, see Fig. 19.

Marking light as a technical system based on LED technology was introduced into the market in 2011 in upper class vehicles. In the research study of D. Schneider, a reaction time reduction of the order of 0.42 s of the test drivers using marking light was registered in comparison to those using high beam (Schneider 2011a, b). In the same study, dynamic field tests with pedestrians standing at different positions at the roadside were conducted with marking light functions and Xenon lamp low beam systems. It was found that the visibility distance with marking light was 34 m longer than the visibility distance with low beam. This value means a time advantage of 1.2 s at a vehicle speed of 100 km/h.

3.3.2 Variable Cut-Off Line Lighting Function

The primary aim of this lighting function is the achievement of maximally possible visibility distance in actual traffic situations. Depending on the distance between one's own car and other traffic participants, the cut-off-line is varied in such a manner that glare load is avoided. This technical principle was realized (according

Fig. 19 Principle of marking light (Image source: Hella (Kleinkes et al. 2007))

to Kalze and Schmidt (2007)). It is visualized in Fig. 20. If no oncoming and preceding vehicle is detected then high beam is activated for maximal visibility distance. In the case of detecting other vehicles in traffic space, the cut-off-line is lowered depending on the relative position between one's own car and the vehicle with the shortest distance. If other vehicles come closer then the cut-off line is adjusted into the position of the low beam (Sprute and Khanh 2007).

3.3.3 Glare-Free High Beam

In this case, the headlamp is in the high beam operation mode and the camera system records and controls traffic space in a real-time process. The actual angular positions and distances of all vehicles are being calculated continuously. Luminous intensities are reduced or switched off in the angular positions of these vehicles with a sufficient angular resolution so that other drivers within these light cones cannot be glared. In the other light cones, light distribution and light intensity have the same characteristics as in a normal high beam so that maximal visibility distances can be achieved. The advantage of this glare-free high beam principle in comparison to the lighting mode of "*variable cut-off line*" is that high beam lighting characteristics can be maintained for a number of light cone segments. This comparison is illustrated in Fig. 21.

In the time period between 2013 and 2016, the glare-free high beam in Fig. 21 (Matrix beam, right image) has consisted/will consist of two parts:

- Below the cut-off line in the low beam area, illumination is realized by a normal LED-low beam;
- In the high beam area, the high beam system consists of several vertically positioned LED light segments with a limited horizontal angular resolution.

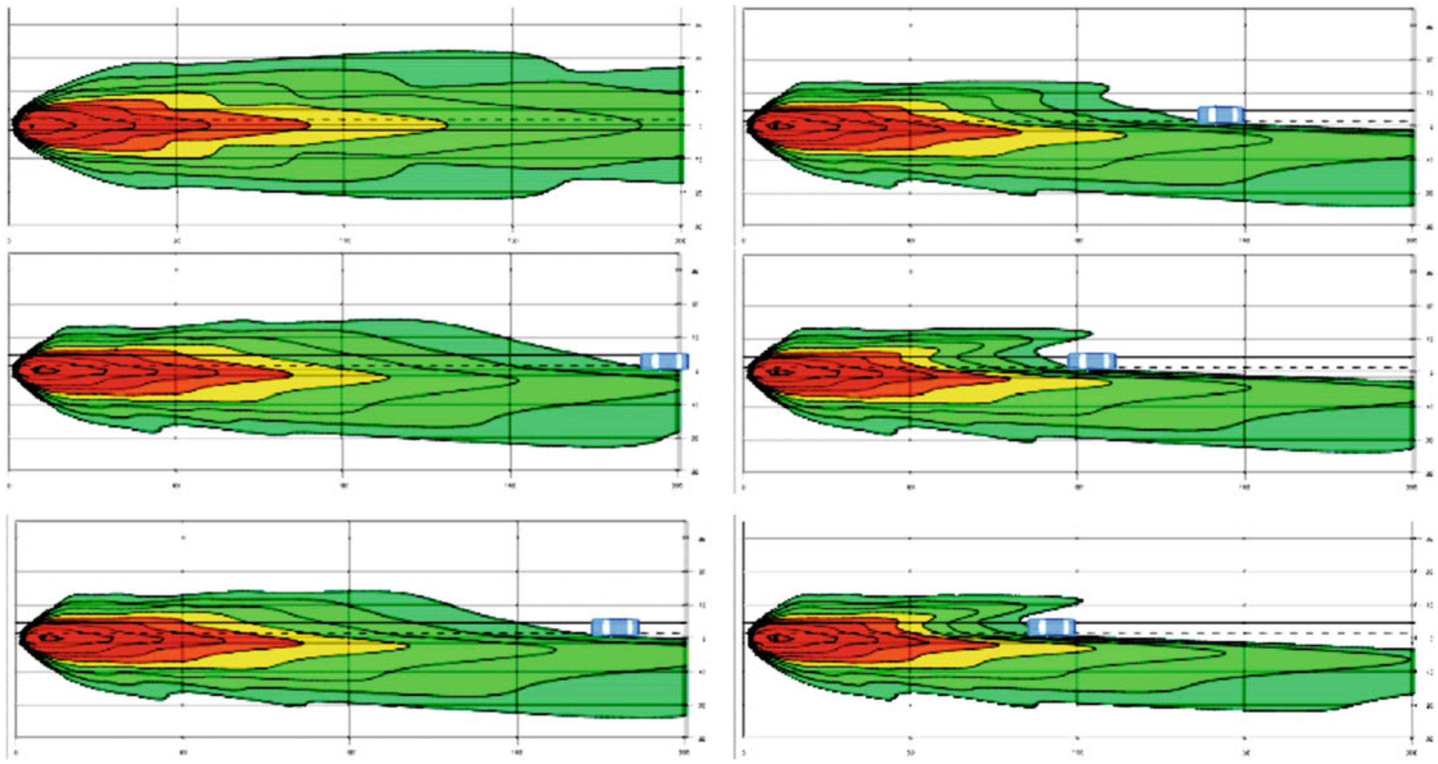
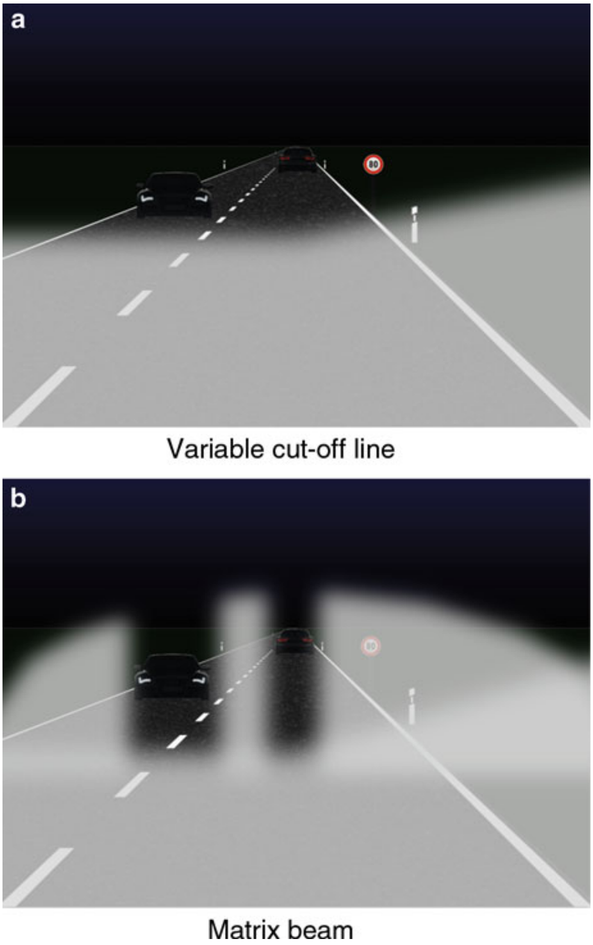


Fig. 20 Variation of light distribution on the road for oncoming traffic by the principle of variable cut-off line (Kalze and Schmidt 2007)

Fig. 21 Glare-free high beam with vertical light segment (matrix beam, **b**) in comparison to variable cut-off line (Kalze and Schmidt 2007; Totzauer 2013), **a**)



With the limited horizontal angular resolution and with only vertical light segments, the advantage of the high beam's maximal visibility distance can only be exploited in those light segments that are switched on. In the light segments which are switched off, only the limited visibility distance of maximal 88 m is possible, see Fig. 21 (right). From 2016 on, the whole head lamp will consist of an LED pixel array with a suitable optical system (micro lens optics) so that the whole traffic space can be divided into numerous horizontal and vertical pixels (or pixel groups). In the angular positions (with accurate angular resolution) in which other traffic participants are detected, luminous intensity can be reduced or switched off in order to avoid glare. At other pixels and angle positions, the visibility distance of high beam can be offered for the driver of the vehicle. The number of LED pixels on an LED array shall be increased along the technological time axis from 2016 to 2020 from 88 pixels to 256 pixels or more.

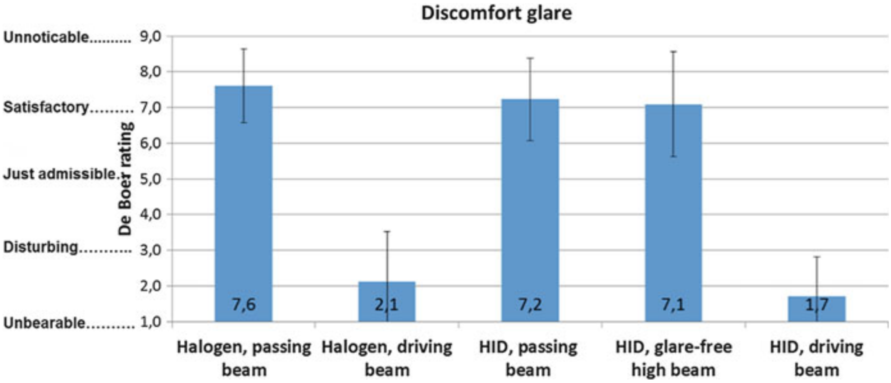


Fig. 22 Subjective glare evaluation in a field test (De Boer scale) for different front lighting systems with Xenon-HID lamps, tungsten halogen lamps and HID glare-free high beam (Zydek et al. 2013)

Principally, the glare-free high beam lighting function should evoke similar glare load like a low beam (if correctly aimed) and similar visibility distance like the best current xenon lamp high beam. This is the current aim of automotive lighting industry. Zydek et al. (2013) studied the performance of low beam and high beam as well as glare-free high beam systems in a series of dynamic field tests along the runway of the Griesheim Airfield (Germany). These experiments were conducted with test cars driven at 80 km/h. Test drivers had both a detection task and a driving task. Visibility distance results were shown in Figs. 7 and 8. In Fig. 22, the subjective glare evaluation of the participants of this study (Zydek et al. 2013) is shown.

The following conclusions can be drawn from Fig. 22 (Zydek et al. 2013):

(a) High beam produces disturbing to unbearable discomfort glare and this finding is independent of light source type;
(b) Comparing HID low beam and halogen low beam, it can be seen that there is almost no difference regarding the subjective ratings of discomfort glare;
(c) Glare-free high beam produces almost no increase of discomfort glare in comparison to low beam with tungsten halogen lamps and Xenon lamps.

In Kleinkes (2013), a comparison of different light-based functions was reported regarding visibility distance in the framework of a static test in a street with 45 test subjects. These results are listed in Table 3.

It can be summarized that the lighting systems to be used to improve visual conditions and visual performance at night can be developed in the future when the following technical principles are considered:

- Development of the light sources: the maximal visibility distance of the best Xenon lamp high beam and LED glare-free high beam is of the order of

Table 3 Visibility distance of different lighting functions according to Kleinkes (2013)

Light function	Mean visibility distance/m	Standard deviation/m
Low beam	85	14.3
Variable cut-off line	100	12.1
Glare-free high beam	130	13.0

145 m. This value can be increased with laser as a light source and with high-current white LEDs with more than 3 A;

- Intelligent and dynamic performance of headlamp systems with tungsten halogen lamps and Xenon lamps;
- Optimization of the LED light sources (optimization of blue chip structure, optimization of phosphor system and packaging technology) and optimization of the LED's optical and thermal properties;
- Fast and safe information processing on board and also regarding car-to-car-communication;
- Intelligent utilization and fusion of all sensor signals.

4 Conclusion

At night, the velocity of processing relevant visual information is low in the human subject and the risk of an accident is increased compared to daytime. For good object detection and recognition performance, the luminance level on the road and its surrounding is very important which can be ensured by the vehicle's lighting system. Otherwise dangerous objects cannot be detected by the driver. To avoid accidents, front lighting illumination systems shall be optimized and controlled in an intelligent and innovative way. To develop high-quality front lighting systems, advanced light source technologies, adaptive light distributions and so-called "light-based" lighting functions shall be used in order to achieve long visibility distances in all traffic situations. The following technological principles shall be considered: development of the light sources (e.g., LED high beam with high-current white LEDs); intelligent and dynamic performance of headlamp systems; car-to-car-communication and the intelligent utilization and fusion of all sensor signals.

References

Eckert M (1993) Lichttechnik und optische Wahrnehmungssicherheit im Straßenverkehr. Verlag Technik, Berlin/München

Fratty H et al (2014) Automotive lighting technologies. Trends and Vision 2025, DVN

Grimm M, Casenave S (2007) DBL: a feature that adds safety to night time traffic. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 355–363

Kalze FJ, Schmidt C (2007) Dynamic cut-off-line geometry as the next step in forward lighting beyond AFS. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 346–354

Khanh TQ (2013) Grundlagenvorlesungen der Lichttechnik. Technische Universität Darmstadt, Fachgebiet Lichttechnik, Darmstadt, Germany

Kleinkes M (2013) New automotive lighting technology: benefit or Mayfly? In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 361–366

Kleinkes M, Eichhorn K, Schiermeister N (2007) LED technology in headlamps – extend lighting functions and new styling possibilities. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 55–63

Langwieder K, Bäumler H (1997) Characteristics of nighttime accidents. In: Progress in automobile lighting (PAL), Technische Universität Darmstadt, Utz Verlag, München, pp 326–339

Lerner M, Albrecht M, Evers C (2005) Das Unfallgeschehen bei Nacht, Bericht der Bundesanstalt für Straßenwesen (BASt), Heft M 172

Rosenhahn EO (2007) AFS-frontlighting on the basis of LED light sources. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 80–87

Rosenhahn EO, Hamm M (2003) Motorway light in adaptive lighting systems. In: Progress in automobile lighting (PAL), Technische Universität Darmstadt, Utz Verlag, München, pp 868–882

Rumar K (2001) Night traffic and the zero vision. In: Progress in automobile lighting (PAL), Technische Universität Darmstadt, Utz Verlag, München, pp 849–858

Schiller C (2007) Lichttechnische Tests an ersten LED-Scheinwerfer-Autos. Technische Universität Darmstadt, interner Bericht des Fachgebietes Lichttechnik

Schiller C, Khanh TQ (2007) First field tests of cars with completely built-In LED headlamps under realistic driving conditions. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 131–138

Schiller C, Khanh TQ (2008) Psychologische Blendung mit Xenon- und Halogenscheinwerfer-Autos – Ergebnisse realer Tests, Zeitschrift Verkehrsunfall und Fahrzeugtechnik. Vieweg Verlag, Heft 9, München

Schneider D (2011a) Marking light- safety enhancement by marking light systems and their technical implementation. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 320–326

Schneider D (2011b) Markierungslicht-eine Scheinwerferverteilung zur Aufmerksamkeitssteuerung und Wahrnehmungssteigerung von Fahrzeugführern. Dissertation, Technische Universität Darmstadt

Sprute H (2012) Entwicklung lichttechnischer Kriterien zur Blendungsminimierung von adaptiven Fernlichtsystemen. Technische Universität Darmstadt, Darmstadt

Sprute JH, Khanh TQ (2007) Approval requirements for a front-lighting-system with variable cut-off line in Europe. In: International symposium on automotive lighting (ISAL), Technische Universität Darmstadt, Utz Verlag, München, pp 31–37

Totzauer A (2013) Kalibrierung und Wahrnehmung von blendfreiem LED-Fernlicht. Dissertation, Technische Universität Darmstadt

UN/ECE Regulation No. 123 (2007) Uniform provisions concerning the approval of adaptive front-lighting systems (AFS) for motor vehicles

Zydek B, Schiller C, Wagner M, Polin D, Khanh TQ (2013) Performance evaluation of vehicles with adaptive headlamp systems. In: ISAL – international symposium on automotive lighting, Darmstadt

Parking Assistance

44

Heinrich Gotzig

Contents

Abstract

For many people parking a car is more and more a chore task. Due to aerodynamic requirements, increasing the size of vehicles, the visibility to rear and front end of the vehicle decreases, and in parallel, it is more and more difficult to find a suitable parking space. While in the 1990s Parking Aid Systems were not seen as necessary in the meantime, systems supporting the parking process are either standard in some vehicles or have a high take rate. While Parking Aid Systems of the first generations were mainly informing systems, nowadays, these systems help in finding a suitable parking place and support steering during the parking maneuver, and future systems will be more and more autonomous, finally resulting in systems which find their parking place by themselves – valet parking.

H. Gotzig (✉)
Valeo Schalter und Sensoren GmbH, Driving Assistance Product Group, Bietigheim-Bissingen, Germany
e-mail: heinrich.gotzig@valeo.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_45

For many people parking a car is more and more a chore task. Due to aerodynamic requirements, increasing the size of vehicles, the visibility to rear and front end of the vehicle decreases, and in parallel, it is more and more difficult to find a suitable parking space. While in the 1990s Parking Aid Systems were not seen as necessary in the meantime, systems supporting the parking process are either standard in some vehicles or have a high take rate. While Parking Aid Systems of the first generations were mainly informing systems, nowadays, these systems help in finding a suitable parking place and support steering during the parking maneuver, and future systems will be more and more autonomous, finally resulting in systems which find their parking place by themselves – valet parking.

International standards for Parking Aid System are described by MALSO (Maneuvering Aid for Low Speed Operation) (ISO 17386 2010), ERBA (Extended Range Back Up Aid) (ISO 22840 2010), and APS (Assisted Parking Systems) (ISO/DIS 16787 under development).

Figure 1 shows basic operation of a Parking Aid System: Support the driver with information on relevant obstacles in the rear and front of his vehicle during parking maneuver.

1 Versions of Parking Assistance Systems

To support the driver during parking maneuvers, there are several technical solutions possible:

- Parking Aid Systems which provide distance information to relevant obstacles in the front and/or rear of the vehicle
- Parking Aid Systems which provide either a camera view to the rear of the vehicle or a camera view from the top of the vehicle (bird's-eye view)

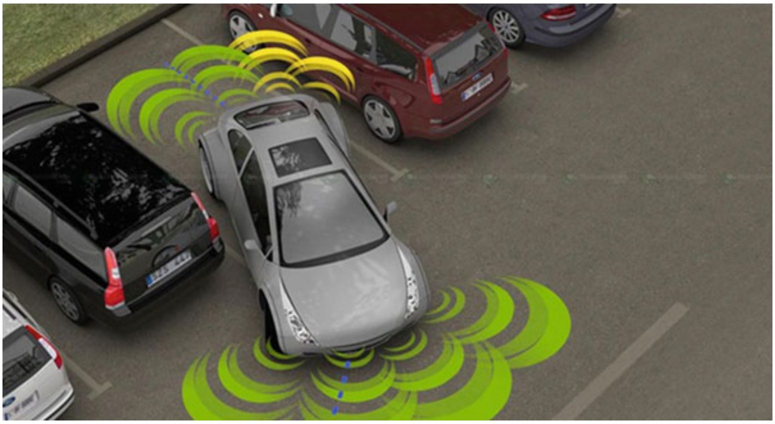


Fig. 1 Parking Aid System

Fig. 2 Versions of Parking Aid System

- Systems which provide in addition to distance or image information on the length of the parking slot relative to the length of the owned car
- Guided Parking Aid Systems which either provide steering instructions to the driver or fully take over the lateral control of the vehicle during the parking maneuver
- Remote Parking Aid Systems, either with the driver in the car or driver outside of the car

Figure 2 shows some examples of Parking Aid Systems. Park Assist provides distance informations to the driver, Park Fit provides information on the length of the parking slot relative to the length of the car, UPark™ gives steering instructions to the driver during the parking maneuver, and Park4U™ steers automatically during the parking maneuver. Park Assist, ParkFit, UPark™, and Park4U™ are Valeo trademarks.

2 Requirements to Parking Aid Systems

Depending on system configuration and degree of support, there are different requirements for sensing the environment and signal processing. System needs to be user friendly. This means that driver information needs to be in such a way that system is seen as support. In pure parking mode only relevant information are provided. This requires sophisticated signal processing as typically parking takes place in a very noisy environment (e.g., street sweeper generates high noise level). Reaction time has to be within a few 100 ms, and when the car is in parking slot search mode, the driving speed should be less than 30 km/h.

Regarding environmental sensor technology, the following requirements apply:

- Full coverage of the detection field (as defined in (ISO 17386 2010)) (Fig. 3)
- Robust against environmental impact (rain, snow, ice, mud, etc.)
- High resolution and accuracy of distance and parking slot measurement (high detection rate only of suitable parking slots)
- Short response time
- Low total system cost
- Small dimensions

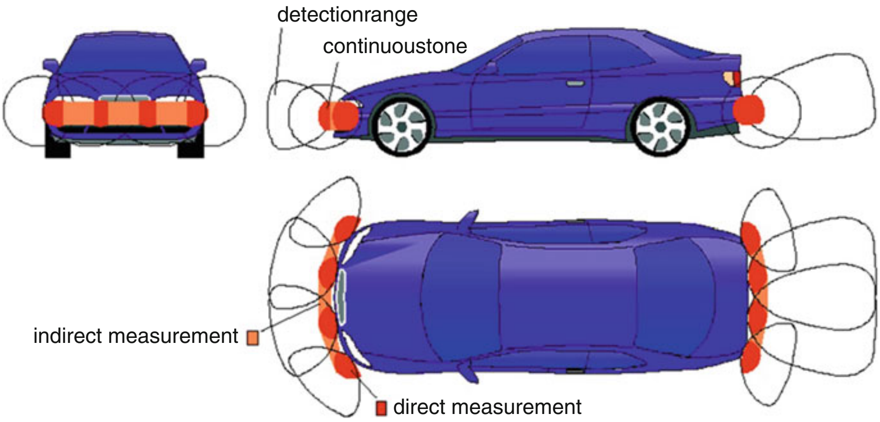


Fig. 3 Requirement to detection field

With guided or (semi)autonomous parking systems, additional requirements need to be considered:

- The trajectory needs to be similar to the one of a human driver in order to get acceptance.
- The trajectory needs to be collision-free.
- During the parking maneuver the driver needs to get warnings on relevant obstacles (park aid functionality).
- End position of the vehicle in the parking slot needs to be fitted; this implies distance measurement to curb and alignment to cars in the rear and front.
- Short duration of parking maneuver.
- Simple and customized human-machine interface.

3 Parking Aid System Architecture

Figure 4 shows a typical system architecture of a Parking Assist System (rear and front):

- Four ultrasonic sensors per bumper → sensing the environment in the rear/front
- Electronic control unit (ECU) → sensor signal processing, calculation of distance to relevant obstacles, communication with car network, and diagnostic
- HMI, in this example one rear and one front sounder → driver information

4 Technical Implementation

Parking Assistance Systems either provides information to the driver, give guidance for steering operations, or take over longitudinal and/or lateral control of the vehicle during the parking maneuver. The technologies which are used are either ranging sensors

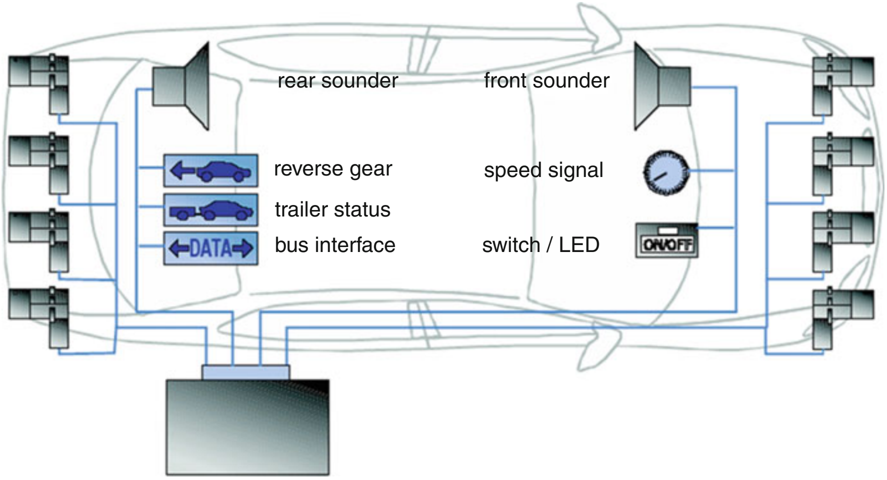


Fig. 4 Parking Aid System architecture

Fig. 5 Rear view camera system

(ultrasonic, RADAR sensors) or vision-based sensors (camera systems with opening angles between 130 and 190° and resolution between 320 × 240 and megapixel) (Fig. 5).

4.1 Informing Parking Systems

Parking Aid Systems were introduced into the market in 1991 (BMW E32 PDC, 1991). This Parking Aid System was based on ultrasonic technology. Depending on

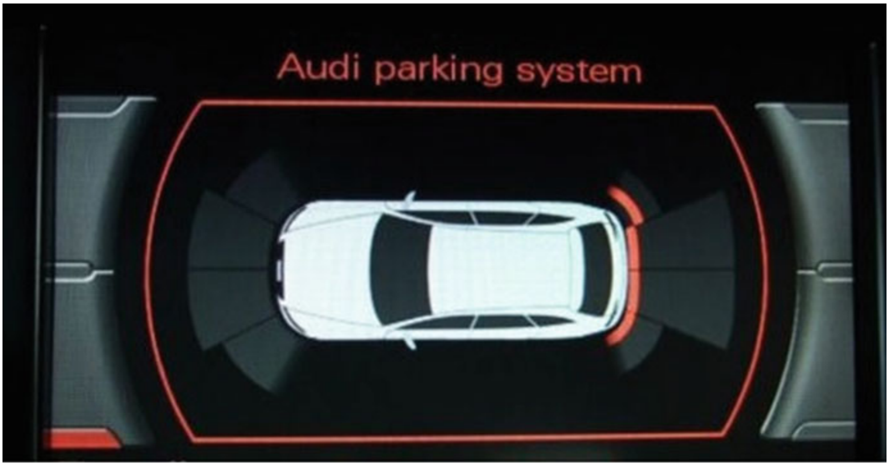


Fig. 6 Example of parking aid in an Audi with additional optical distance information

the specifications, cars with such system are equipped between 2 (only on the rear corner) and 12 sensors (6 per bumper). The driver receives information on the closest object relative to his car, which is within the detection field. The information is mainly provided by an acoustical beep tone. The shorter the time between the pulses, the closer the distance. Continuous tone typically means a distance of 0.3 m or below. To distinguish between front and rear, frequency of the tone for rear is below 1 kHz and front in the range of 1.5 kHz. The auditive information can be supplemented by visual distance information (Fig. 6).

Ultrasonic technology for distance measurement is the preferred technology as it is rather cheap and robust against almost all environmental influences (works even with heavy wind) and a proven technology. Even if the technology can be used up to a max speed of approximately 130 km/h, Parking Aid Systems normally operate in a range below 10 km/h. High speed is only possible if the detected obstacles are not in the driving path (ultrasonic-based blind spot and parking slot measurement). The reason is the speed of sound: 343 m/s at 20 °C. With a worst case system reaction time of 500 ms and a speed of the car of 10 km/h, the driven distance is 1.39 m, which is for worst cases the maximum detection range. Typically speed while parking is much lower (2–3 km/h), so driver can react and brake.

Other technologies like ultra-wideband short-range sensors are only used on very few cars. Frequency homologation which results in a limitation to max 7 % penetration rate per country (European Commission 2014; Tristant 2009) and system costs do not justify the improved performance (versus ultrasonics), namely, reaction time, detection range, and additional information on speed of the detected obstacles. In the past one additional argument for this technology, namely, hidden integration, is in the meantime also possible for ultrasonic sensors.

Other informational parking systems are applications which provide information on the relative length of a parking slot (Fig. 2, UPark™). While passing parking

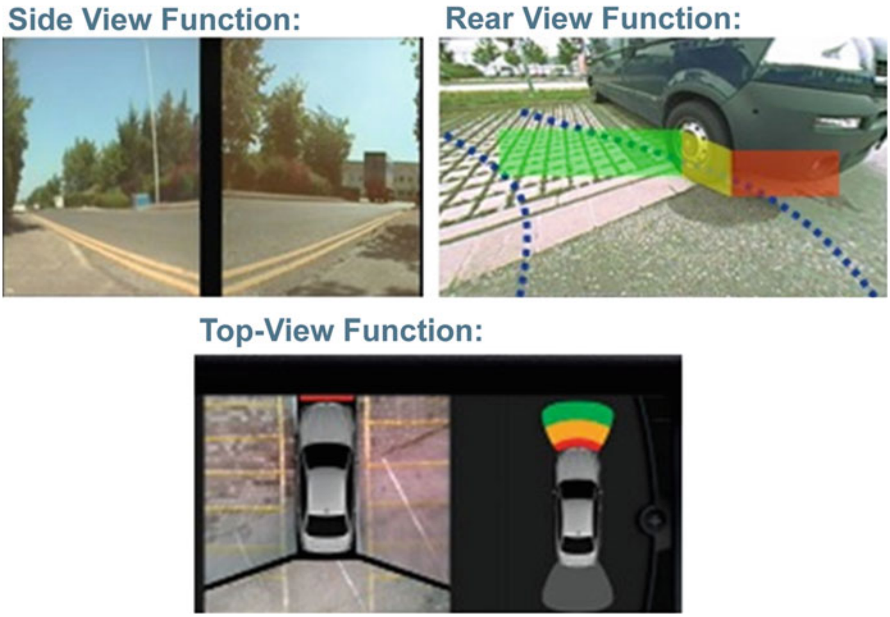


Fig. 7 Examples of viewing systems: side, rear, and top view system

slots, the system measures their length and compare it with the length of the own vehicle. Information provided to the driver are: (1) slot is big enough for parking, (2) slot is only a little bit bigger than own car, parking is possible but difficult, and (3) slot length is too small for parking.

Besides ultrasonic-based Parking Aid Systems, there are camera-based systems on the market, so-called rear, top, front, and surround view systems (Figs. 5 and 7).

Side view means that the driver gets a view to the front side of a vehicle which helps in situations when he is entering into a street but the visibility is obstructed by parked cars. Rear view systems provide a view to the rear of the car. In the market one can find two types of rear view systems. To cover the full area behind the vehicle, one has to use lenses with a big opening angle up to 190°. This results in a fish-eye effect. Very simple rear view systems only provide the image recorded by the camera; more intelligent rear view systems perform a distortion correction. In Fig. 8 one can see the difference with/without correction. Even with correction the estimation of the distance is very difficult for human people.

Top view system provides a bird's-eye view from the vehicle. Images from four (rear, side left, side right, front) or five (rear, side left, side right, front left, front right) wide-angle cameras are processed and merged to a picture which is shown to the driver (Fig. 7).

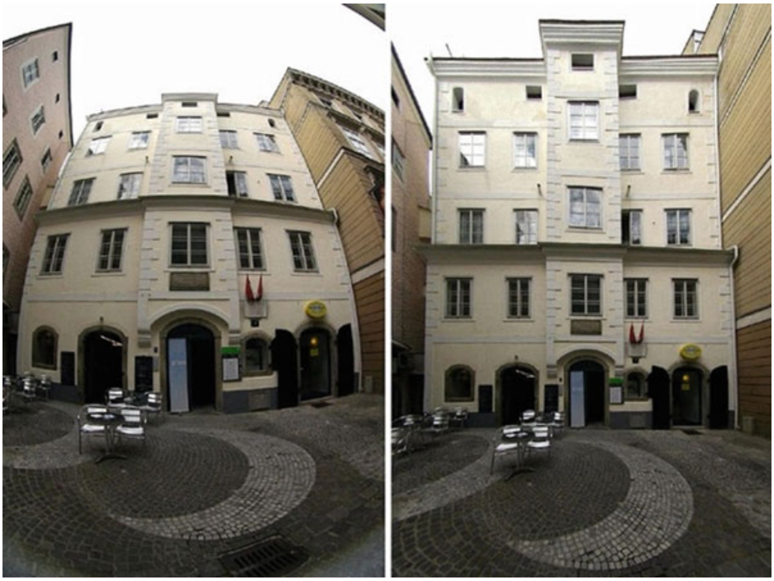


Fig. 8 Rear view system with or without distortion correction

4.2 Guided Parking Systems

Besides parking systems which provide only information to the driver, there are systems which give guidance and additional supporting information to the driver. Such guidance can be achieved by steering instructions during the parking maneuver. The system performs a parking slot measurement as described in Sect. 3, and when a suitable slot is found and the guidance function is activated, it informs the driver in which direction he has to drive for a stress-free parking. Mainly these systems are used when electrical power steering is not available.

Standard parking systems just provide either distance information or a camera view. With guided parking systems information from the steering angle sensor are either used to adopt the detection field or to display the driving path into the image provided to the driver. When a driver is, for example, backing into a parking slot, he gets guidance how to steer correctly avoiding any additional movement and too close distance to parked cars.

Parking systems with a high-performance rear view camera may also provide trailer hitch support. Here the driver gets guidance and sees how he has to steer in order to dock correctly (Fig. 9).

4.3 Semiautomatic Parking Systems

Even if the informing or guided parking systems give helpful guidance to the driver like driving forward, backward is possible, or you have to steer in a certain

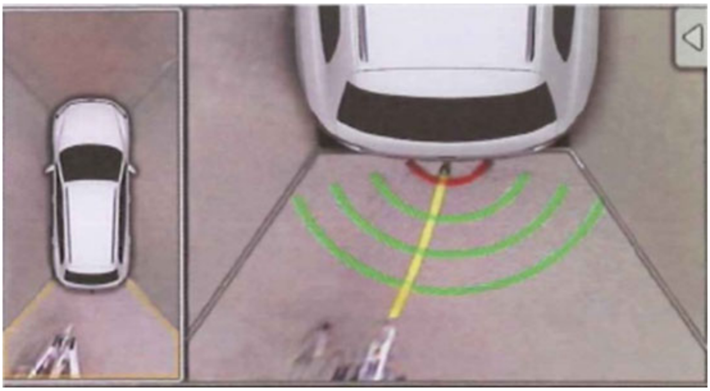


Fig. 9 Principle of semiautomatic parking system

direction, there is no active support by the system. When backing into a parking slot, a lot of drivers have difficulties in steering, so these maneuvers can result in annoying situations which take a lot of time. With semiautomatic parking systems, steering during the parking maneuver is done by the system (Fig. 10).

Nearly all semiautomatic parking systems in the market are either based on ultrasonic, ultrasonic + camera, or camera systems only.

The minimum performance and basic system description of such system are described by ISO/DIS 16787 (final release expected in 2016). Legal requirements regarding steering are defined by ECE-R 79 “Automatically commanded steering [. . .] shall be automatically disabled if the vehicle exceeds the set limit of 10 km/h by more than 20 % or [. . .].” (Gasser 2013).

4.3.1 System Description

Figure 11 shows a typical system description of semiautomatic parking system which needs the following components:

- Electronic control unit (ECU) for signal processing, trajectory planning, steering interface, HMI, and error handling
- ECU of steering system
- Button to activate the system
- Input from wheel pulses, acceleration, steering angle, yaw rate, steering torque, speed of vehicle, gear and trailer status, ambient temperature, and turn signal
- Distance information from parking slot measurement and parking aid sensors
- Rear and front speaker for driver information

4.3.2 Application of Semiautomatic Parking Systems

Principally semiautomatic parking systems need to do three different tasks: slot measurement, calculation of the trajectory, and steering. These tasks are realized by four SW modules: (1) slot measurement, (2) odometry, (3) path calculation, and

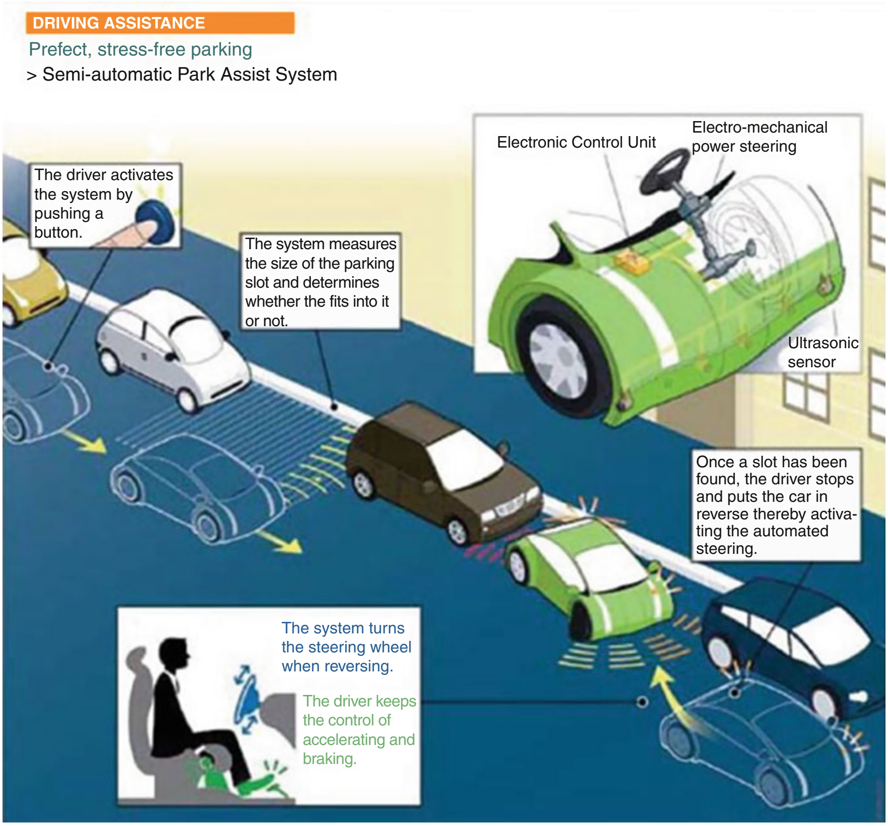


Fig. 10 Principle of semiautomatic parking system

(4) steering interface (Fig. 12). Besides detection of suitable parking slots, the main target is correct end position within the parking slot. Correct end position means basically well oriented between the boundaries (typically a car in the front and rear). Factors are distance to curb and alignment inside the parking slot. For path calculation the following information is necessary: (1) slot length, (2) slot depth, (3) orientation of the car relative to the parking slot, and (4) position of the car relative to the parking slot. These values need to be very precise as typically an inaccuracy of orientation by 1° results in an offset of $\approx$10 cm.

The main difficulty regarding measurement of slot length is the shape of boundary obstacles. Figure 13 shows two examples of cars with different bumper shape. By analyzing the pathway of the detected distance information, algorithms are able to determine the correct length.

Table 1 shows the required vehicle data for correct distance measurements, odometry, and path calculation.

The most critical is the exact knowledge of tire diameter as it shows a big variation: mileage, pressure, different types, and summer/winter tire.

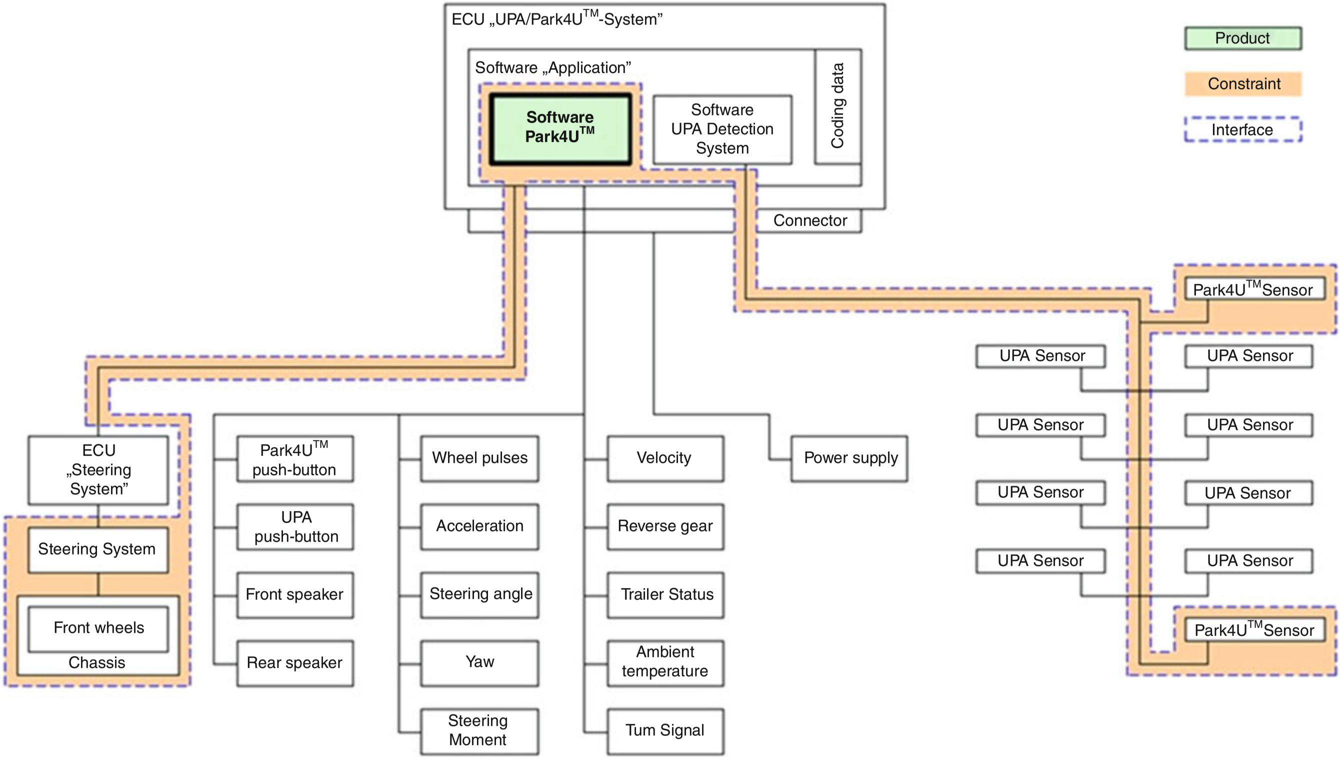


Fig. 11 System description of semiautomatic parking system (ultrasonic based)

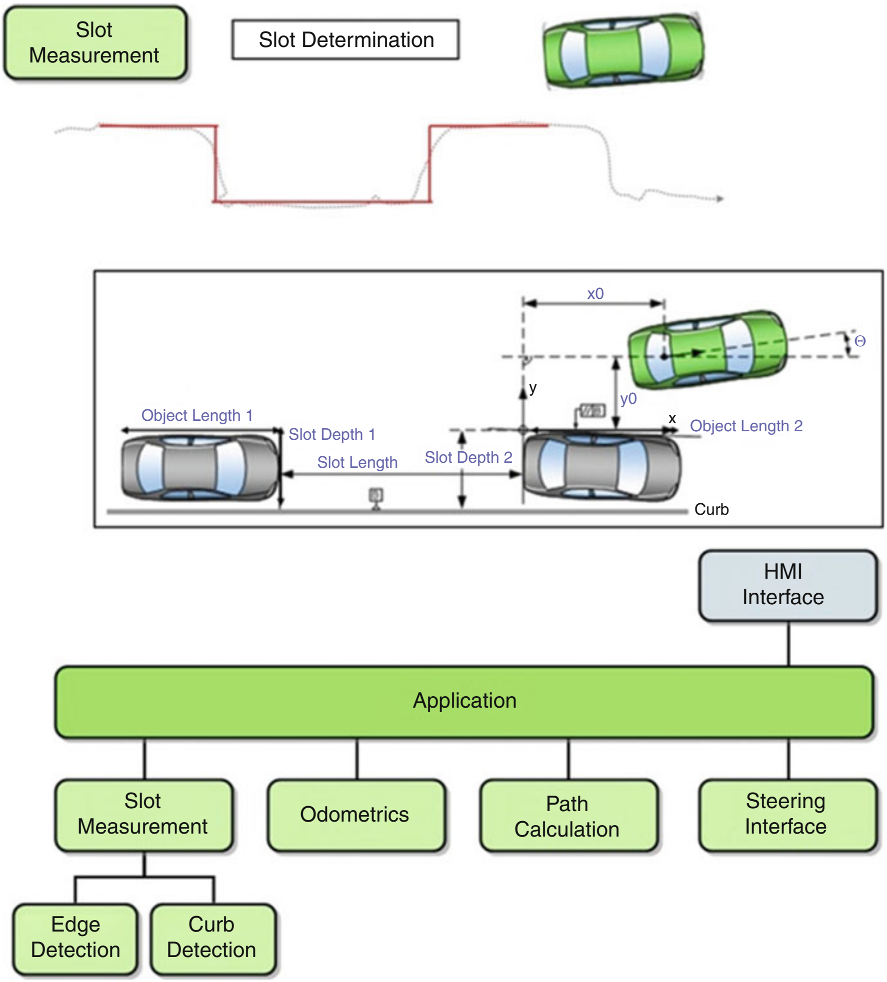


Fig. 12 Park slot measurement and SW architecture of semiautomatic parking systems

First semiautomatic parking systems were introduced in 2003 into the market (Wikipedia 2015a, b). A car could steer with a little input from the driver. Since 2007 full active semiautomatic systems are in production. While the first generation offered only parallel parking, driving backward in one move in the following years on the one hand performance was continuously improved, and as well as the system could be used for perpendicular parking, park-me out, and within some limitations, the system also provides braking support (Fig. 14). Further developments are ongoing. Recent systems use an environmental map which evaluates input from different sensors and vehicle data and then shows a much higher information content realized by data sensor fusion. This enables, for example, parking even in very small parking slots, parking relative to lines on the road, etc.

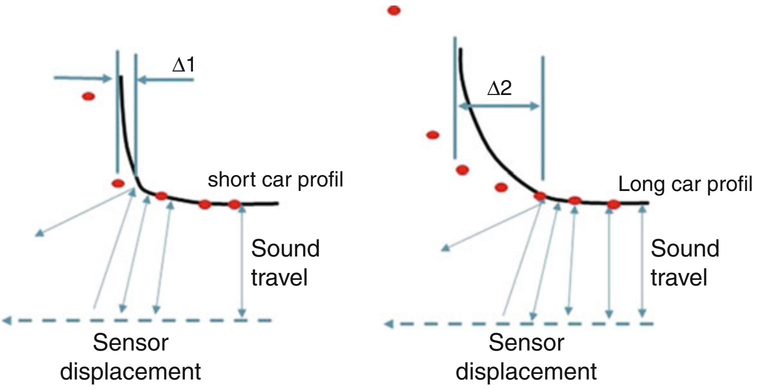


Fig. 13 Data analysis measurement of slot length

Table 1 Required vehicle data for semiautomated parking systems

Requirement		Update (ms)
Wheel pulses	25 mm/pulse	20
Vehicle speed	0.1 km/h	20
Steering angle	0.5°	20
Driving direction	True at any pulse	20
Longitudinal acceleration rate	0.05 m/s^2	20
Yaw rate	0.5°/s^2	20
Steering moment	0.1 Nm	20
Outside temperature	0.1 °C	100
ESP/ABS intervention	–	100
Reverse gear	–	100
Turn signal	–	100
Trailer status	–	100

5 Outlook

Already in 1990 a fully automated parking car was presented (Walzer and Grove 1990). But even if the necessary technology is now available, "steering, braking, acceleration can be electronically controlled," there is no fully automated parking car available in the market, only several research projects (Institute of Measurement, Control and Microtechnology 2015). In the near future there will be only one system available with remote parking and automatization, level 2 "Partial Automation" (Walker Smith 2013). This means that the driver still monitors the driving environment (Fig. 15). Here it is realized by a bird's-eye view of the remote parking

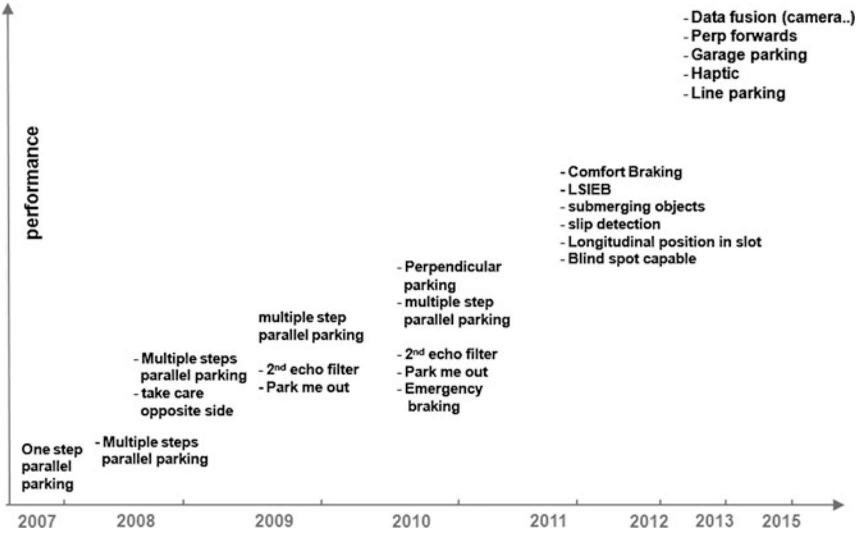


Fig. 14 Evolution of semiautomatic parking systems

Fig. 15 Remote parking – level 2 automatization

car. During the maneuver the driver (or operator) has to observe the situation, and longitudinal movement is only possible if, e.g., the driver press a finger on the connected smartphone.

The next step, "level 3 automatization," would mean that the system performs all necessary driving tasks, is able to detect all errors, and informs the driver appropriately. These are applications like valet parking (BMW AG 2014), where the driver leaves his car, for example, at the entrance of a parking garage and the car searches a slot and parks by himself. When the driver comes back, he calls his car and the car drives automatically to his position. Valet parking combined with information on available parking will have a big impact on traffic flow especially in cities. Studies show that the part of the urban traffic caused by vehicles searching for a parking space is situated between 5 % and 10 % in cities and can reach 60 % in small streets (Gantelet and Lefauconnier 2006). But from a product liability point of view, valet parking is not yet possible today, even if several companies are working to make it ready for the market. Further improvements in sensor technologies, signal processing, and connectivity will more and more improve parking systems toward low-speed maneuvering.

References

BMW AG (2014) BMW Innovations at the 2015 Consumer Electronics Show (CES) in Las Vegas. 360-degree collision avoidance and fully-automated parking in multi-storey car parks. https://www.press.bmwgroup.com/global/pressDetail.html?title=bmw-innovations-at-the-2015-consumer-electronics-show-ces-in-las-vegas-360-degree-collision-avoidance&id=T0198231EN. Accessed 30 June 2015

European Commission (2014) A radar for your car. https://ec.europa.eu/digital-agenda/en/radar-your-car. Accessed 30 June 2015

Gantelet E, Lefauconnier A (2006) The time looking for a parking space: strategies, associated nuisances and stakes of parking management in France. http://abstracts.aetransport.org/paper/download/id/2373. Accessed 30 June 2015

Gasser T (2013) Legal aspects of active safety systems. http://www.interactive-ip.eu/dyn/mediaout.dhtml/-/mime/FEpresentations/TomGasserBASt/4.interactIVe_20130909-v13-FE_SP7_session_Tom_gasser_v2.pdf. Accessed 30 June 2015

Institute of Measurement, Control and Microtechnology (2015) Autonomous driving Uni Ulm. https://www.uni-ulm.de/in/automatisiertes-fahren. Accessed 30 June 2015

ISO 17386 (2010) Transport information and control systems – Manoeuvring Aids for Low Speed Operation (MALSO) – performance requirements and test, procedures. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=51448. Accessed 30 June 2015

ISO 22840 (2010) Intelligent transport systems – devices to aid reverse manoeuvres – Extended-Range Backing Aid Systems (ERBA). http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=36521. Accessed 30 June 2015

ISO/CD 16787 (under development) Intelligent transport systems – Assisted Parking System (APS) – performance requirements and test procedures. http://www.iso.org/iso/catalogue_detail.htm?csnumber=63626&utm_source=ISO&utm_medium=RSS&utm_campaign=Catalogue. Accessed 30 June 2015

Tristant P (2009) Automotive 24 GHz UWB short-range radars. http://www.itu.int/md/dologin_md.asp?id=R09-SEM.WMO-C-0022!!PDF-E. Accessed 30 June 2015

Walker Smith B (2013) SAE levels of driving automation. http://cyberlaw.stanford.edu/blog/2013/12/sae-levels-driving-automation. Accessed 30 June 2015

Walzer P, Grove HW (1990) IRVW futura the Volkswagen research car. http://papers.sae.org/901751/. Accessed 30 June 2015

Wikipedia (2015a) Intelligent parking assist system. https://en.wikipedia.org/wiki/Intelligent_Parking_Assist_System. Accessed 30 June 2015

Wikipedia (2015b) Automatic parking. https://en.wikipedia.org/wiki/Automatic_parking. Accessed 30 June 2015

Adaptive Cruise Control

45

Hermann Winner and Michael Schopper

Contents

H. Winner (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

M. Schopper
Daimler AG, Sindelfingen, Germany
e-mail: michael.schopper@daimler.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_46

Abstract

Adaptive cruise control (ACC) has reached a new quality in driver assistance. For the first time, a large part of the driver's tasks can be assigned to an automatic system and the driver relieved to a substantial degree. Based on cruise control, ACC adjusts the vehicle speed to the surrounding traffic. It accelerates and decelerates automatically when a preceding vehicle is traveling at less than the speed desired by the driver.

ACC is a key functional innovation and represents a new system architecture with a high degree of function distribution. The different operating modes and system states are described along with function limits and transition conditions.

From the many elements of this overall function, target selection and longitudinal control are addressed in detail because of the special challenges they present. Target selection is based on the actual road curvature being determined by the ESC sensor signals that describe the driving dynamics, of which several options are assessed. Predicting and selecting a suitably shaped corridor is explained using an example. Major sources of error for the individual steps, their severity, and possible countermeasures are described.

The prerequisite for vehicle-following-distance control is the selection of a target. An example shows that the basic control principle is simple, but it conflicts with comfort and convoy stability. Details of additional control functions in curve situations and approaches are provided.

The driver perspective is addressed in terms of control and display functions and in terms of satisfaction as ascertained by use and acceptance studies, also taking into account an extended driver familiarization phase.

1 Introduction

Adaptive cruise control, abbreviated to ACC, describes a method of vehicle speed control which adapts to the traffic situation. Active cruise control, automatic distance control, automatic cruise control, or autonomous intelligent cruise control tends to be used as synonyms. DISTRONIC and automatic distance control (ADR, Automatische Distanz-Regelung) are registered trademarks.

The relevant international standards are ISO 15622 (transport information and control systems, adaptive cruise control systems, performance requirements and test procedures) (ISO TC204/WG14 2010) and ISO 22179 (intelligent transport systems, full-speed-range adaptive cruise control (FSRA) systems, performance requirements and test procedures) (ISO TC204/WG14 2009), with the former describing the first functionality, often referred to as the standard ACC, while the second describes an extension of the functionality for the low-speed range, known as a full-speed-range ACC.

In ISO 15622 (ISO TC204/WG14 2010), the ACC function is described as follows:

> An enhancement to conventional cruise control systems, which allows the subject vehicle to follow a forward vehicle at an appropriate distance by controlling the engine and / or power train and potentially the brake.

ACC is derived from the long-standing driving speed control, referred to in English-speaking countries as cruise control which is widely used in North America and Japan (abbreviated to CC) or commonly in German-speaking countries as “tempomat.” Its role is to control a desired speed v_{set} set by the driver, and it is included as part of the ACC function (Fig. 1 top).

The main extension concerns adjusting the speed to the speed of the immediately preceding vehicle, here in addition to v_{to} (to: target object identified by ACC as target for control) (Fig. 1 center).

Although ISO 15622 leaves it open as to whether the brake is used for the control, the application of the brake to increase the deceleration has become established as a de facto standard. The appropriate distance mentioned in this standard is determined by τ, the time gap that is often colloquially referred to as distance in seconds. It is defined as:

> Time gap τ: “Time interval for travelling a distance, which is the clearance d between consecutive vehicles. Time gap is related to vehicle speed v and clearance d by: $\tau = d/v$.”

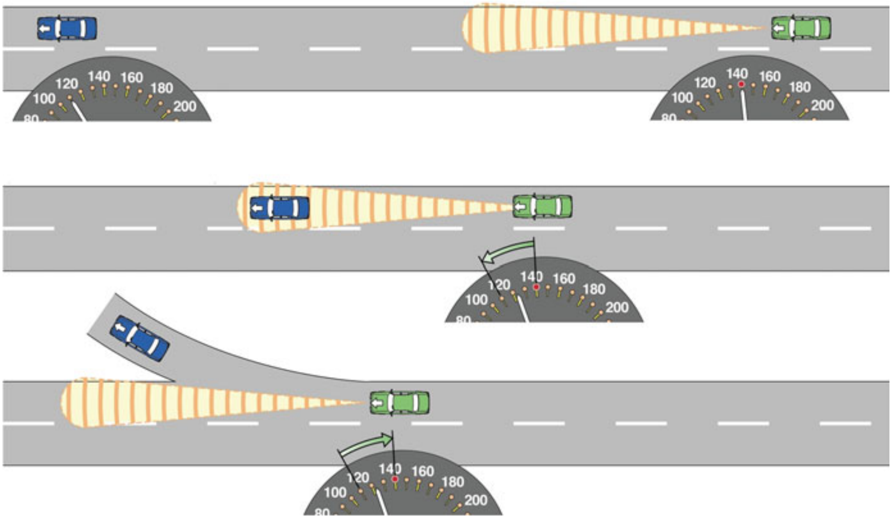


Fig. 1 Situation-adapted change from free driving to following and back (Source: BOSCH)

The use of the temporal rather than the spatial reference follows the basic idea that to prevent rear collision, it is sufficient to have a distance which is in accordance with the reaction time, assuming the same deceleration capability for both the preceding and the subject vehicle. Therefore, in the presence of a preceding vehicle which is moving slower than one's own desired speed, the control task of the ACC is to adapt one's own speed to that of the preceding vehicle to obtain compliance with a clearance that ensures a constant reaction time.

However, as soon as the target leaves the immediate driving corridor and no other vehicle is designated as the target, ACC restores the set speed without further action by the driver (Fig. 1).

2 A Look Back at the Development of ACC

A prototype description of this function was first documented in 1980 in Ackermann (1980). It was the result of a research project conducted in the 1970s, in which several firms collaborated and also competed to develop RADAR sensors in the frequency range of 35 GHz, which was the range technically possible at that time. However, because of their technical performance, size, and manufacturing costs, they were not suitable for series application. After a hiatus which lasted until the end of the 1980s, the European PROMETHEUS (PROgramMe for a European Traffic with Highest Efficiency and Unprecedented

Safety) project, which ran from 1986 to 1994, again stimulated the development of system functionality and also of sensor systems. This program gave rise to the designation AICC (autonomous intelligent cruise control), which was also the title of a Common European Demonstrator Project (CED5).

Two additional developments were crucial to the market launch of ACC: installation of exhaust emission systems, which had become necessary under the EURO III emission regulations, and introduction of ESC onto the market. ESC meant that the yaw-rate sensor was available for detecting curves (see Sect. 7.1), and the active buildup of brake pressure, together with E-Gas or the diesel engine equivalent, could provide electronic diesel control and speed control at practically no extra cost.

Despite the momentum provided by PROMETHEUS or any of the factors mentioned above, the first systems were introduced outside Europe. Mitsubishi launched ACC in its Diamante model as early as 1995 (Watanabe et al. 1995), followed by Toyota 1 year later. Common to both of them was that they did not include a braking intervention, and they both used laser-scanner-based LIDAR sensors. While the system installed by Mitsubishi was regarded as a prototype, the Toyota system, equipped with a LIDAR made by Denso (Furui et al. 1998), was a real series system and sold in larger numbers. In Europe, we had to wait until 1999 before ACC was available to buy. These systems had to be much more elaborate to satisfy European customers. They had to include a braking intervention, in order to take account of the greater speed differences on German motorways, a higher maximum set speed $v_{\mathrm{set,max}}$, and a mm-wave RADAR sensor that was still very reliable even in bad weather.

The first system with a braking intervention and mm-wave RADAR was introduced in the Mercedes-Benz S-Class with an ADC RADAR sensor. This was followed by systems from Jaguar in the XKR with a Delphi RADAR sensor and, 1 year later, by BMW in its seven series with a BOSCH sensor and control unit. Since then, RADAR-sensor-based ACC systems have dominated the European market, while in Japan LIDAR sensors (e.g., made by Omron) continued to be used for a long time. For further details about the history of ACC development up until 2003, please refer to the conference paper (Winner 2003).

Although, at the moment, it looks as if RADAR has won the competition between LIDAR and RADAR, which has been ongoing for more than 20 years, LIDAR is still being developed and is expected to be useful for future applications (see ▶ Chap. 18, “Automotive LIDAR”). Essentially, both principles are suitable for ACC, even if there are still differences in certain areas.

The market success of ACC is still lagging far behind expectations. Now that a wide range is available and costs have come down significantly, ACC is also becoming big business. Another factor that contributes to its success is the functional extension to use in traffic congestion offered in the Mercedes-Benz S-Class (W 221) since 2005, especially in more high-end vehicles, which traditionally include a high proportion of automatic gearboxes.

3 Requirements

3.1 Functional Requirements for Standard ACC Pursuant to ISO 15622

The function definitions in Sect. 1 give rise to the following functional requirements:

- Under free cruising conditions:
    - Constant speed control and high control comfort, i.e., minimal longitudinal jerk and no swinging but high control quality (with no obvious deviation from the set speed)
    - Cruise control with brake intervention in case of a lowered desired speed or incline travel
- When following another vehicle:
    - Vehicle-following control with swinging damping adjustment to the speed of the vehicle ahead so that the speed fluctuations are not copied
    - Time-gap control to maintain the set time gap τ_{set} and gradual "falling back" when the interval is greatly shortened by vehicles cutting in, in line with standard driving behavior
    - Control with the dynamics expected by the driver
    - Convoy stability of the control when following other ACC vehicles
    - Adequate acceleration capability for dynamic following
    - Ability to decelerate for the majority of pursuit driving situations (90 %) in moving traffic
    - Automatic target detection when approaching cut-in or cut-out situations within a defined distance range, i.e., determination of a target-seeking corridor
- When approaching:
    - For slow approaches, prompt speed control to the desired distance.
    - For faster approaching, predictable deceleration course in order to facilitate an assessment by the driver of whether to intervene because of inadequate ACC deceleration.
    - If the vehicle has become closer than the desired clearance, "falling back" in a standard driving manner.
- Functional limits:
    - No control at very low speeds, i.e., hand over to the driver while the speed is below a minimum speed (ISO 15622: below $v_{\text{low}} \leq 5\,\text{m/s}$ no positive acceleration).
    - Minimum set speed $v_{\text{set,min}}$ above 7 m/s ($\geq$30 km/h speedometer values).
    - The time gap shall not be below $\tau_{\min} = 1\,\text{s}$ in the steady state.
    - Priority given to driver's intervention, i.e., deactivation when brake pedal is depressed and override when accelerator pedal is depressed.
    - Driver to set desired speed v_{set} and desired time gap τ_{set}.

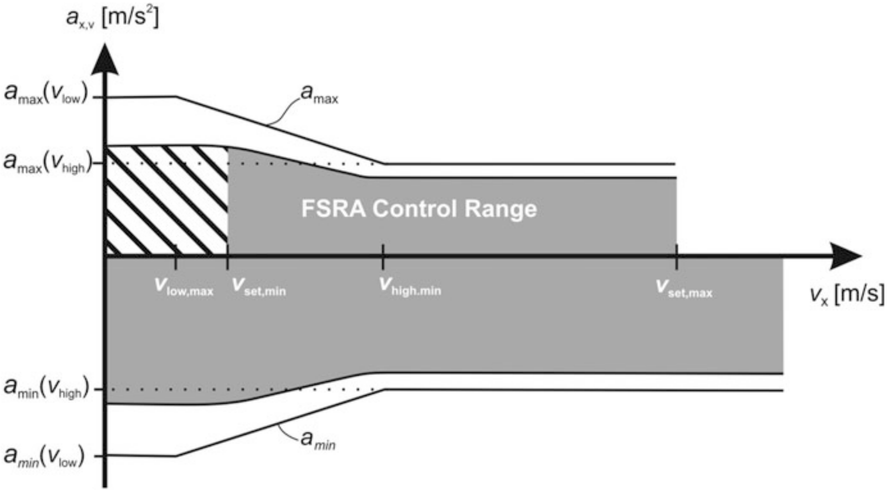


Fig. 2 Functional limits of FSR-ACC according to ISO22179

- Appropriate handover of longitudinal control to the driver in the event of system failure, in particular, when this occurs during deceleration.
- Acceleration within the limits of $a_{\min} = -3.5\ \mathrm{m/s^2}$ to $a_{\max} = 2.5\ \mathrm{m/s^2}$.

3.2 Additional Functional Requirements for FSR-ACC Pursuant to ISO22179

In addition to the requirements for the standard ACC function, full-speed-range ACC has the following additional requirements:

- When following another vehicle:
  - Control throughout the entire speed range down to 0 km/h, particularly in the creep speed range (with increased requirements for the coordination of drivetrain and brakes)
- When stopping:
  - Control of appropriate stopping distance (typ.: 2–5 m).
  - Greater deceleration capability at low speeds (Fig. 2).
  - Safe stopping with service brake in active system mode.
  - In the case of system shutdown to a standstill without driver intervention, transition into a safe holding state without power supply is required.
- Functional limits:
  - Above of $v_{\mathrm{high,\,min}} = 20\,\mathrm{m/s}$, an acceleration within the limits of $a_{\min}(v_{\mathrm{high}}) = -D_{\max}(v_{\mathrm{high}}) = -3.5\ \mathrm{m/s^2}$ up to $a_{\max}(v_{\mathrm{high}}) = 2.0\ \mathrm{m/s}$ is permitted.

- Below $v_{\mathrm{low,\,max}} = 5\,\mathrm{m/s}$ acceleration within the limits of $a_{\min}(v_{\mathrm{low}}) = -D_{\max}(v_{\mathrm{low}}) = -5.0\ \mathrm{m/s^2}$ up to $a_{\max}(v_{\mathrm{low}}) = 4.0\ \mathrm{m/s^2}$.
- Between $v_{\mathrm{low,max}}(5\ \mathrm{m/s})$ and $v_{\mathrm{high,min}}(20\ \mathrm{m/s})$, the acceleration shall be between the speed-dependent limits of $a_{\min}(v) = -D_{\max}(v) = -5.5\ \mathrm{m/s^2} + (v/10\ \mathrm{s})$ up to $a_{\max}(v) = 4.67\ \mathrm{m/s^2} - (2v/15\ \mathrm{s})$.
- The rate of deceleration γ below 5 m/s shall not exceed the jerk limit of $\gamma_{\max}(v_{\mathrm{low}}) = 5\ \mathrm{m/s^3}$ and above 20 m/s of $\gamma_{\max}(v_{\mathrm{high}}) = 2.5\ \mathrm{m/s^3}$. Between these parameters, the limit depends on the speed: $\gamma_{\max}(v) = 5.83\ \mathrm{m/s^3} - (1v/6\ \mathrm{s})$.

4 System Structure

The various tasks of ACC relate to the modules shown in the diagram in Fig. 3. The modules can in turn be subdivided, and different hardware units can be assigned. The information interfaces between the modules can vary significantly. This affects both the physical content and the data rate and bit representation. The four layers and their modules are described in the following Sects. 5, 6, 7, 8, 9, and 10, unless they are covered in depth in other chapters of this handbook. In such cases, the specific requirements made of these components by ACC are listed.

4.1 Example Mercedes-Benz DISTRONIC

Whereas a variant of *DISTRONIC* still used separate units for the long-range RADAR sensor and for processing the sensor signals, for calculating speed and distance control, and for calculating actuator control signals, as outlined in Fig. 4, in another variant all functional units from the sensor through to actuator control are incorporated in one housing, thus forming a so-called sensor control unit (SCU). ESC plays a particularly important role in the structure of this system example. As with most of the known systems, it supplies the vehicle dynamic variables for path prediction (see Sects. 7.1 and 7.2), but, in addition to brake control, which is often included in ESC, it also takes over the task of monitoring the ACC system and acts as a communication and coordination center for drive control. By separating the generation of sensor data from the subsequent traffic situation analysis and longitudinal dynamic control, it is possible to use sensors from different suppliers, and/or it is easier to adapt or extend them by adding new sensors, since only the sensor data is processed in the sensor itself (cf. (Pasenau et al. 2007)).

4.2 Function Stages

ACC was the first system to influence vehicle dynamics which, as a distributed system, would lose its core functionality in the event of the failure of a peripheral

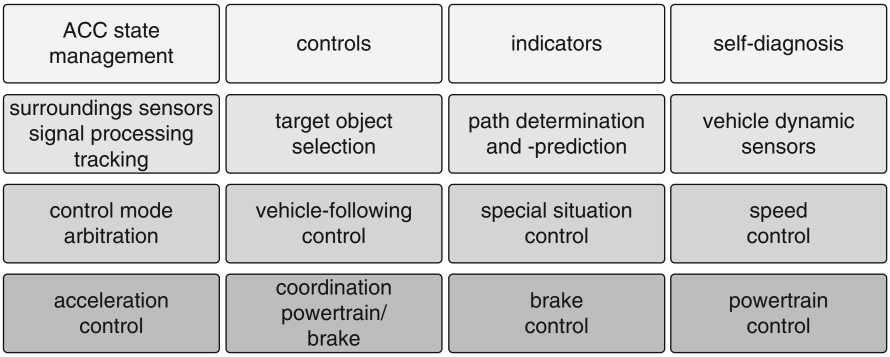


Fig. 3 Function modules of ACC systems

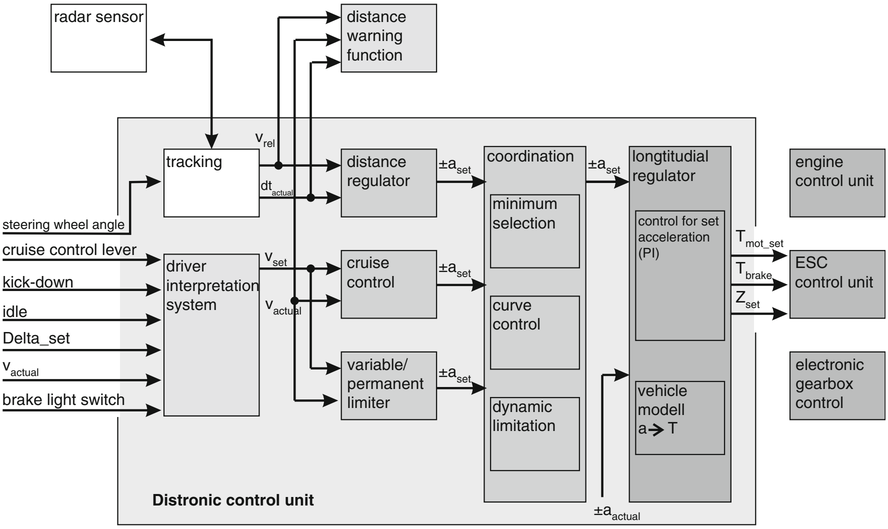


Fig. 4 Function module of the DISTRONIC system

system component. The other vehicle speed control functions (cruise control CC) with no adaptive capability can still be provided if all the necessary systems for speed control, i.e., engine, brakes, indicators, and controls, are available. However, this is not recommended, since, for example, after a longer period of unimpeded driving, drivers cannot directly recognize that their habitual distance control function is not available as they approach an object. The vast majority of systems offered, DISTRONIC included, do not, therefore, reduce the ACC function to a cruise control function if the surrounding sensors fail.

5 ACC State Management and Human-Machine Interface

5.1 System States and State Transitions

The system states of the standard ACC are illustrated in Fig. 5. The on state is the ACC off state, which, after a successful automatic function test, can be switched automatically to the ACC standby state or directly by the driver with the main switch. This standby state, insofar as the defined criteria for activation (see Table 1) are met, enables activation of the ACC active state.

If the ACC has been successfully activated, two major control states exist: speed control in situations of free cruising and ACC time-gap control for following a vehicle ahead that is traveling at a lower speed than the set speed v_{set}. If this is not the case, the speed is adjusted to the desired speed v_{set}. Transition between these states usually takes place automatically without driver intervention, purely by the detection of a target object and its distance and speed by the forward ACC sensor, as illustrated in Fig. 1.

Deactivation, i.e., transition from ACC active to ACC standby, is usually initiated by actuation of the brake pedal or intentionally switching off via the control button. The various systems on the market have even more deactivation criteria which are shown in the right-hand column of Table 1. Transition into the ACC off state is effected when malfunctions are detected by the main switch, if present.

For implementation of the ACC as a full-speed-range ACC, basically just one state is added: the FSRA hold and its transitions. This is described in Fig. 6.

The FSRA hold state marks the holding of the vehicle at a standstill by the FSRA system. A transition from the speed control to the hold state would require a desired speed 0 km/h to be permitted. It makes sense to limit the minimum desired speed $v_{\mathrm{set,min}}$ to a value > 0, e.g., 30 km/h.

In the hold state, some special features are observed. Even if it is possible to transfer the holding function to the driver with appropriately signaled instructions, it is good practice to keep the vehicle on, even by simply actuating the brake pedal. The system does not switch off, but the vehicle is kept safe from inadvertently rolling and so prevents critical system states. For safety reasons and except for very short stops, the transition from a hold state into one of the two driving states is only allowed with driver's confirmation, because it is difficult for the system alone to give clearance to drive off safely.

Similarly, the presence of the driver is monitored, as she/he can leave a stopped vehicle at any time. Upon detection of an intention to exit (e.g., open door, released belt, or no seat occupancy detection), a suitable system shutdown with a safe hold state is initiated, even in the event of power failure, for example, by activating an electromechanical parking brake. If this is not possible, the driver must be warned and/or the system switched off before the driver leaves the vehicle so that she/he is able to secure the vehicle against rolling away.

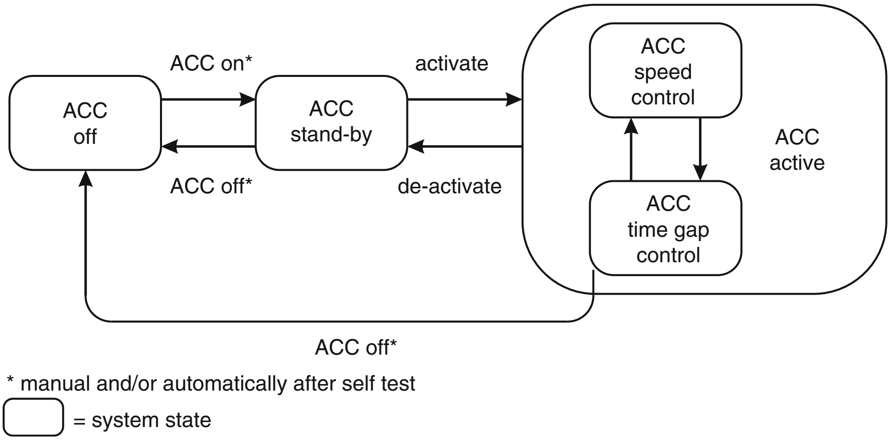


Fig. 5 States and transitions according to ISO15622

Table 1 Activation and deactivation criteria (for activation all criteria have to be fulfilled, for deactivation just one)

Activation only if simultaneously (selection)	Deactivation, e.g., when
	Deactivation by control switch
	Driver brakes at $v > 0$
$v \geq v_{\mathrm{set,\,min}}$	$v < v_{\mathrm{min}}$ (only standard ACC)
Power train ready for use	Engine speed significantly below idle speed
Forward gear engaged	Non-valid gear
ESC in full operation	ESC passive
Slip control not active	Slip control active for more than a given time (depends on the cause, e.g., 300 ms for yaw control, ca. 600–1000 ms for traction control)
Parking brake released	Parking brake activated
No ACC system failure	ACC system failure
Additionally for FSRA:	
Driver's door closed	$v = 0$ AND at least 2 of 3 signals active: door open, no belt, seat not occupied
Driver seatbelt is fastened (if available using seat occupancy sensors)	
Brake pedal acting AND $v = 0$ AND target object detected	Note: At ($v = 0$ AND driver brakes) no deactivation
Target object detected AND $0 < v < v_{\mathrm{set,\,min}}$	

Once stopping is detected, the responsibility for safe holding is transferred to the ESC system. For a short time, the braking pressure must be increased to ensure sufficient lock and permanent holding without power, controlled by an electric parking brake (EPB).

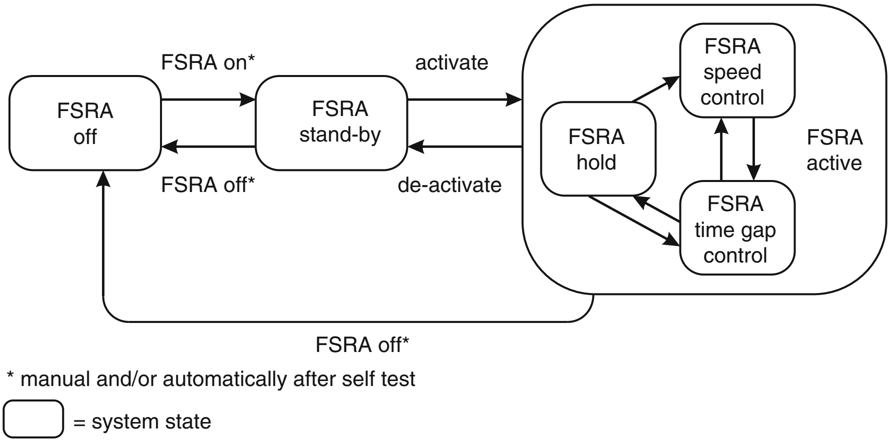


Fig. 6 States and transitions for FSR-ACC according to ISO22179

5.2 Control Elements

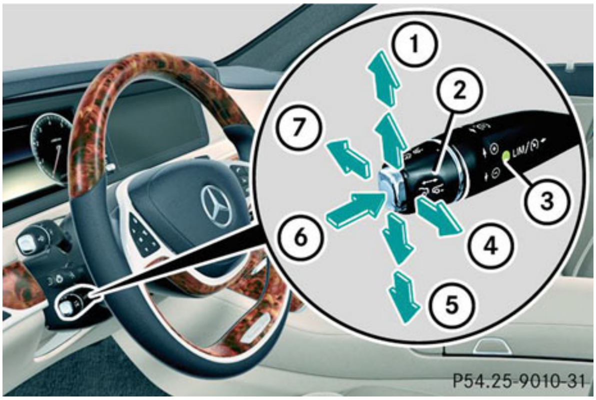
The ACC elements are used to implement the transitions from one state to another and adjust the preset control values, namely, the desired speed and the desired time gap:

- Control element to switch from ACC off to ACC standby state. There are two options:
  - A switch which is activated only once and then remains permanently in the ON position.
  - A push button which activates the controls once per ignition.
- Control element for the activation of the ACC system. This control is often also used for active control to increase the current set speed.
- Control element to reduce the current set speed.
- Control element to activate the ACC system, using the last set speed (resume).
- Control element for setting the desired time gap. Here again, there are two fundamentally different power-up states:
  - A constant initial state with a default setting, which usually corresponds to a time gap of 1.5–2 s
  - The last selected state, such as a mechanical lock

Often, the controls are arranged in groups or integrated into control levers, as the following examples illustrate.

The cruise control lever illustrated in Fig. 7 combines seven functions for the DISTRONIC PLUS of Mercedes-Benz vehicles. The actions 1 (up) and 5 (down) activate the ACC, initially adopting the current speed as the set speed. With further upward movements, the set speed is changed in small increments of 1 km/h, with

Fig. 7 Control element for DISTRONIC PLUS (Mercedes-Benz W222) with seven functions



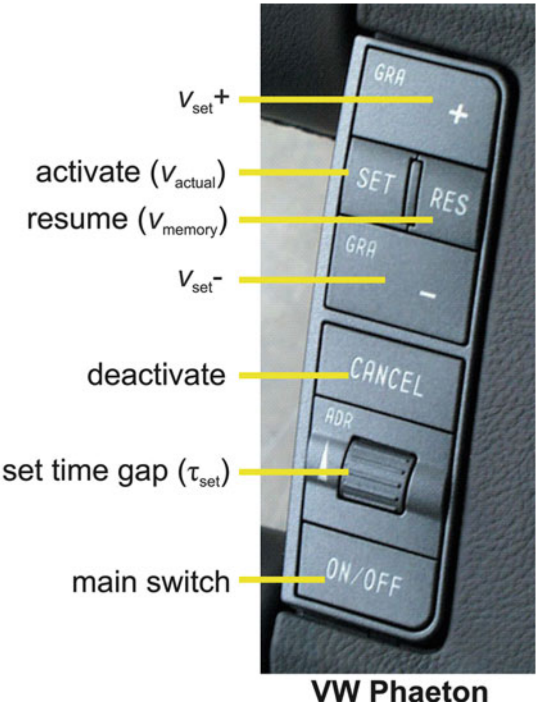
large strokes in 10 km/h increments. With corresponding downward movements, the set speed is reduced in the same way. Movement 4 (toward the driver) also activates ACC, but it resumes at the previously used set speed (resume function). The first time you activate the ACC, the current speed is adopted. Resume from standstill starts with this function. Movement direction 7 (forward) deactivates the ACC, while the movement 6 switches between cruise control and speed limiter function. The operation of the speed limiter function is analogous to the operation of cruise control/DISTRONIC PLUS. On activation, the LED 3 lights up in the cruise control lever. Operating element 2 is turned to set the desired time gap. The last set rotational position is thus available for a new driving cycle, which can use the former setting.

In addition to this typical version of a lever arrangement for the operating interface, there are others and also versions incorporated in the steering wheel, such as the example shown in Fig. 8. Despite more than 10 years' experience, there is no sign of standardization of the different versions, so that, when they change their cars, today's drivers have to adapt to the new mode of operation and, as is shown later on, also to the display.

5.3 Display Elements

Although most ACC states can be determined from the current control action, clear feedback of the states, especially during state transition, is also important for monitoring the system. However, the response of the desired speed and the desired time gap are also essential for user-friendly operation. We now differentiate between two types of display: *permanent* (p) and *situational* (s). The latter appears only when a certain event occurs for a certain time when the driver has operated a control element. The situational display has, on the one hand, the advantage that the display space can be shared with other situational display functions and, on the other, that attention can be focused better.

Fig. 8 Steering wheel ACC control in VW Phaeton (Source: Volkswagen AG)



A further distinction concerns the importance (I) of the display, distinguishing between the stages *essential*, *important*, and *helpful*. According to this classification, *essential* displays are found in all systems and important displays are found in *most*. Even while *helpful* displays are only implemented in some cars, they improve the intuitiveness of the system functions and give the driver detailed expectations and a better understanding of the system responses. The display of distance and relative speed of the newly detected object enables the driver to easily spot a false detection and assign the system responses plausibly.

As is often the case with control functions, the activation state and the set speed are also combined in the display. Table 2 shows the best-known display functions and the recommended display technology (T), where in this case initial differentiation is only between optical (o) or audio elements (a). Haptic elements are not considered since haptic display functions are not used for ACC apart from the inherent kinetic feedback of the handling.

Corresponding to the control element examples from Sect. 5.2, display concepts are presented. The display of the Mercedes-Benz DISTRONIC PLUS shown in Fig. 9 contains details about the activation state, depicting the subject vehicle (4), the desired distance (3) (orange bar “next to the road”), and target detection (vehicle 1), indicating the position of the actual clearance (2). Not listed in the picture are the speed band, which is limited at the bottom end by the speed of the preceding vehicle

Table 2 Conventional display functions for ACC (see text for abbreviations)

States	Type	I	T
Activation state	p	e	o
Relevant target object detected	p	i	o
Override by the driver	s	h	o
Below a critical distance (e.g., when cutting in and out)	s	i	a (+o)
Go possible (FSRA only)	s	h	o
Transition autom. go → driver triggered go (FSRA only)	s	h	o
System settings			
Desired speed (set speed)	p	e	o
Desired time gap (set time gap)	p, s	i	o
Speed of preceding vehicle	p	h	o
Actual distance to preceding vehicle and/or deviation between set time gap to actual time gap	p	h	o
State transitions			
ACC off - > ACC standby, if provided	p	e	o
Handover request when a system limit is reached	s	i	a + o
System shutdown	s	e	a + o

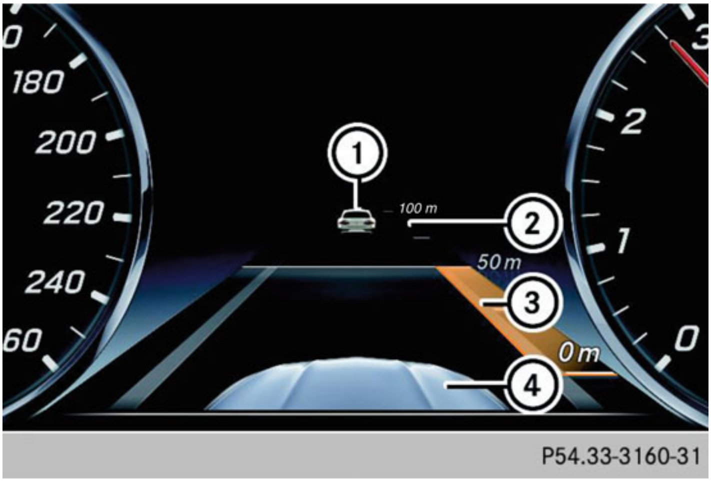


Fig. 9 Displays of DISTRONIC PLUS (Mercedes-Benz W222)

and at the top end by the desired speed that has been set, and the symbols for the take-over request and the display when the driver overrides the ACC (DISTRONIC PLUS passive).

An optionally available head-up display (Fig. 10) presents the actual speed, the desired speed, and the combined activation/time gap/target object display.

Fig. 10 ACC symbols on the head-up display (BMW E60)

6 Target Object Detection for ACC

6.1 Requirements of the Surrounding Sensor(s)

The ACC functionality succeeds or fails with the detection of the relevant target vehicle, which is the basis for the control. The first prerequisite is a set of surrounding sensors which are necessary to detect and then to decide, of the vehicles in the relevant area, whether and which of the detected objects is to be selected as the target object. RADAR and LIDAR are successfully used as surrounding sensor technologies. The requirements listed below apply equally to both. A technical description of the sensors can be found in ▶ Chaps. 17, “Automotive RADAR,” and ▶ 18, “Automotive LIDAR.”

6.2 Measurement Ranges and Accuracies

6.2.1 Distance

According to subdivisions defined for ISO 15622 (ISO TC204/WG14 2010), the standard ACC function requires that objects are detected from the minimum detection distance $d_{\mathrm{min0}} = \mathrm{MAX}\left(2\ \mathrm{m},\ (0.25\ \mathrm{s} \cdot v_{\mathrm{low}})\right)$, and also the distance must be determined from $d_{\mathrm{min1}} = \tau_{\mathrm{min}}(v_{\mathrm{low}}) \cdot v_{\mathrm{low}}$ (Fig. 11). $\tau_{\mathrm{min}}(v_{\mathrm{low}})$ is the smallest time gap at the smallest allowed ACC operation speed. As the time gap is increased at low speeds, $d_{\mathrm{min}\ 1}$ is about 10 m. There is no need for distance measurement below this distance because the ACC will always decelerate in this situation or in any case the driver is asked to take over at speeds of below

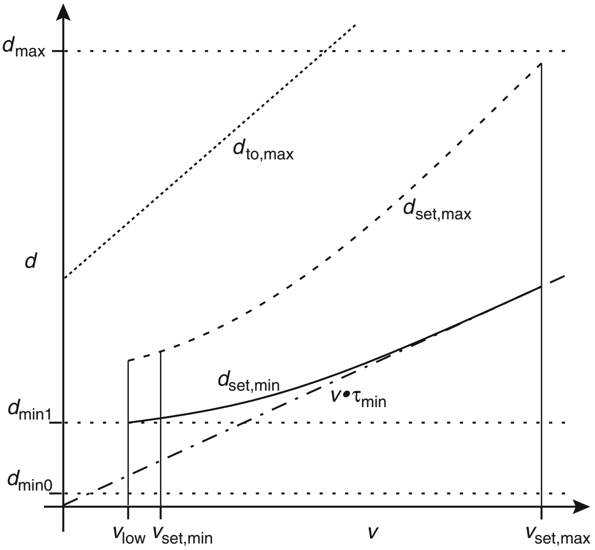


Fig. 11 Requirements of the distant range pursuant to ISO15622 plus requirements as a function of driving speed

Table 3 Distance requirements of typical setups

	$v_{low} = 5\,m/s$ (= 18 km/h)	$d_{min0} = 2\,m$
$\tau_{set,\,min}(v_{low}) = 2\,s$		$d_{min1} = 10\,m$
$\tau_{set,\,max} = 2\,s$	$v_{set,\,max} = 50\,m/s$ (= 180 km/h)	$d_{max} = 100\,m$

v_{low}. If speed falls below $d_{min\ 0}$, it can be assumed that a control process will be interrupted before reaching such a small distance from the driver. The same applies in the case of a cut-in in close proximity to the subject vehicle, for which drivers will not rely on the ACC function but will resolve the situation through their own brake operation.

The maximum distance d_{max} required must, of course, enable control with the maximum target distance, i.e., the distance for setting the largest time gap at the maximum setting speed $v_{set,max}$. A control margin is usually reserved for the comfort of the system control. Since a set time gap of at least $\geq$1.5 s is required, the maximum time gap can only be reduced as far as this limit.

The above requirements (ref. Table 3) are minimum requirements and apply only to stationary pursuit. A greater distance is desirable for an approach, particularly if the speed difference is considerable. As shown in Sect. 7, the more difficult the target selection is, the greater the distances involved; consequently, in many cases, a deceleration reaction is experienced as negative at a distance greater than 120 m even if the target selection is working without error. This is particularly the case if you intend to overtake the target vehicle. Overtaking is impeded by the ACC deceleration reaction before the lane change has started.

In practice (ref. (Winner and Olbrich 1998)), a restriction of the reaction range has been useful. Particularly, in the lower and middle speed range, there is no benefit from reacting throughout the entire range since objects have no effect on one's own vehicle at a great distance. An exemplary boundary curve $d_{\mathrm{to,max}}$ is shown in Fig. 11. In another case, $d_{\mathrm{to,\,max}} = \mathrm{MAX}\left(v \cdot 3.6\ \mathrm{s},\ d_{\mathrm{range,\,min}}\right)$ is used, that is, to say exactly the same number of meters as km/h, but at least $d_{\mathrm{range,\,min}}(\approx 80\,\mathrm{m})$.

No high demands are made on the accuracy of distance measurement as the system responds only weakly to distance variations, as is shown below. With the control loop gain given below, distance errors d_{err} of 1 m (effective value in the band from 0.1 to 2.0 Hz) propagate to acceleration amplitudes of at maximum $a_{\mathrm{set,\,err}} = 0.1\,\mathrm{m/s^2}$ and, therefore, remain below the notice threshold of 0.15 m/s^2 (Meyer-Gramcko 1990) when vehicle following. Gain errors of the distance ε_{d} can therefore be tolerated up to a level of 5 % without any noticeable impact upon the driver. However, the minimum set time gap should be selected with an appropriate margin, so that the minimum time gaps defined for steady vehicle following are not less than $\tau_{\mathrm{min}} = 1\,\mathrm{s}$, because of the gain error associated with the relative error.

6.2.2 Relative Speed

The accuracy of the relative velocity must fulfill far higher requirements than that for distance. Any deviation of the relative velocity leads to a change of acceleration (see Sect. 8). A static offset leads to a steady deviation of the distance, with an offset of 1 m/s leading to an approximately 5 m distance deviation. Fluctuations in the speed of $v_{\mathrm{rel,\,err}} = 0.25\,\mathrm{m/s}$ (rms in the 0.1–2 Hz band) are still accepted, as the resulting subsequent acceleration fluctuations remain below the driver's sensitivity. While filtering the speed signal can reduce the fluctuations effectively, an excessive delay has to be avoided as otherwise the control quality is adversely affected. As a guideline, a maximum delay time of 0.25 s can be used, wherein for stable control with the smallest time gap of $\tau_{\mathrm{min}} = 1\,\mathrm{s}$, 0.75 s remains for the control time constant and the actuator delay.

Relative errors $\varepsilon_{\mathrm{vrel}}$ of the relative speed up to 5 % are largely unproblematic for vehicle-following control, since the consecutive acceleration control with the control systems for brake and drivetrain produces similarly large deviations, and thus the relative distortions of the control set point caused by the speed error are hardly noticeable.

Greater challenges for the accuracy of the relative velocity are posed by the classification of objects, whether they are moving in the same direction, at a standstill, or in the opposite direction. For this classification, tolerances must be smaller than 2 m/s and 3 % of v_{rel}. The relative error can also be calibrated with stationary objects because they are measured much more frequently and are thus seen as an accumulation in a statistical measurement. This allows even those errors to be compensated that result from vehicle speed determination based on the rolling circumference, whose accuracy is usually limited to 2 %.

6.2.3 Lateral Detection Area for Standard ACC Function

The requirements for the lateral detection range are derived from the initial assumptions:

- $\tau_{\max}$, the maximum time gap for following distance control
- $a_{y,\max}$, the maximum assumed lateral acceleration for cornering
- $R_{\min}$, the smallest curve radius specified for the ACC function

For a given curve radius $R \geq R_{\min}$, the maximum cornering speed can be determined by the maximum lateral acceleration. If this is multiplied by the time gap $\tau_{\max}$, we obtain the required maximum range $d_{\max}(R)$. The offset of the curve line $y_{\max}$ at $d_{\max}$ (Fig. 12), however, is independent of the curve radius and speed:

$$y_{\max} = \frac{\tau_{\max}^2}{2} \cdot a_{y,\max} \tag{1}$$

The maximum azimuth angle $\phi_{\max}$ is determined by the ratio of the maximum offset $y_{\max}$ and the maximum range $d_{\max}$ at $R = R_{\min}$:

$$d_{\mathrm{Rmin}} = d_{\max}(R_{\min}) = \tau_{\max}\sqrt{a_{y,\max} \cdot R_{\min}} \tag{2}$$

$$\phi_{\max} = \arcsin(y_{\max}/d_{\max}(R_{\min})) \approx y_{\max}/d_{\max}(R_{\min}) \tag{3}$$

Because of the observed driver behavior (e.g., see (Mitschke et al. 1991)), the underlying fundamental lateral acceleration is dependent on the driving speed. Implicitly, this results in a dependence on the curve radius, as tighter bends are traversed at lower speeds. This is reflected by the different values for the standard curve classes defined in ISO 15622. Thus, $a_{y,\max} = 2.0\,\mathrm{m/s^2}$ is adopted for $R_{\min} = 500\,\mathrm{m}$ and $a_{y,\max} = 2.3\,\mathrm{m/s^2}$ for $R_{\min} = 250$ and $R_{\min} = 125\,\mathrm{m}$, respectively. In Fig. 13, for a maximum time gap of $\tau_{\max} = 2\,\mathrm{s}$, the necessary (unilateral) angle $\phi_{\max}$ is illustrated for three different maximum lateral acceleration assumptions. Despite this highly idealized real cornering view, measurements from the field (Winner and Luh 2007; Luh 2007) show that the formula above and the assumptions can be used to determine the requirements for the opening angle curve for a given capability. The two empirical values refer to the curve radius at which half of the following distance-controlled runs occurred without a target loss.

Another outcome of the investigations (Winner and Luh 2007; Luh 2007) showed that with an opening angle of $\Delta\phi_{\max} = 16^\circ\ (\pm 8^\circ)$, both subjectively and objectively the standard ACC function is covered to a sufficient extent and a further increase of the azimuth angle range results in less improvement of the standard ACC function, as long as the detection of cut-in vehicles by the dynamics of target selection is specified (see Sect. 7). As evident in the subsequent consideration of the overall error, a small azimuth alignment error leads to significant functional impairment. Since the tolerance limit depends on many individual factors, particularly on back-scattering properties of the objects, no definite value can be given.

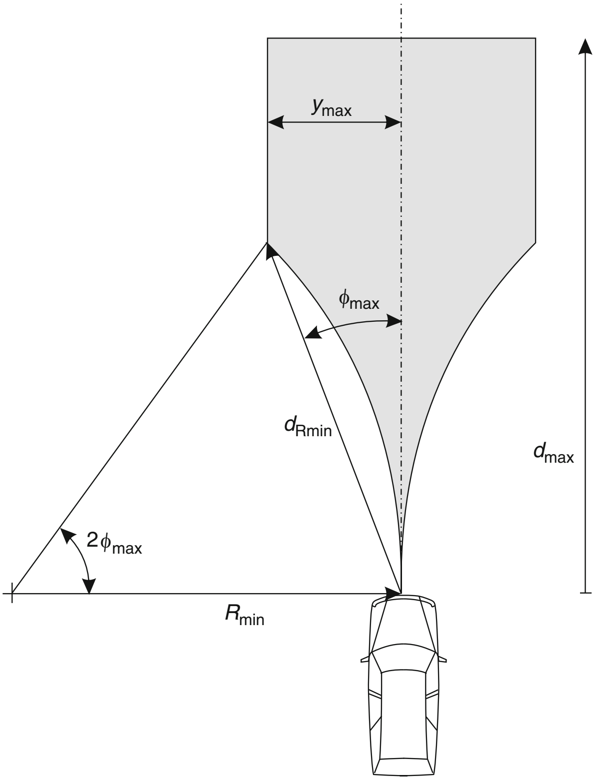


Fig. 12 Required detection range (in azimuth angle) depending on the curve radius at constant lateral acceleration and time gap

Static error of 0.25° should be avoided while dynamic noise-like errors should be smoothed with filters and may amount to 0.5° without serious detrimental effects on the system function.

6.2.4 Lateral Detection Range for FSRA

One hundred percent coverage of the area directly in front of the vehicle is aimed at in order to enable an automatic go function. As this is difficult to realize in practice, the minimum requirements are much lower in the emerging FSRA ISO standard 22179 and can be met with a centrally positioned sensor with an opening angle $\Delta\phi_{max} = 16° (\pm 8°)$. Standing starts in such systems are therefore limited to a corresponding driver-enabling function, even after very short stops.

An excess of ±8° coverage of the area in front of the vehicle up to about 10–20 m is required for close and staggered following distance control at low speeds. This most likely occurs in congested conditions when not driving directly behind the vehicle ahead but also helps to achieve a better view. Even if the target object changes lanes slowly, coverage will become increasingly difficult if one's own required driving corridor is not free. A sensor with a too narrow azimuth angle range will lose the target object, although it could not be passed

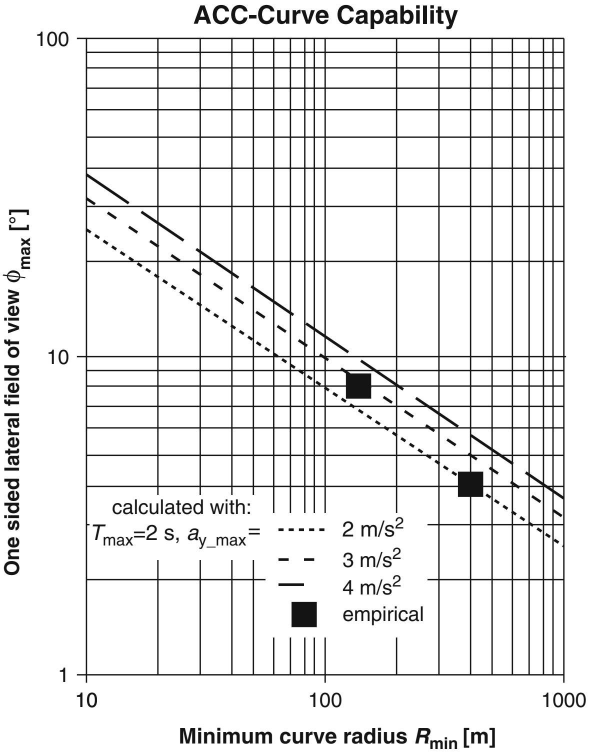


Fig. 13 Required azimuth angle range depending on assumed lateral acceleration and time gap. *Lines*, theoretical path; *dots*, experimental results for two angle ranges

without colliding. So the driver has to intervene in these cases. Figure 14 shows the desired range for complete coverage.

From the minimum distances which can be typically expected for the cut-in, i.e., about 2–4 m at very low speeds, coverage of neighboring lanes also makes sense (at least up to half the width) to ensure an early detection of cut-in vehicles. Reliable angle determination is very important, as only by calculating the lateral movement from the angle values can ACC respond to cut-in vehicles. The earlier FSRA systems from Mercedes and BMW therefore used two forward-pointing 24 GHz UWB RADAR sensors (UWB, ultra wideband; see ► Chap. 17, "Automotive RADAR"), thereby providing a good compromise between range (approx. 20 m), azimuth angle (approx. 80°), and the angle resolution dictated by the sensor principle. The wide included angle provides a large area of overlap in the detection ranges, resulting in stable object detection. Greater ranges are not necessary, since the long-range RADAR sensors that are used partially cover the neighboring lanes at these distances.

6.2.5 Vertical Detection Range

Vertical coverage requires the detection of all relevant objects for ACC (trucks, cars, motorcycles). Since the objects are not very high off the ground or are lower

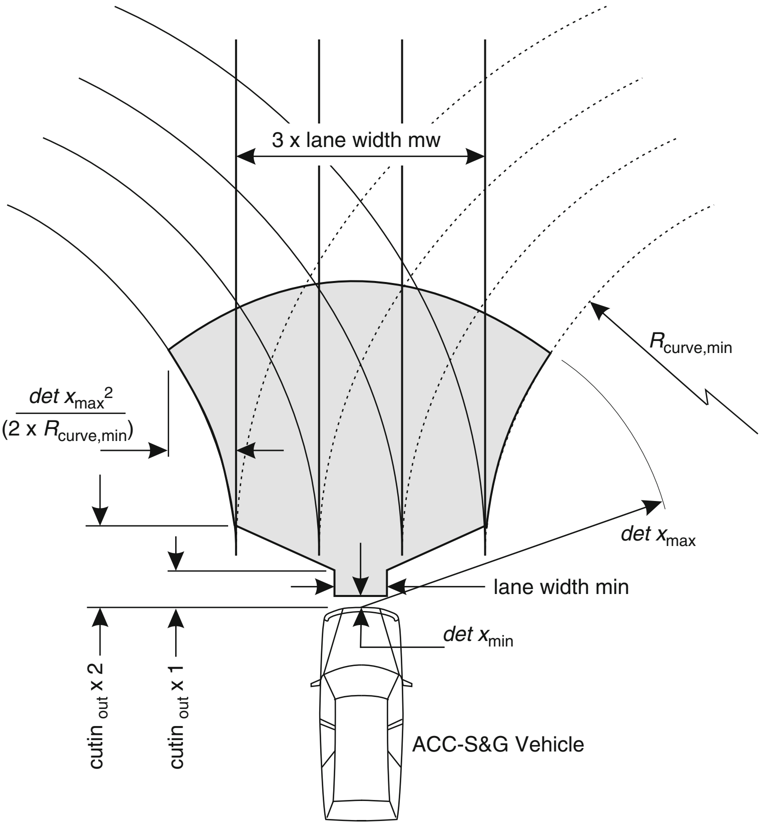


Fig. 14 Desired range for complete close-range coverage

than the normal sensor installation heights, only the slope parameters and static and dynamic changes within the pitch utilized by ACC dynamics are considered. In practice, requirements of $\Delta\vartheta_{max} = 3° \ (\pm 1.5°)$ have resulted.

Incorrect elevation angles usually have only a small negative effect, as the elevation is rarely used as a measured variable, for example, 2D scanning LIDAR scans the environment in several superimposed horizontal lines. However, in the case of RADAR sensors, changes of the antenna pattern with deviating elevation angles of greater than 0 can be expected. Furthermore, it is necessary to prevent the elevation of the available area from being reduced by a misalignment to the extent that the above requirement is no longer guaranteed.

6.2.6 Multi-target Capability

As several objects may be present in the sensor field, a multi-target capability is very important. This particularly means the ability to differentiate between relevant objects in the driving corridor and irrelevant objects, such as those on the adjacent

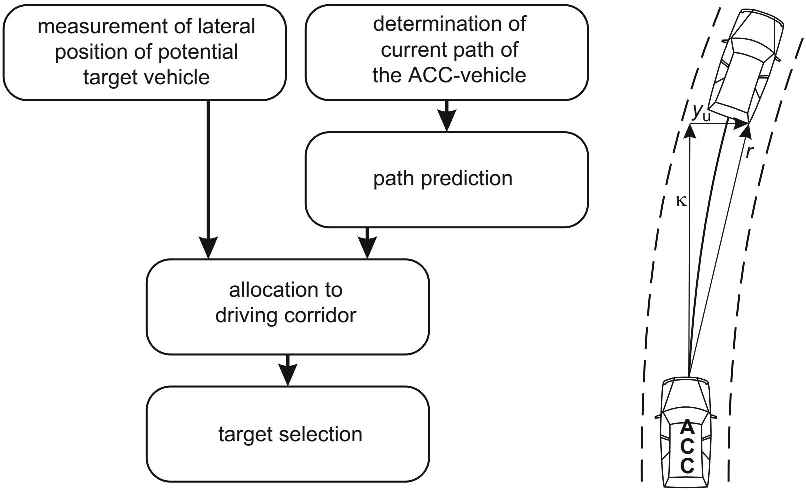


Fig. 15 *Left*, steps for target selection; *right*, definition of the variables

lane. This differentiation can be achieved by a high degree of differentiation of at least one measured variable (distance, relative velocity, or azimuth angle). However, the requirement for a high degree of differentiation must not be at the expense of association problems in which the objects are identified repeatedly as new objects. As described in ▶ Chap. 17, "Automotive RADAR," this can be combated by association windows in the tracking, which are adjusted to the dynamics of the objects and of the subject vehicle carrying the sensor.

7 Target Selection

Target selection is of very great importance to the quality of the ACC, as both relevant objects can be overlooked and false targets may be selected. In both cases, the user's expectations of this system are not fulfilled.

The following error analysis is based on the need for target selection, as shown in the steps in Fig. 15 (left). The measurement of the lateral position $Y_{U,i}$ of the object i is carried out by the ACC sensor with an uncertainty of $\varepsilon_Y \approx \varepsilon_\phi \cdot r$, which results from the inaccuracy of the angle determination ε_ϕ.

7.1 Determination of the Path Curvature

The curvature κ describes the change of direction of a vehicle as a function of the distance traveled. The constant part of a curve is the reciprocal of the curve radius $R = 1/\kappa$. The curvature of the vehicle trajectory can be determined by various onboard sensors and assumes for all calculations that they are used outside dynamic vehicle limits. So, they are not valid for skidding situations or in the presence of significant wheel slip.

7.1.1 Curvature Calculated from the Steering Wheel Angle

In order to calculate the curvature κ_s from the steering wheel angle δ_H, three vehicle parameters are required: the steering ratio i_{sg}, wheelbase l, and the characteristic speed v_{char} derived from the understeering behavior in the linear dynamic range, characterized, therefore, at low lateral accelerations. Under typical ACC conditions of very good approximation, κ_s could be determined according to

$$\kappa_s = \frac{\delta_H}{\left(i_{sg} l\right)\left(1 + \frac{v_x^2}{v_{char}^2}\right)} \tag{4}$$

7.1.2 Curvature Calculated from the Yaw Rate

To calculate the curvature κ_Ψ from the yaw rate, the driving speed v_x is required and the slip angle rate is disregarded:

$$\kappa_\Psi = \frac{\dot{\Psi}}{v_x} \tag{5}$$

7.1.3 Curvature Calculated from the Lateral Acceleration

The calculation of the curvature κ_{ay} from the lateral acceleration a_y also uses the driving speed v_x:

$$\kappa_{ay} = \frac{a_y}{v_x^2} \tag{6}$$

7.1.4 Curvature Calculated from the Wheel Speeds

For the curvature κ_v from the wheel speeds, the relative difference of the wheel speeds $\Delta v / v_x$ and the width of the track b are required. In order to minimize driving influences, the difference $\Delta v = (v_l - v_r)$ and driving speed $v_x = (v_l + v_r)/2$ are determined by the speed of the non-driven axle.

$$\kappa_v = \frac{\Delta v}{v_x b} \tag{7}$$

Although all these methods can be used for the determination of curvature, they all have different strengths in different operating conditions. They differ particularly in crosswind, lateral road inclination, wheel radius tolerances, and in terms of sensitivity in different speed ranges.

Table 4 Comparison of the different approaches for curvature determination

	κ_s	κ_Ψ	κ_{ay}	κ_v
Robustness against crosswind	− −	+	+	+
Robustness against lateral road inclination	− −	+	− −	+
Robustness against wheel radius tolerances	o	+	+	−
Sensitivity at low speeds	+ +	o	− −	−
Sensitivity at high speeds	−	o	+ +	−
Offset drift	+	− −	− −	+

As Table 4 shows, the curvature from the yaw rate is best. However, a further improvement in signal quality can be achieved if some or all signals are used for mutual comparison. This is especially possible because the ACC vehicle is equipped with ESC, and therefore all the above sensors are part of the system. At standstill, there is offset adjustment of the yaw rate, but this requires a halt phase that does not occur when driving on highways with no traffic jams. In this case, statistical averaging methods can be used, as the average of the yaw-rate sensor supplies the offset over long distances.

7.2 Path Prediction

To predict the future path, the system needs to know the (future) path of the carriageway and the future driving lane choice of the ACC vehicle and also, in fact, those of the potential target vehicles. Since this information is not always available without image processing or vehicle-to-vehicle communication, working hypotheses are used, which employ simplified assumptions.

One simple hypothesis is the assumption that the current curvature will be retained. This basic hypothesis continues to be used until further information is available. This approach disregards entries and exits to bends, changes in the lane markings, and also drivers' steering errors. If past lane marking assignments are available, the hypothesis that the objects and the ACC vehicle will remain within their lanes will be used. However, this is invalidated if objects cut in or out or if the driver changes lane. Nor does it help for the initial assignment.

The compromise approach is to delay the object data by half of the time gap and to allocate it on the basis of the then-current path curvature. This is very robust at the start and end of curves, because, due to the delay, the curve of the road between the objects and the ACC vehicle is used, and thereby good assignment is permitted even when curvatures change. This method does not, however, constitute a replacement for the first one in the case of initial assignment. Additional options for path prediction are offered by GPS navigation in conjunction with a digital map and the curvature information stored therein. Unfortunately, such maps are not always up-to-date and roadworks are not marked. The method which uses static objects at the roadside to determine curvatures is also only partially useful in the absence of such objects, but is presumably nevertheless included in most ACC path prediction algorithms. The lateral movement of vehicles in front can also be used to improve path prediction, because in most cases this is an early indication of an impending bend in the road.

The use of lane marking information from the processing of camera images is obviously very promising. However, an improvement at distances of over 100 m cannot be really expected today, since today's standard camera pixels correspond to around 0.05°, a value which, at around 120 m, corresponds to a width of 10 cm and is therefore barely adequate for lane marking detection. In addition, image-based path prediction outside the headlight beam is impossible in darkness, especially when roads are wet.

The aforementioned algorithms are used in different ways and to different degrees by different manufacturers, but overall they deliver as a starting value a predicted trajectory curvature κ_{pred}. The trajectory can then be extrapolated depending on the distance. Instead of the circular function, a parabolic approximation is sufficient in the case of the normal opening angle:

$$y_{\text{c, u}} = \frac{\kappa_{\text{pred}}}{2} d^2 \tag{8}$$

The cross-track error values $y_{\text{i,u}}$ of the object detected by the ACC sensor can then be related to this trajectory resulting in the relative offset:

$$\Delta y_{\text{i,c}} = y_{\text{i,u}} - y_{\text{c,u}} \tag{9}$$

Errors of the predicted curvature therefore increase quadratically with the object distance d. At high speeds ($v_{\text{x}} \geq 150\,\text{km/h}$), a curvature error κ_{err} of less than 10^{-4} /m is acceptable, but it will cause an error at 100 m of $\Delta \text{y}_{\text{i,c,err}}(100\,\text{m},\ 150\,\text{km/h}) \approx 0.5\ \text{m}$. At 140 m, the error doubles due to the quadratic propagation. At low speeds ($\approx 50\,\text{km/h}$), the curvature error pursuant to Eq. 5 is around three times higher. The distance for $\Delta y_{\text{i,c,err}}(57\ \text{m}, 50\ \text{km/h}) \approx 0.5\ \text{m}$ is just 57 m from which a reduction of the maximum target selection distance at low speeds is derived.

7.3 Driving Corridor

The driving corridor is a term frequently employed by experts for the corridor used for the ACC target selection. In its simplest form, it is determined by a width b_{corr} (not dependent on distance) with the predicated path as a center line (Eq. 8). Initially, one might equate the driving corridor width with the lane width b_{lane}. However, this assumption has been found to be inappropriate.

The example in Fig. 16 shows that there is an area in which clear assignment is impossible based on the measured lateral position alone.

Since it cannot be assumed, however, that the measured lateral position of the object corresponds with the center of the object, both the right-hand and the left-hand object edges must be taken into account. Vehicles traveling off-center present

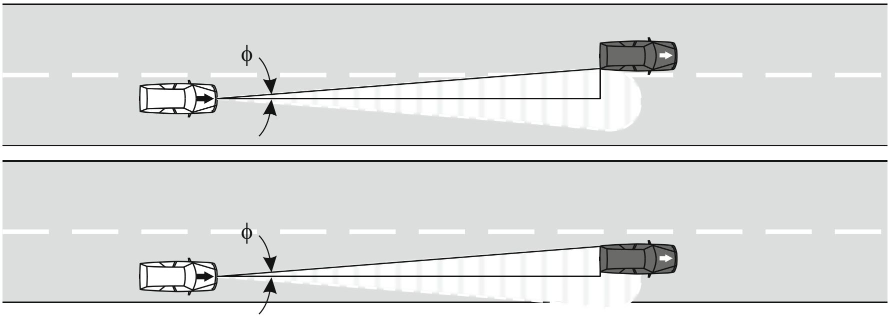


Fig. 16 Example of differing assignments despite equivalent relevant data

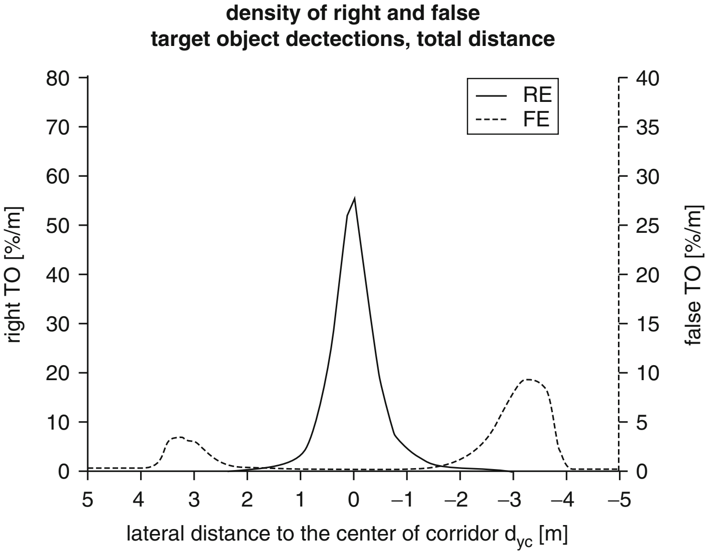


Fig. 17 Temporal distribution of lateral target object position to the center of the corridor for an 8 m wide corridor (left *y*-axis and *solid line*, true positive detection; right *y*-axis and *dashed line*, false positive detection) (Luh 2007)

further uncertainty of assignment, both of the ACC vehicle and of the potential target vehicle. Therefore, assignment to the actual lane is only certain if the measured lateral position (without errors) lies within ±1.2 m of the predicted path center (without errors). The assignment of the object to the neighboring lane is only certain if its position is at least 2.3 m from the path center. The values relate to a lane width of 3.5 m.

It follows from the statistics recorded by a RADAR sensor and illustrated in Fig. 17 that some misrecognition must be expected even in a lane width of 3.5 m but, on the other hand, that in a narrower driving corridor, target losses can be expected.

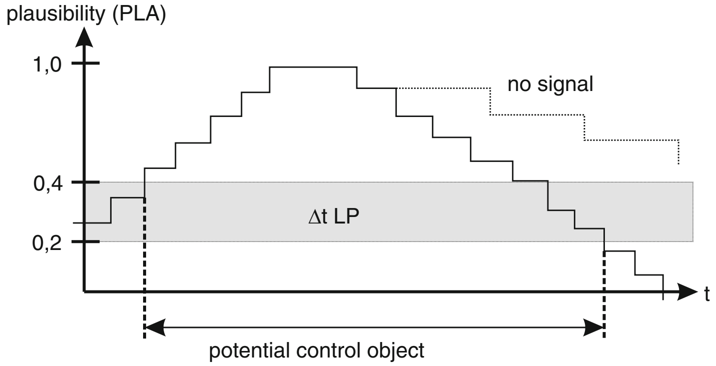


Fig. 18 Formation of target plausibility (illustration from Luh (2007))

Three measures are employed to improve target selection: a variable driving corridor width depending on the type of road, an approximate driving corridor contour, and a local and temporal hysteresis function for the target selection.

Two pieces of information are important for a variable driving corridor width: Are there neighboring lanes to the left or right? If not, the selection area can be very wide on the respective side of the driving corridor (around 2 m to the respective side, i.e., 4 m if there is no driving lane in the same driving direction on either side). Information about the presence of neighboring lanes can be obtained from the observation of static targets at the roadside and by oncoming vehicles, whereby changes, e.g., broadening to two lanes in a single direction, can only be detected with a time delay. If a neighboring lane is detected, e.g., by the observation of vehicles in the same direction with a lateral position outside one's own lane, a statistical observation of the lateral positions can be used to adjust the driving corridor so that roadworks involving narrower lanes can be negotiated.

A further measure is local hysteresis, which means that an object marked as a control object is assigned a wider driving corridor than all other objects. Typical differences are around 1 m, i.e., around 50 cm on either side. This prevents misrecognition of objects in the neighboring lanes particularly during changing conditions (entering and exiting bends, uneven steering) while nevertheless keeping the target object stable in these situations.

Temporal hysteresis is also used, as shown in Fig. 18.

As compared with assignment reliability (LP, lane probability), target plausibility (PLA) increases in the case of positive LP. Above an upper threshold (in this case 0.4), the object becomes the target object unless other criteria suggest otherwise. The target plausibility may increase to a maximum value (in this case 1) and decrease based on two options: In the case of the absence of detection (no signal) and if allocated to the neighboring lane (negative LP). Once below the lower threshold (in this case 0.2), the object loses the characteristic of being able to be selected as the target object.

The assignment measurement LP can be approximately mapped as shown in Fig. 19. The further away the object is, the less clear is the transition between the

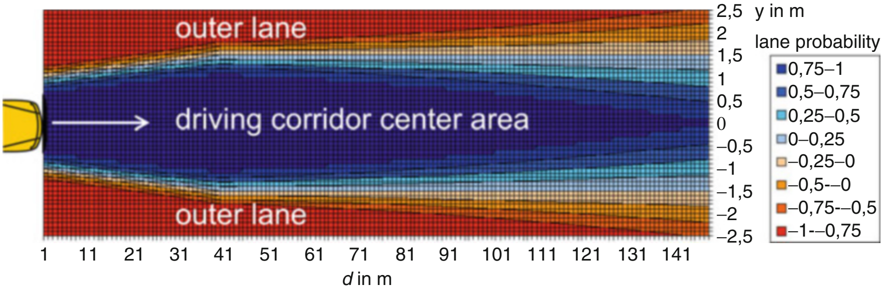


Fig. 19 Fuzzy corridor contour to avoid assignment errors

lane assignments. Therefore, account must be taken of the fact that errors of location determination and path prediction increase with distance. In addition, other estimated uncertainties can dynamically restrict the core range, e.g., a major bend in the road.

7.4 Further Criteria for Target Selection

It can make sense to use other criteria in addition to the lane assignment. The most significant criterion for target selection is the object speed. Oncoming vehicles are completely ignored for control purposes. Nor are static objects selected as target objects, with the exception of those already detected as objects moving in the travel direction (so-called halted objects). These are relevant particularly for the full-speed-range ACC function in the same way as objects traveling in the same direction are relevant. Permanently static objects are often used for other functions (see also Sect. 9.3) and therefore are subject to separate filters. For basic ACC functions, however, they play only a minor role.

Another simple but very effective approach is to limit the distance as a function of the travel speed (cf. Fig. 11). Thus, at a speed of 50 km/h, a reaction to targets which are more than 80 m away is neither necessary nor expedient since the danger of mistaken assignment greatly increases with distance. Empirical values suggest a distance value $d_{\mathrm{to},0} = 50\,\mathrm{m}$ and a slope of $\tau_{\mathrm{to}} = 2\,\mathrm{s}$.

$$d_{\mathrm{to,\,max}} = d_{\mathrm{to},0} + v \cdot \tau_{\mathrm{to}} \tag{10}$$

If several objects meet the criteria for a target object, the following decision-making criteria are considered individually or in combination:

- The smallest longitudinal distance
- The smallest distance to the path center (minimum $|\Delta y_c|$)
- The smallest set acceleration

The final criterion assumes a connection to the ACC or a multi-target object interface, but improves the transition in the case of target objects cutting out.

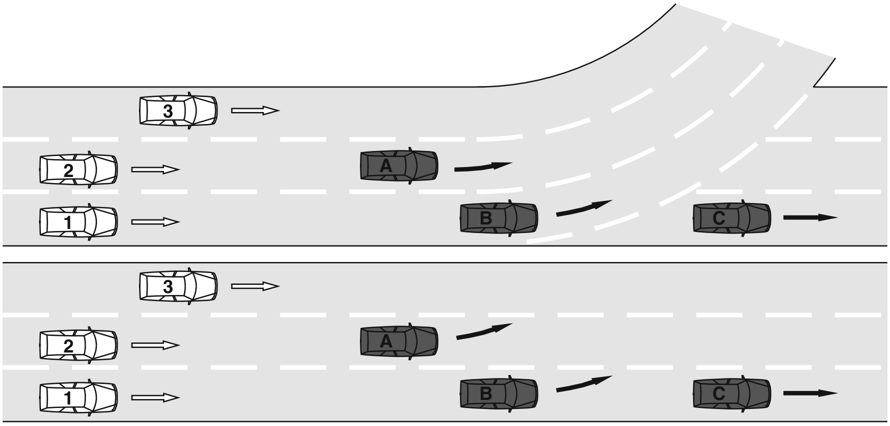


Fig. 20 Situation example for ambiguous target assignment (vehicle positions and movements are identical in both pictures)

7.5 Target Selection Limits

The approaches illustrated in the latter sections are highly effective and have reached a high quality level. There are, however, situation-dependent constraints which require explanation as in the following two examples. The vehicles in Fig. 20 move identically in the moment illustrated, assignment as the "correct" object, however, proceeds differently due to the differing progress of the road. Another example is the "overtaking dilemma" when approaching at high speed. For comfortable braking to follow a significantly slower moving vehicle, deceleration needs to start at a great distance away. On the other hand, the probability that the preceding vehicle must be overtaken is particularly high if the difference in speed is large. Early deceleration would considerably hinder the overtaking process. Since the overtaking process is rarely indicated more than 6 s before reaching the vehicle to be overtaken (Winner and Olbrich 1998), the dilemma exists between too early a response for unhindered overtaking and too late a response for a comfortable or even adequate approach.

Another parameter is the late detection of cutting-in vehicles. On the one hand, temporal and local hysteresis in the driving path assignment leads to a delayed reaction of around two seconds with respect to the moment the vehicle cutting in crosses the lane marking. Since drivers are aware of the cut-in even before the marking is crossed, due to the situation and the indication of the lane change by the direction indicators, the late reaction is once again a critical point for the user. The same phenomenon occurs with the detection of vehicles cutting out, although target approval is objectively correct once the neighboring lane has been fully entered.

If lane information is available (e.g., from a camera system), the temporal and spatial hysteresis can be much smaller, but nevertheless significant improvement of cutting-in and cutting-out detection on the part of ACC can only be achieved by situation classification, which interprets the indicators also seen by people. However, there is then a danger that this might adversely affect the transparency of the ACC functionality.

Improving the target selection in the case of lane changing on the part of the ACC vehicle can be achieved by interpreting the direction indicator, resulting in a shift of the driving path to the indicated direction. The digital map in combination with the search and detect function also enables an adaptive driving corridor function.

On average, modern ACC systems perform an incorrect assignment only once every hour, a value which is surprisingly small in the light of the many potential errors and one which is hard to improve, as described in detail in Winner (2005).

8 Vehicle-Following Control

Although the ACC vehicle-following control is often described as a distance control, it is anything other than a means of difference-guided distance control. As a point of departure for further considerations, it is assumed that the controller output is manifested in direct vehicle acceleration with no time delay and, moreover, that the ACC vehicle follows the target vehicle at the set time gap τ_{set}. Disregarding vehicle lengths, it therefore follows that the ACC vehicle will reach the position of the target vehicle after a time gap of τ_{set}. If the ACC vehicle now echoes the position of the target vehicle with a time lag, the time gap will be retained irrespectively of speed. In the same way, the speed and acceleration of the preceding vehicle are imitated with a time lag. Thus in the steady state, a simple control principle can be derived which even avoids feedback:

$$\ddot{x}_{i+1}(t) = \ddot{x}_i(t - \tau_{\text{set}}) \tag{11}$$

The index $i + 1$ stands for the ACC vehicle in a convoy with a continuous index i. The notation is introduced with reference to the observation of the convoy stability as its measurement of the quotient $V_{\text{C}} = \hat{\ddot{x}}_{i+1}(\omega)/\hat{\ddot{x}}_i(\omega)$ of the (complex) acceleration amplitudes. The convoy is stable precisely when the condition

$$|V_{\text{C}}| = \left|\hat{\ddot{x}}_{i+1}(\omega)/\hat{\ddot{x}}_i(\omega)\right| \leq 1,\ \text{für}\ \forall\omega \geq 0 \tag{12}$$

is fulfilled. Otherwise, from a disturbance which is quite small, the frequency components of the frequencies for which this condition is not fulfilled will be greater with each subsequent convoy position. Convoy stability obviously applies for the idealized control principle represented in Eq. 11 because

$$|V_{\text{C}}| = \left|e^{-j\omega\tau_{\text{set}}}\right| = 1 \tag{13}$$

even if semi-stable without damping. This approach is not suitable in practice, but it illustrates a basic controller design. The disadvantages of this approach are the numerically unsuitable detection of the acceleration of the preceding vehicle (differentiation of the relative speed and the driver's actual travel speed, the required filtering leads to phase delay) and the fact that there is no correction opportunity if the speeds do not match or in the case of deviations in the distance.

For this purpose, the following is a control design based on relative speed:

$$\ddot{x}_{i+1}(t) = (\dot{x}_i(t) - \dot{x}_{i+1}(t))/\tau_v = v_{rel}/\tau_v \tag{14}$$

or in the frequency range

$$\hat{\ddot{x}}_{i+1}(\omega) = \hat{\ddot{x}}_i(\omega)/(1 + j\omega\tau_v) \tag{15}$$

With few steps, this approach can be transferred to an acceleration-led approach such as in Eq. 11, wherein the acceleration value of the preceding vehicle is not delayed by a fixed time but is filtered in a PT1 element and thereby delayed implicitly by τ_v. The application of Eqs. 14 and 15 is obviously convoy stable, but it only meets the requirements of a constant time gap if τ_v equals τ_{set}. Moreover, this control approach is not suitable for reducing any distance deviations. For this purpose, the controller is extended by an additive correction component for the relative speed which is proportional to the difference between the set and actual distances:

$$\ddot{x}_{i+1}(t) = \left(v_{rel} - \frac{d_{set} - d}{\tau_d}\right)/\tau_v \tag{16}$$

or in the frequency range

$$\hat{\ddot{x}}_{i+1}(\omega) = \hat{\ddot{x}}_i(\omega)\frac{1 + j\omega\tau_d}{1 + j\omega(\tau_d + \tau_{set}) - \omega^2\tau_d\tau_v} \tag{17}$$

The stability condition $|V_C| \leq 1$ is now only met if τ_v is small enough:

$$\tau_v \leq \tau_{set}(1 + \tau_{set}/2\tau_d) \tag{18}$$

So far, the choice of the distance control time constant τ_d has been left open. For this purpose, a reference scenario can be used, namely, falling back in a cut-in situation. In this case, it is assumed that the cutting-in vehicle cuts in without a speed difference at a distance which is 20 m smaller than the set distance. An appropriate reaction would be a delay of around 1 m/s^2 which equates to taking one's foot off the accelerator pedal or very slight braking. For such a response according to Eq. 16, the product must be $\tau_v \cdot \tau_d = 20$ s^2. This result is the fundament for the following considerations.

It follows from Eq. 16 that the smaller the vehicle-following time gap, the higher must be the loop gain defined by τ_v^{-1} for the relative speed. However, a high loop gain also means a minimum damping of speed fluctuations of the target vehicle, as shown in Fig. 21 for frequencies above 0.05 Hz.

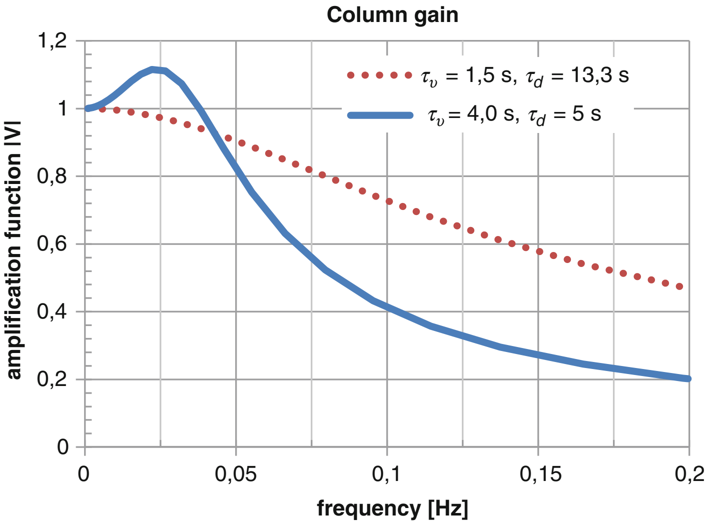


Fig. 21 Column gain for various loop gains

As is demonstrated by Witte's measurements in Witte (1996), these variations noticeably occur at "driver-regulated" driving speeds, due to the fact that drivers only start to use a constant accelerator setting as a correction when there is a significant variation from their wishes, and this is only changed to another value when they again notice a deviation.

It is impossible to achieve both stability on the one hand and a high degree of uncoupling from the speed variations of preceding vehicles on the other. An implicit differentiation between cases is a possible way out of this dilemma. Drivers are most sensitive to variations when they are driving in steady vehicle-following mode with only slight speed differences. The problems associated with convoy instability only occur where there are significant deviations from the steady state. Obviously, therefore, the control loop gain must be selectively designed to cater for such a difference. This can be done very simply via a characteristic curve with two kinks at $\pm\Delta v_{12}$ (Fig. 22), a rounded transition also being possible. This approach makes it possible to damp the speed variations with high control time constants within the control differences of $\Delta v \approx 1\,\mathrm{m/s}$, but, if greater dynamics are required, such as a greater deceleration step, stability can be achieved on the large signal level.

Only the basic principle is transferred to the actual controller function, since additional influencing variables require a modification. This can be expressed via lookup tables or more complex mathematical functions. Moreover, in the above analysis, all other system delay times are ignored, and this is justified neither for the surrounding sensors nor for the subordinate acceleration control circuit. As a set value, the control circuit time constants must be reduced by the delay times in order to fulfill the stability conditions.

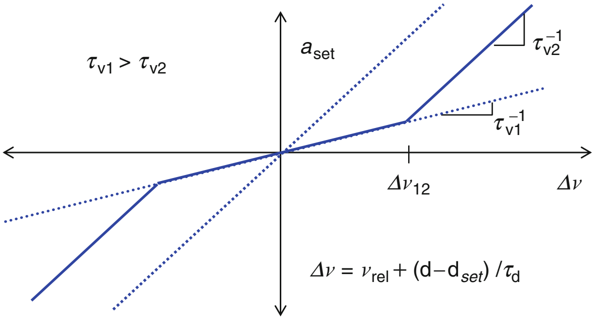


Fig. 22 Control loop gain curve of a nonlinear distance and relative speed controller

9 Target-Loss Strategies and Curve Control

When negotiating bends, target loss is possible because the maximum azimuth angle of the ACC sensor is inadequate (see Sect. 6.2) to detect the target object. Even when traveling straight ahead, short-term target loss is possible if, e.g., reflectivity is low (e.g., motorbike) or object separation fails. In these cases, immediate acceleration to a set speed, after the cutting-out of the target vehicle, would be inappropriate. A differentiation is often made between these two scenarios in that in the cutting-out case, the target plausibility (see Sect. 7.3) deteriorates due to a negative assignment measurement (LP < 0) to the driving corridor, and the object is still detected in the case of this "loss of target." Conversely, when driving round narrow bends or in the case of other target losses, the response to which should not be rapid acceleration; the target loss is related to object detection errors and a positive assignment (LP > 0) to the driving path in the last known measurement. The response design differs when this differentiation is made: In the first case, acceleration after loss of target is brisk unless a new target object limits acceleration, while in the second case acceleration is initially suppressed. Yet, how long should this continue and what strategies then follow? The time gap preceding the loss of target is used for the duration of the suppressed acceleration. Should the target object disappear from the measurement range because it has entered a bend, this can be verified by the ACC vehicle after traveling the distance corresponding to the time gap, because the curvature will be different from that at the time of the target loss. If this curve criterion is met, the acceleration suppression strategy can be replaced by a curve control. In the other case, it is assumed that the target object is no longer in the driving corridor, and the speed is then adjusted to the new situation.

In the case of curve control, two aspects are important: Lateral acceleration and the effective range $d_{\max,\text{eff}}$ of the ACC sensor. This is given by the curvature κ and the maximum azimuth angle $\phi_{\max}$ and equals approximately:

$$d_{\max,\,\text{eff}} = 2\phi_{\max}/\kappa \tag{19}$$

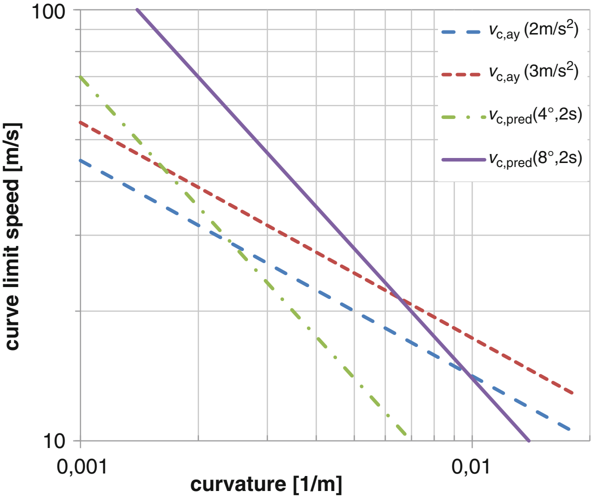


Fig. 23 Curve limit speeds for curve control depending on the curvature κ ($v_{c,ay}$ limit speed resulting from the limit lateral acceleration $a_{y,max}$, derived $v_{c,p}$ from the maximum azimuth angle ϕ_{max} and the width of the look-ahead region)

From this, a speed $v_{c,\,p} = d_{max,\,eff}/\tau_{preview}$ can be deduced for the minimum time gap $\tau_{preview}$ available for an approach:

$$v_{c,\,p}\left(\kappa, \phi_{max}, \tau_{preview}\right) = 2\phi_{max}/\left(\kappa \cdot \tau_{preview}\right) \tag{20}$$

This speed enables the decision to be made as to whether to accelerate further. This strategy results in an appropriate driving strategy precisely for very tight bends, e. g., cloverleaf junctions.

The second criterion, presumably used in all ACC systems, is lateral acceleration. As in the derivation of the curve classification (Sect. 6.2), a lateral acceleration limit $a_{y,max}$ that defines the comfort range is assumed which is between 2 m/s^2 (at higher speeds) and 3 m/s^2 (at lower speeds). From this, in turn, a curve limit speed $v_{c,ay}$ can be derived:

$$v_{c,\,ay}\left(\kappa, a_{y,\,max}\right) = \sqrt{a_{y,\,max}/\kappa} \tag{21}$$

Both limit speeds are illustrated for two typical values respectively in Fig. 23. If the actual driving speed is above the set speeds, a positive acceleration is either reduced or even reversed (negative) without falling into the range of significant decelerations of more than 1 m/s^2.

In combination with the "target-loss acceleration suppression" described above, astoundingly good "blind flights" of a proven quality of 80 % were achieved in a series of tests (Luh 2007), measuring whether the driver continued the run without intervening following target loss.

The reaction to curve-related target losses can be improved using information from a digital map, ideally with precise lane accurate positioning (not yet supplied in production vehicles). The curvature is then detected in advance and adjusted to the control strategy at motorway exits.

Another challenge for the ACC developer is when the target vehicle turns off the carriageway. The change in direction of the speed vector of the preceding vehicle results in a perceptible deceleration for the following vehicle. As the sensor measures only this, the ACC vehicle deceleration is disproportionate and must be reduced in a suitable manner.

9.1 Approach Strategies

The approach capability is defined as the maximum negative relative speed $-v_{\mathrm{rel,\,appr}}$, which can be controlled by ACC with respect to a vehicle traveling at a constant speed before a critical distance $d_{\mathrm{appr,min}}$ is exceeded. It depends on the distance $d_{\mathrm{appr,0}}$ at the start of the deceleration, on the assumed constant maximum increase of the deceleration . . . $x_{\mathrm{v,\,min}} = -\upgamma_{\max}$ and on the maximum deceleration = minimum acceleration $\ddot{x}_{\mathrm{v,\,min}} = -D_{\max}$.

$$-v_{\mathrm{rel,appr}} = \sqrt{2D_{\max}\cdot\left(d_{\mathrm{appr,0}} - d_{\mathrm{appr,min}} + \frac{D_{\max}^{3}}{6\gamma_{\min}^{2}}\right)} - \frac{D_{\max}^{2}}{2|\gamma_{\min}|} \tag{22}$$

$$d_{\mathrm{appr,0}} = d_{\mathrm{appr,min}} - \frac{D_{\max}^{3}}{6\gamma_{\min}^{2}} + \left(-v_{\mathrm{rel,appr}} + \frac{D_{\max}^{2}}{2|\gamma_{\min}|}\right)^{2} / 2D_{\max} \tag{23}$$

The distance required for a noncritical approach increases approximately fourfold with the differential speed and approximately reciprocally with respect to the maximum deceleration. At a distance of 100 m at $D_{\max} = 2.5\,\mathrm{m/s^2}$, around 20 m/s (72 km/h) differential speed can be compensated for, while for an approach capability of $|v_{\mathrm{rel,appr}}| = 100\ \mathrm{km/h}$, $d_{\mathrm{appr,0}} \approx 120\ \mathrm{m}$ and $D_{\max} \approx 3.5\,\mathrm{m/s^2}$ are required.

The ramp of the deceleration reduces the approach capability but increases transparency for the driver, see also Sect. 3.2.

Of course, in the case of dynamic approaches, it is impossible to avoid values falling below the stationary set distance and/or the set time gap. Therefore, a significantly smaller reserve distance value can also be used as the set distance $d_{\mathrm{appr,min}}$ for a successful approach. It should be noted however that undershooting is only permissible over a distance of 250–300 m.

9.2 Overtaking Assistance

Vehicle following and overtaking are incompatible. Vehicle-following control therefore has to be temporarily modified for overtaking assistance to be implemented. If the overtaking maneuver can be accurately predicted, the current preceding vehicle could be ignored and the subject vehicle driven as if there were no vehicle in front. However, the direction indicator is not a clear indication of either the actual intention to overtake or of the desired or possible start of the maneuver. The first case occurs if the direction indicator signals the intention to turn left. Since this situation rarely occurs at high speeds, and also because overtaking is usually associated with a high speed, a compromise solution is that overtaking assistance is only used at speeds in excess of 70 km/h.

It is not feasible to simply blank out the current target, because “signaling a left turn” is often also used to tell the driver in front to keep the overtaking lane free. However, because it is impossible to predict if and when this intention is going to be followed up, only a cautious reduction of the previous nominal distance remains for “swinging out.” Within this phase, the overtaking process should be initiated with a detectable change in direction. The necessary rapid “letting go” of the previous target can be supported by a vehicle path offset to the left. If the overtaking maneuver cannot be followed through as wished, after a few seconds of advance phase, the ACC reverts back to normal vehicle-following mode. However, this function is only suitable in countries with high relative dynamics, e.g., in Germany. In the USA, on the other hand, there are often only slight differences in speed between the lanes, so that this function has to be offered in a much different form or omitted altogether. Alternatively, with corresponding sensor performance (especially azimuthal coverage range and multi-target capability), the speeds of vehicles in the target lane can be analyzed to form a basis for overtaking assistance.

9.3 Reaction to Static Objects

Static objects may be obstacles lying in the driving corridor. However, most are irrelevant targets such as drain covers, bridges, or signs. Even at speeds of 70 km/h, deceleration of around 2.5 m/s^2 should be started some 100 m before the object. However, since the probability of error in the target selection is still very high in this case, a reaction to static objects is recommended only in exceptional cases. The most important exception relates to the history of static objects. If these are measured in advance with an absolute speed which can be distinguished from zero, these are classified as “stopped” objects and may also be treated as potential target objects. Otherwise, the conditions for a response to static objects are limited to an immediate area of up to approx. 50 m. The response may be either suppression of acceleration, in which a speed increase is suppressed as long as the static object is detected in the driving corridor, or a warning that the vehicle will drive over the object. A braking response to static obstacles is only possible, if the error probability of target selection can be considerably reduced. By using several sensors, where possible based on different physical principles, and robust sensor data fusion, it is possible to achieve a quality of perception that is adequate for braking interventions for static targets.

9.4 Stopping Control: Specifics of Low-Speed Control

In principle, it is not necessary to have a different control approach for low-speed control; however, the nominal values must give greater weighting to distance and speed deviations. For example, in a state controller structure, the corresponding controller gains can be adapted to the situation via a fuzzy estimate. This gives a greater degree of comfort and driver-like behavior. Since the distance to the preceding vehicle is short at low speeds, particular attention must be given to situations such as vehicle cutting in close," "object too close," or "stopping" in this speed range. Because of the greater weighting of distance and speed deviations in detecting these situations, the controller reaction is quicker and this ensures dynamics appropriate to the given situation. Compared with "normal" ACC operation, higher decelerations (up to 5 m/s^2) are permitted in low-speed control. This also allows dynamic stopping. However, decelerations in excess of this are not useful, as this would make drivers think that the system can automatically master any situation, so that they cease to be aware of the function limits. An additional feature of low-speed control is traffic jam detection. If a traffic jam is detected on the basis of sensor data (e.g., due to the person in front repeatedly driving off at a low pace, low maximum speed, stopping shortly after setting off), the distance and speed deviations are given a lower weighting, to produce gentler controller behavior. This can then guarantee different behaviors with dynamics appropriate to the situation, e.g., a tailback on a motorway or a queue at traffic lights.

10 Longitudinal Control and Actuators

10.1 Basic Structure and Coordination of Actuators

Longitudinal control presents the challenge of converting adaptive speed control, namely, the ultimate set acceleration obtained from various individual controllers, into an actual acceleration. For this purpose, the sum forces (or sum torques) of respectively self-contained subordinate control circuits of the drive and brake survey systems are adjusted so that the desired acceleration can be implemented. Even though simultaneous actuation of a drive torque and a brake torque is possible, it is generally avoided and the respective elements are controlled independently.

With respect to the design of harmonic transitions between drive and brakes, it makes sense to choose a physical value with which both actuator subsystems can be controlled equally. Wheel torques (but also wheel forces) are an option. In this case, a summary observation is sufficient, i.e., the sum of the wheel torques acting on all four wheels since ACC does not apply torques to individual wheels. In this way, coordination on the basis of the same physical signal can be effected as close to the actuators as possible, as shown in the following sections.

ACC requires the implemented actual sum wheel torque (and/or force) to calculate the driving characteristic equation and therefore, among other things, estimate the road gradient. The drive must also provide the current set maximum and minimum value as the sum wheel torque. In this case, particularly the minimum possible torque, i.e., the achievable torque in the current gearshift operation, is important, since the brake can only be activated when it is no longer possible to decelerate via the drive.

As outlined in more detail in later chapters, ACC control does not require absolute accuracy of the actuators since deviations from the required reference value can generally be well compensated for by the enclosed control loop. Only at the start of the control process and at transitions between the different actuators does good absolute accuracy simplify the control. In principle, however, good relative accuracy is necessary to achieve the desired control comfort.

10.2 Brake

The ACC systems without braking system intervention, which were initially supplied mainly by Japanese car manufacturers, were not well received in Europe because of the minimum deceleration caused solely by the engine drag torque in conjunction with gear shifting meaning that the driver had to actuate the brake too often. The pioneering equipping of top-class vehicles since 1995 with ESC and Brake Assist, which support the driver in emergency braking situations, has greatly simplified the implementation of a suitable brake intervention for ACC systems, and hardly any ACC is now found without brake system intervention.

As an engine torque interface for the set torque requirement had already been introduced for traction (TCS) and stability control systems (ESC), it made sense to use brake torque as interface signal. This has the advantage that the distribution between the various actuators in the ACC controller is very simple, and the specifics of the individual actuators have little or no influence on the controller design, which greatly simplifies transferability to different vehicles and models.

If one observes the transfer function of the brake, one recognizes that pressure changes with the factor 0.1 effect the deceleration, i.e., pressure increases of 1 bar are only just below the human detection threshold of 0.15 m/s^2. Correspondingly high are the metering requirements for the brake actuator.

The following applies for the relationship between deceleration D (= negative acceleration) and braking pressure p_{Br}:

$$\Delta D = \Delta p_{\mathrm{Br}} \frac{A_{\mathrm{K}} \mu_{\mathrm{Br}} R_{\mathrm{Br}}}{m_{\mathrm{v}} R_{\mathrm{dyn}}} \tag{24}$$

Key to symbols:

ΔD	Change in vehicle deceleration
Δp_{Br}	Change of braking pressure
A_{K}	Total area of the brake pistons
μ_{Br}	Sliding friction coefficient between brake pad and brake disk
R_{Br}	Effective radius of the brake disks (average value)
m_{v}	Vehicle mass
R_{dyn}	Dynamic radius of the wheels

$A_K \cdot \mu_{Br} \cdot R_{Br} = 70\ \mathrm{Nm/bar}$ and $m_v = 2100$ kg and $R_{dyn} = 0.34$ m can serve as approximate values. Transmission values of 0.07 . . . 0.14 m/(s^2bar) therefore result depending on the vehicle configuration.

10.2.1 Actuation Ranges

If a deceleration of 2 m/s^2 is required for standard ACC, then, with an even road surface, 20–25 % of the maximum pressure of approximately 80 bar required for full braking with good road conditions is sufficient. However, if we take account of maximum load capacity, unbraked trailer loads and driving downhill, we obtain much higher values of up to 50 % of the maximum pressure. In order to have sufficient reserve to allow for different brake disk friction conditions, as well as fading, the permissible actuation range must be increased at the top end, so that, in extreme cases, the entire available actuation range can be used. The safety of the braking intervention cannot, therefore, be monitored via the value of the required braking torque, but only via the deceleration effected.

10.2.2 Actuator Dynamics

For comfort functions such as ACC, deceleration variations of up to 5 m/s^3 are typically permitted (see Sect. 3.2). This would result in a required pressure variation dynamic of 30–40 bar/s. In order to be able to follow the specified torque and/or pressure trends sufficiently dynamically, the brake system must be able to follow changes of up to 150 bar/s.

Dynamic following of the set value with sufficiently rapid pressure increase up to the start of braking, and as far as possible delay-free following in the case of pressure modulations is required. The maximum time delays in this case should remain <300 ms. The preconditions for this, along with a correspondingly dimensioned pump, are predominantly the choking of the hydraulic system in the pump intake area in order to be able to provide the required volume, largely irrespective of temperature. The control of the set value must be free from vibration and overshooting as this will be perceived by the driver as extremely unpleasant. Together with rapid following of dynamic set values, as far as possible stepless following of small or slowly changing set values is absolutely essential because precisely this kind of control of small control differences is typical for ACC operation. Stationary deviations are also to be avoided as these turn into speed and distance errors and can lead to limit cycle oscillations.

10.2.3 Control Comfort

As already described in the introduction, the vehicle responds very sensitively to pressure changes. To ensure that a sensitive driver perceives the pressure increase as continuous, the brake system must be able to handle braking increments of less than 0.5 bar. As far as possible, braking pressure increase and decrease should be silent, harmonious, and continuous. Inadvertent pressure changes of more than 1 bar are to be avoided. Additional pump elements are beneficial for an even pressure increase, while continuously controlling valves are beneficial for pressure decrease. In terms of acoustics, a low pump speed is desirable, as are suitable location of the hydraulic unit and the suitable placement of brake leads in order to prevent the absorption of vibrations from the chassis. A complicating factor is that in the case of brake intervention, one of the substantial noise emitters in the vehicle, the engine, is reduced to its acoustic minimum, the drag range.

10.2.4 Miscellaneous Requirements

- The brake light has to be actuated independently of the driver applying the brakes. In brake actuator systems with an active booster, this can be achieved at the pedal without changing the brake light switch, whereas, in brake actuating systems with hydraulic pumps, the brake light can be controlled by the control unit as a function of brake pressure and deceleration. Flickering of the brake light must be avoided by means of minimum actuation times or switching hysteresis.
- The distribution of brake pressure between the front/rear axles must be kept identical to normal brake application to prevent overloading of the brakes on one axle or unstable vehicle behavior. Additional brake circuit pressure sensors have proven useful in this context. During protracted braking, they allow leakage to be detected and compensated for in one circuit.
- When the driver brakes during ACC braking, the pedal feedback should be kept to a minimum. In particular, vibrations or even shocks at the pedal must be avoided, and the transition into the normal brake pressure curve should be smooth.
- If the vehicle becomes unstable, the vehicle control (ABS, TSC, ESC) has priority, and transitions into skid control must be suitably designed.
- Safeguards: In the event of errors in the ESC system, the brake pressure must be reduced immediately; in the event of errors in the partner control devices, braking must be terminated or the pressure gradually ramped down, depending on the severity. It is also necessary to ensure that all cut-off signals (in addition to the brake light switch), such as the actuator elements, handbrake operation, wrong gear, etc., are safely processed.

10.2.5 Feedback Information

The brake subsystem is the key supplier of internal vehicle state values, the most important of which are vehicle speed, yaw rate, steering angle, brake light switch, and slip control information. Also, the current actual brake moment is fed back in order that the ACC can carry out a gradient estimate. The ESC system provides binary state information (flags) (e.g., ABS active, TSC active, ESC active) for an appropriate response to control states.

10.2.6 Additional Requirements for FSRA

- When braking at low speeds, greater demands are placed on the brake control noises, mainly because of the absence of driving and engine noise. Similarly, braking noises such as squealing or judder must be minimized.
- Because of the higher decelerations, initial braking behavior is altered; the pedal should not become too stiff.

– Standstill management: Once it detects that the vehicle is at standstill, the FRSA hands over responsibility for keeping it at standstill to the ESC, giving rise to the following tasks:
  - Increasing (or even decreasing after sharp decelerations) the brake pressure for keeping the vehicle at standstill, road gradient detection is advantageous for this.
  - Permanent roll monitoring and, if necessary, increasing brake torque.
  - Slip detection at very low friction condition, possibly releasing the brake, to maintain steerability.
  - Safe transition to no energy holding (actuation of the electric parking brake, EPB) on recognizing the driver's intention to alight from the vehicle.
  - Monitoring the temperature of the hydraulic system because of heating due to permanent valve currents, and switching off with driver warning when overheating is expected.
  - With engine start-stop systems, it is necessary to ensure that all necessary functions remain active during the voltage drop that accompanies engine start-up, in particular, correct closing of the hydraulic valves that are responsible for maintaining the necessary brake pressure.

10.3 Drive

A review will now be made of the combination of the combustion engine and the automatic transmission, with the manual transmission treated as a special case. Combinations with hybrids are also conceivable. In principle, we can say that transitions between electric and combustion engines must be imperceptible as far as possible for the ACC as for the driver. The drive is, furthermore, only a torque generator for the ACC since how the torque is generated is irrelevant for the system function. With respect to recuperated braking with an electric motor, it is important to ensure corresponding coordination with the brake system, which has to implement the changeover to the friction brake.

It has proven useful to view the engine and transmission as one unit from the perspective of the ACC and to specify direct total wheel torques as set values and to leave it to the drive system to decide how this torque is to be applied, either by changing the engine torque or by changing the gear ratio.

Consequently, a change of acceleration as observed for braking Δa is proportional to the sum wheel force change and/or the sum wheel torque change:

$$\Delta a = \frac{\Delta F_{\mathrm{R}\Sigma}}{m_{\mathrm{v}}} = \frac{\Delta M_{\mathrm{R}\Sigma}}{m_{\mathrm{v}} R_{\mathrm{dyn}}} \tag{25}$$

Key to symbols:

Δa	Change of vehicle acceleration
m_{v}	Vehicle mass

R_{dyn}	Wheel radius
$\Delta F_{\text{R}\Sigma}$	Total wheel force variation
$\Delta M_{\text{R}\Sigma}$	Total wheel torque variation

While direct actuation of the engine via engine torque set value is possible, specific action is required to influence the transmission in order to maintain an adequate dynamic while avoiding unwanted gearshifts. Simply using a heavily gearshift-based characteristic as in CC is inadequate because the ACC-following controller must be far more dynamic in design than a vehicle speed controller purely designed for constant travel.

Equally, a direct conversion of the engine torque set values into virtual driving pedal angles in order to control the transmission logic is not suitable because the ACC attempts to emulate a preset acceleration precisely and, other than with the driver, deviations are directly reflected in set value changes which at certain operating points could lead to toggling gearshifts.

10.3.1 Engine Control (Control Range, Actuator Dynamics, Steps/Accuracy, Feedback Information (Loss Torque of Ancillary Units))

As with the brake, the ACC requires access to the entire possible torque range for the necessary control range in order to cover all relevant driving situations. The required servo dynamics corresponds to the dynamics required by the driver, which should not present a problem in most modern systems because the driver set values can also be transferred electronically; driver and ACC settings are therefore principally transferred via the same path.

The drive optimally applies the required sum wheel torque of the ACC function (similar to the accelerator pedal) at the respective operating point. The states of engine, transmission, and auxiliary units will be considered for calculating the set value. Coordination is autonomous in the drive system as far as possible. If this is not supported, conversion to the engine torque by the ACC control unit or a longitudinal dynamic module is required, for which the current transmission ratio must be known.

The ACC function differentiates between different operating modes for comfort purposes which relate to the coordination of the drive's different actuation options (e.g., fuel cut-off, gearshifts, incorporation of ancillary units). Thereby, minor inconstancies in the torque application such as occur, for example, when activating the fuel cut-off, have to be avoided or tolerated. In addition, more serious inconstancies in the torque application such as occur, for example, during additional gear shifting in automated multistep reduction gears, are avoided or permitted.

Examples

– Triggering the fuel cut-off, but no additional gearshifts (only rolling gearshifts), on approaching a slowly moving target object or on reducing the desired speed.

- Triggering the fuel cut-off and additional gearshifts during static downhill travel to support the brake system in downhill mode.
- Suspending the fuel cut-off during static downhill travel to cancel previously effected gearshifts. This prevents "fuel cut-off toggling" and allows the gearbox to cancel shift-in, if there is a change in gradient during static downhill travel.

Special Features of Combination with Manual Transmissions

The engine control determines the drive ratio of the wheel torque/crankshaft torque via the transmission ratio of the different gears and thereby calculates an engine torque from the drive requirement of the ACC function and implements this in the best possible manner.

During the shift process, i.e., after activation of the clutch by the driver, the engine control adjusts the crankshaft speed for synchronization of the crankshaft/transmission input speed in the target gear. The crankshaft speed set value is determined depending upon the target gear predicated in the engine control unit.

The engine control evaluates the crankshaft speed and advises the driver, taking into account the situation, to select a lower gear. The request to select a higher gear is not necessary.

To prevent the engine from stalling, the engine must have the opportunity to switch off the ACC function if the driver does not respond to the shift instruction. ACC is also switched off if the clutch process exceeds a time limit (e.g., 8 s), or a suitable gear has not been engaged.

10.3.2 Transmission Control

The ACC state control requires, as one of its activation criteria, information from the transmission that a valid (forward) gear is engaged.

If engine torques are to be preset, the ACC needs the current torque amplification V_S from the transmission; this is the ratio of the force $F_{R\Sigma}$ on the drive axle to the engine torque M_M and is given by the product of converter amplification μ_W, the gear ratio i_G of the current gear, and the axle gear transmission i_A divided by the dynamic wheel radius R_{dyn}:

$$V_S = \frac{F_{R\Sigma}}{M_M} = \frac{\mu_W \cdot i_G \cdot i_A}{R_{dyn}} \tag{26}$$

In this case, the hydrodynamic converter amplification is generally incorporated as a curve in a lookup table which may need to be temperature compensated.

FSRA may also use electronic gearshift systems as additional protection for stopping management. In this case, the parking lock is engaged on detection of the intent to leave the vehicle. This is sufficient in combination with a multistage driver early warning system which advises the driver of the responsibility to make the vehicle safe when it is stopped.

It is not sufficient, however, as the sole safety means for a fully automated stopping management system (without driver intervention), since the parking lock blocks only the propeller shaft. At corresponding *d* split conditions, the wheels could turn and the vehicle could roll away. Also, in the case of a late request or in the case of an error when the vehicle is already rolling, it is not possible to safely engage the parking lock above approx. 3 km/h, while an EPB can function at any speed in principle.

11 Use and Safety Philosophy

11.1 Transparency of the Function

Transparency of the system responses is essential for acceptance of the ACC system. Only when users are quickly able to predict the system responses will they also use the system effectively. This presents the developer with the problem of making the control as simple as possible and sometimes leaving out features which an experienced user and of course the developer would value. As the driver gives up a part of the vehicle management task to the system when the ACC function is active, and needs only to monitor this, the transparency of the system plays an important role. Because current ACC systems only perform a part of the longitudinal control, it is recommended and in fact necessary to sample system limits during normal use of the system so that they are reached and/or exceeded with a certain regularity. This ensures that the driver knows the system limits at any one time and is versed in reassuming control from the system.

Adaptive cruise control is not a safety function. It is designed first and foremost to increase driver comfort. Of course, a comfort system should not pose any dangers; therefore, the ACC system must guarantee a level of safety corresponding to this requirement. Fault tree analyses have shown that dangerous situations can only occur if the driver does not employ intervention options. Two consequences then ensue:

1. The driver must not find resuming of control too demanding. In particular, the driver must recognize the need to resume control and take the corresponding steps and select the correct operating method in sufficient time.
2. The driver control take-over option must also be fault tolerant so that this option, e.g., switching off the control, strong deceleration, or strong acceleration, can only be blocked in an extremely improbable manner.

Prompt recognition of the need to take over the controls is derived from the driver's perception gained from past experience. In particular, too much confidence in the technology because only error-free functioning is experienced would be problematic because the driver would be unprepared both for the occurrence and also the response. In the case of ACC, this difficulty is not an issue because, as explained above, it is impossible to perfect its function. This negative aspect therefore has the benefit that the driver is permanently trained for the error situation. The driver remains aware that she/he may have to resume control in the case of unwanted performance and is well practiced in when and how this is carried out.

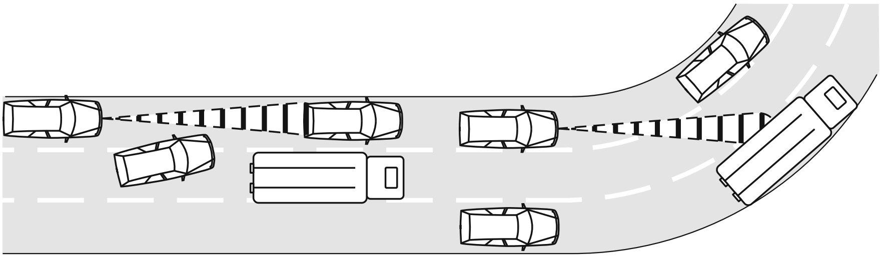


Fig. 24 Typical problem situations: *left*, late reaction to driver cutting in; *right*, difficulty in assigning objects when entering a bend

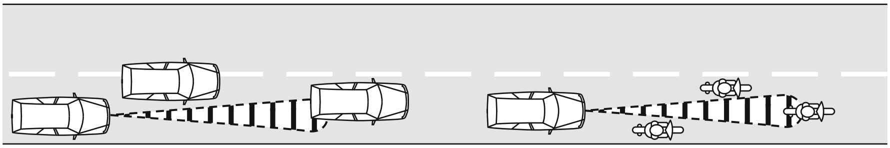


Fig. 25 Typical problem situations: ambiguity in the case of off-center motor vehicles and motorcycles

11.2 System Limits

Active sensors such as RADAR or LIDAR sensors offer precise detection of distance and relative speed, and at least the RADAR sensors are largely robust to weather influences. Also, due to the limited opening angle and the difficulties involved in the lane assignment of the detected object, especially in curve situations, there are limitations that sometimes lead to unexpected or incomprehensible system responses which must be explained to the user via suitable media.

Because of the narrow detection range of the ACC sensors, drivers cutting in directly in front of the actual vehicle are recognized late (Fig. 24 left). The assignment of the detected objects is still problematic when entering a bend, especially when curve travel cannot yet be detected because of the imminent vehicle signals (steering wheel angle, yaw rate) (Fig. 24 right).

We can expect help from the use of cameras which are able to detect lane markings and by using information from modern navigation systems about the expected route of the road. Extremely off-center driving methods can also lead to failures in detection. Particularly, in the case of motorcycles, detection is a problem due to their narrow silhouette (Fig. 25).

Some of the weaknesses above relate to first-generation ACC systems and have been at least partially compensated for by the wider focal range of sensors of the successive generation or by the use of additional sensors with a smaller reach but a significant lateral detection range, as are increasingly used in FSRA systems.

12 Safety Concept

The failure tolerance of the driver control take-over option has been eased by the distribution of the system. One actual example is the reading of the brake pedal switch both by the engine control and by the ESC. On detection of pressure on the brake pedal, the torque request of the ACC longitudinal control is ignored by the engine control. Deceleration requests to the brake control are also suppressed when pressure is applied to the brake pedal. Both the actuation of the brake pedal and the accelerator pedal are detected as redundant, so that both the actuation and the subsequent response states for single faults are protected even if the ACC-ECU or the data network breaks down.

Since the partitioning of the tasks can vary greatly, as the abovementioned examples show, there are no real standard solutions. Instead, the guaranteed driver intervention capability has to be demonstrated by a fault tree analysis.

Along with the permanent availability of the driver intervention options, inherent safety of the ACC system is essential. Again, the dislocation of the system proves to be an advantage. For example, the ESC system as an evidently inherently fail-safe system can be used to monitor the ACC control. If one selects the resulting vehicle acceleration as the monitoring value, all theoretically possible fault sources are included. Since first-generation ACC systems have kindly small operation limits of acceleration, generally +1 m/s^2 and/or −2 m/s^2, this kind of acceleration monitoring is easy to implement. The only disadvantage is that, in this kind of monitoring, the acceleration and/or deceleration is applied to the vehicle for a short time before it is suppressed by the ESC. However, the limits can be selected to ensure that over 95 % of normal drivers are able to cope with this.

13 Users and Acceptance Studies

From the very outset, the development of adaptive cruise control (ACC) has been accompanied by tests with test-persons. The first large-scale test was conducted by TÜV Rheinland at the beginning of the 1990s (Becker and Sonntag 1993). It addressed the general questions about the handling and acceptance of the function, which was still in its infancy. Subsequently, several basic variants with different deceleration capability and different time gaps were analyzed (Nirschl and Kopf 1997; Nirschl et al. 1999). Somewhat later, UMTRI conducted a very extensive field test (Fancher et al. 1998), which allowed long-term predictions to be made for the first time, even though the technical basis used was a long way from that used in series production today. Near-series systems were investigated in Abendroth (2001), Filzek (2002), and Weinberger (2001). In addition to this, the industry conducted other studies with ACC, but these were never published. In total, a large number of results (see also (Winner et al. 2003)) were collated, and only some of these are presented here for a few selected categories.

13.1 Acceptance

The results of the testers in all studies of acceptance carried out so far are unambiguous.

Becker and Sonntag (1993) describe how the testers in the pilot study perceive driving with ACC subjectively as safer, more relaxing, and less stressful than manual driving. This conclusion was arrived at despite the prototype status of the test vehicles, which exhibit certain serious sensor deficits. Nevertheless, the expectations of the system of the test participants were fully met and at times exceeded. It is clear, therefore, that the results of the testers with respect to acceptance and comfort at this level of maturity of ACC are largely robust.

Even with ACC systems without brake intervention, testers in the UMTRI study expressed themselves highly satisfied, which Fancher et al. (1998) attribute to the reduction of “throttle stress.”

By investigating the quality of processing ancillary tasks, Nirschl and Kopf (1997) detect less mental stress on the driver when using ACC. In subjective statements, high acceptance is expressed and ACC is seen more as a comfort than a safety system.

And with global satisfaction and acceptance on the part of drivers, Weinberger (2001) analyzes the temporal trends in long-distance journeys. All aspects such as “enjoyment of the system,” “intuitiveness in use,” “confidence in the controls,” “feeling of well-being,” and “stress” are rated in principle as good to very good. Over the duration of the trial, initial euphoria is replaced by a phase of relative disenchantment which nevertheless results at the end of the test in a significantly better evaluation than at the start.

13.2 Use

The object of any investigation is the time-gap behavior of drivers in comparisons between manual driving and driving with ACC. In straightforward following situations (Abendroth 2001), the average minimum time gap is 1.1 s with both manual driving and ACC. In contrast, Becker and Sonntag (1993) find a proliferation of time gaps of 1.7 s with manual drivers, albeit with significant scatter of results. A possible explanation to this may be the more winding test route. In ACC mode, the average time gap is 1.5 s, which was specified in the pilot study as a basic setting. Filzek (2002) finds that when given free choice of levels of 1.1, 1.5, and 1.9 s, testers choose an average ACC time gap of 1.4 s.

A significantly shorter average time gap of 0.8 s in manual driving is reported by Fancher et al. (1998). This apparent contradiction reveals the difficulty of transference between studies carried out in different traffic systems, in this case, the USA and Germany.

Significantly, in all studies, polarization takes place with respect to the selected ACC time gap. While at the start, the testers “play” with the levels, the frequency of adjustment decreases over the duration of the test. Respectively, around half of the testers then choose either slightly lower or slightly higher levels. With respect to the frequently selected short time gaps, a limit to at least 1.0 s appears to make sense for safety reasons.

Fancher et al. (1998) investigated the time-gap behavior in more detail and found that similar levels of 1.1, 1.5, or 2.1 s were selected depending on the age of the tester, i.e., older drivers selected correspondingly larger ACC time gaps.

Both Filzek (2002) and Fancher et al. (1998) report that significantly fewer drivers choose very small time gaps of below 0.6 s with ACC (Fancher et al., 6 out of 108 testers).

Mercedes-Benz market researchers gad interviewed customers in the USA about the use of DISTRONIC (see Fig. 26). The details relate to the S-class (W220, 1998–2005) and the SL (R230, since 2001). The rate of use in multiple-lane highways is expected to be substantially higher than in other road categories. The discrepancies between sports cars and saloons are surprisingly low in terms of the rate of use. Differences in types of use are somewhat greater. Since in the case of the DISTRONIC operating concept, the time gap stays at the old value purely mechanically, changing the time gap is only necessary if there is a reason for a change. Little or no use is made of this option. The distance setting is usually specified as average.

13.3 Compensating Behavior

Becker et al. (1994) studied the compensating behavior of drivers by evaluating the time gaps during which complex secondary tasks had to be processed in parallel. While testers automatically observe greater time gaps when driving manually, they do not change the desired time gaps for ACC operation. An analysis of visual distraction shows that drivers looked away for longer periods when driving with ACC, with a maximum of eight seconds being stated. It is worth noting that drivers subjectively perceive a lower safety risk when driving with ACC than without it. The authors conclude that, because of this riskier driving behavior, automated distance control can only be expected to increase safety if the technical system is capable of handling safety-critical situations better than the average driver.

13.4 Habituation Effects

Investigations conducted by Weinberger et al. (2001) with frequent drivers (>1000 km/week) indicate that steady driver behavior can only be assumed after they have been using ACC for 2 weeks. The characteristics used for determining the learning period were subjective assessment of the simplicity of operation and transparency of control take-over situations, as well as measuring the time (related to the time to collision, TTC) of the driver's intervention in control take-over situations, using a data recorder. This reveals that drivers not only have different driving styles but

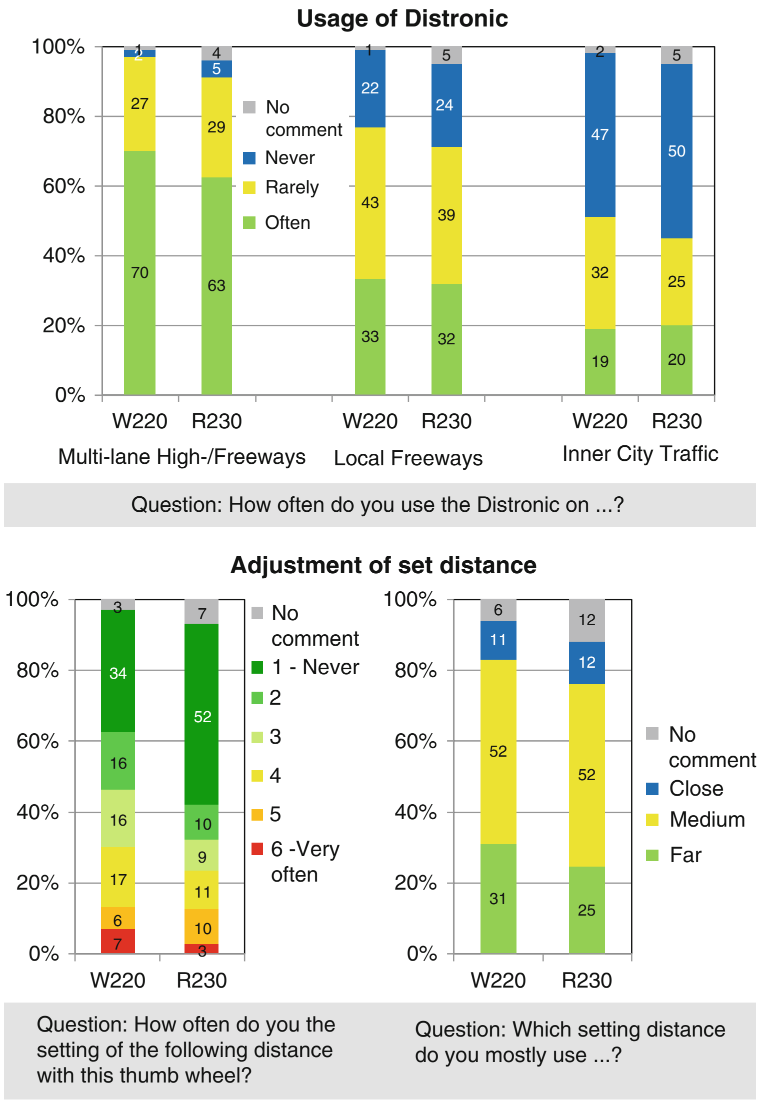


Fig. 26 Information about the use of an ACC system in the USA based on the DISTRONIC (Source: Mercedes-Benz Market Research 2005)

also different learning strategies. Drivers who consider themselves to be more sporty tended to intervene later, i.e., at shorter TTC, at the beginning of the tests than at the end, in order to establish the limits of the system, whereas drivers who considered themselves to be more cautious started off with an early “mistrustful” intervention and gradually started to intervene later as the learning phase progressed.

In summary, the above characteristics can only be assumed to be representative of stable driver behavior once this learning phase has been completed. The results of an assessment made after only a short time are only partially transferable to the main period of use, at least for the above characteristics.

Similarly, Nirschl and Kopf (1997) confirm a reduction in the mental load on the driver, which is consistent with refinement of the mental model over the period of use.

13.5 Driver Control Take-Over Situations

The principle simplicity of the driver's mental perception of ACC according to Becker et al. (1994) is also attributable to the fact that a correct response to system limits in control take-over situations is possible even after using the system for a very short time. Fancher et al. (1998) explain that 60 % of testers felt able to recognize control take-over situations promptly and correctly after just 1 day. 95 % of testers agreed to this statement after 1 week.

Nirschl et al. (1999) also report that most testers were able after a short time to estimate which ACC situations required intervention. However, the average of the three analyzed ACC variants with a somewhat smaller brake delay of 1 m/s^2 leads to greater uncertainty in the estimation than the variants with stronger braking intervention and/or with no braking intervention.

Weinberger (2001) explains that the estimation of control take-over situations was perceived subjectively by testers as noncritical and was perceived as simpler by the drivers as the period of use extended. The testers also stated that decisions are particularly easy in those situations which ACC cannot manage in principle (e.g., braking behind a halted vehicle). It was shown that after the driver takes over the controls, in 80 % of cases, the mean deceleration of the vehicle is below or equal to 2 m/s^2. This period is also covered by ACC so that one may conclude that these situations are also noncritical from an objective standpoint.

With a few exceptions, the first ACC vehicle tests show a uniform trend although there would be sufficient reason to justify differentiation in the results:

- The technology of the systems analyzed differed considerably both in terms of the scope of functions and level of maturity.
- The traffic conditions in the USA are only partly comparable with those in Europe.
- Short-term and long-term tests were carried out, and in the long-term tests, clear learning effects were found which invalidate the predictions of some of the results of short-term tests.

Obviously, at least in its basic functions, ACC appears to be robust with respect to the stated differences in the performance of the tests. The core function was understood by the drivers at the start, irrespective of the limitations of the predecessor systems.

With respect to the FSRA control take-over situation, Neukum et al. (2008) analyzed a crossroad problem situation. A vehicle which had been at a standstill for a long period at a crossroad is initially concealed by the target vehicle and then overtaken shortly before the approach so that it is suddenly in the driving corridor of the FSRA vehicle. As a static vehicle which has not yet moved into the focal range of the RADAR, this was not accepted by the FSRA system as a control object, i.e., the driver must intervene and brake in order to avoid a collision. All testers were able to do so without the co-drivers (present for this purpose) having to initiate the action. Nevertheless, this situation is rated by many drivers as threatening when encountered for the first time.

13.6 Comfort Assessment

The focus of the analyses documented in Didier (2006) was the investigation of comfort. For this purpose, two vehicles from different manufacturers with different ACC systems were driven by a total of 36 test persons. The subjective assessments, which were obtained by means of a questionnaire on selected comfort criteria, were compared. Although the systems in both vehicle series carried out the same functions, it was possible to detect even slight differences between the two systems with respect to comfort. However, the simultaneous analyses of the objective quantitative characteristics "Frequency of driver override by actuation of accelerator" and "Interruption of the control by driver brake intervention" could not be adequately equated with the comfort evaluation.

13.7 Effectiveness Analyses

Usually, commercially available ACC systems are offered in combination with collision avoidance or collision mitigation systems. The reason for this is that the comfort and safety systems share the surrounding sensing system (usually a RADAR sensor). The analysis of the effectiveness of ACC systems to increase road safety based on real accident figures always takes account of the contribution of collision avoidance or collision mitigation systems in these cases. An analysis of the ordering of spare parts for all Mercedes-Benz S-Class models in Germany over a period of 3.5 years shows that fitting the vehicles with DISTRONIC PLUS systems reduced the number of forward collisions by 35 % (Schittenhelm 2013). As is shown in Fig. 27, the real field numbers are surprisingly close to the calculated forecasts (Schittenhelm 2013).

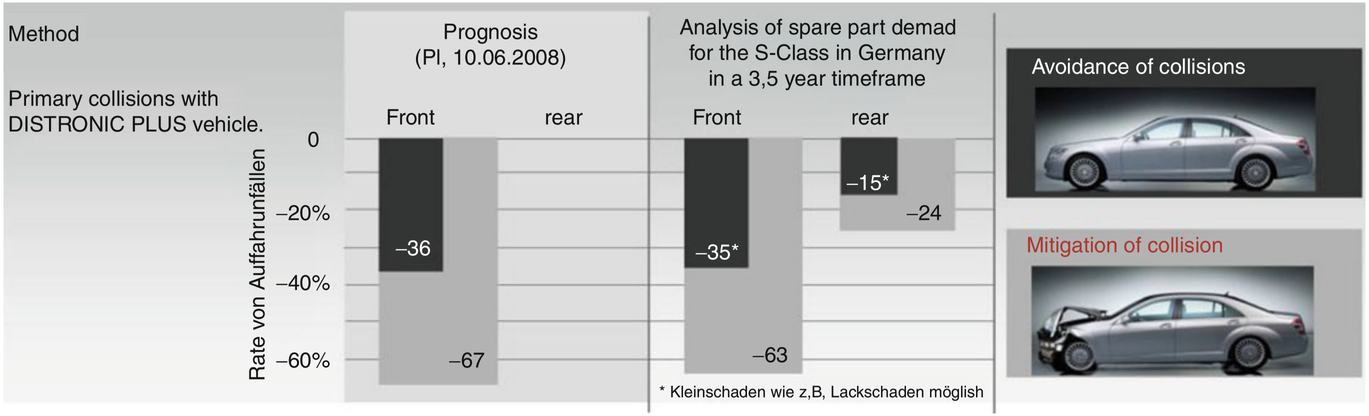


Fig. 27 Analysis of the effectiveness of DISTRONIC PLUS in increasing driving safety using the example of the Mercedes-Benz S-Class (Schittenhelm 2013)

14 Conclusion and Outlook

14.1 Current Developments

With the series introduction of FSRA, everyday traffic situations are covered by the function range. Now that ACC has also reached classes of vehicle primarily fitted with manual shifts in Europe, it is not clear whether FSRA will stumble at this hurdle and be limited to the existing niche of vehicles with automatic gearboxes or whether it is capable (if necessary, flanked by energy-saving measures) of increasing the proportion of vehicles with automatic gearboxes.

Essentially, alternatives to the currently prevailing RADAR sensor principle are also conceivable in future. These include LIDAR systems and camera systems. Although the latter will certainly not have the functional scope that is standard today, they could be a sound option for other applications because of their multiple uses.

14.2 Function Enhancements

Although FSRA already relieves drivers of a lot of work, there is naturally an interest in it also taking over lateral guidance. For the low-speed range, traffic jam assist (▶ Chap. 51, “Traffic Jam Assistance and Automation”) is the follow-on system, with which driving can be completely automated, even if measures such as hands-on monitoring are required to ensure that the driver monitors this automatic function adequately. In technical terms, stereo cameras and RADAR or scanning LIDAR are suitable for this task.

In the medium- and high-speed range, the lane-keeping function and ACC combine to become a third-generation vehicle guidance assistant and are activated or deactivated together. Since its introduction in the Mercedes-Benz S-Class (W222), such functionality is now also available in other models. This assistance follows the philosophy of the two preceding systems and restricts itself to “gentle” comfort-oriented interventions, whereby parallel protective functions (forward collision warning, automatic emergency braking, lane change warning, lane departure warning) flank and safeguard this functionality if limits are reached where gentle interventions are no longer adequate.

References

Abendroth B (2001) Gestaltungspotentiale für ein PKW-Abstandsregelsystem unter Berücksichtigung verschiedener Fahrertypen (Design potentials for an automotive distance control system under consideration of different types of drivers), Dissertation TU Darmstadt. In: Schriftenreihe Ergonomie. Ergonomia-Verlag, Stuttgart

Ackermann F (1980) Abstandsregelung mit Radar (Distance control with radar). In: Spektrum der Wissenschaft Juni 1980. pp 24–34, Heidelberg

Becker S, Sonntag J (1993) Autonomous intelligent cruise control – Pilotstudie der Daimler-Benz und Opel Demonstratoren. In: Prometheus CED 5, TÜV Rheinland. Köln

Becker S, Sonntag J, Krause R (1994) Zur Auswirkung eines Intelligenten Tempomaten auf die mentale Belastung eines Fahrers, seine Sicherheitsüberzeugungen und (kompensatorischen) Verhaltensweisen (Effects of an intelligent cruise control on the mental load of drivers, their safety conviction and (compensatory) behaviour). In: Prometheus CED 5, TÜV Rheinland. Köln

Didier M (2006) Ein Verfahren zur Messung des Komforts von Abstandsregelsystemen (ACC-Systemen) (A method for measuring the comfort of distance control systems (ACC systems)), Dissertation TU Darmstadt. In: Schriftenreihe Ergonomie. Ergonomia-Verlag, Stuttgart

Fancher P et al (1998) Intelligent cruise control field operational test. University of Michigan Transportation Research Institute (UMTRI) final report, Michigan

Filzek B (2002) Abstandsverhalten auf Autobahnen – Fahrer und ACC im Vergleich (Distance behaviour on autobahns –driver and ACC – a comparison), Dissertation TU Darmstadt. In: VDI Fortschritt-Berichte Reihe 12, Nr. 536. VDI-Verlag, Düsseldorf

Furui N, Miyakoshi H, Noda M, Miyauchi K (1998) Development of a scanning laser radar for ACC. Society of Automotive Engineers SAE paper no. 980615, Warrendale

ISO TC204/WG14 (2009) ISO 22179:2009 Intelligent transport systems – full speed range adaptive cruise control (FSRA) systems – performance requirements and test procedures, Geneva

ISO TC204/WG14 (2010) ISO 15622:2010 Transport information and control systems – adaptive cruise control systems – performance requirements and test procedures, Geneva

Luh S (2007) Untersuchung des Einflusses des horizontalen Sichtbereichs eines ACC-Sensors auf die Systemperformance (Investigation of the impact of the horizontal detection area of an ACC sensor on the system performance), Dissertation TU Darmstadt. In: VDI Fortschritt-Berichte Reihe 12. VDI-Verlag, Düsseldorf

Meyer-Gramcko F (1990) Gehörsinn, Gleichgewichtssinn und andere Sinnesleistungen im Straßenverkehr (Auditory, vestibular and other perception sensories in road traffic). In: Verkehrsunfall und Fahrzeugtechnik, Nr. 3. pp 73–76, München

Mitschke M, Wallentowitz H, Schwartz E (1991) Vermeiden querdynamisch kritischer Fahrzustände durch Fahrzustandsüberwachung (Avoidance of lateral-dynamics-critical driving states by driving state monitoring). In: VDI Bericht 91. VDI-Verlag, Düsseldorf

Neukum A, Lübbeke T, Krüger H-P, Mayser C, Steinle J (2008) ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen (Driver behaviour at functional system limits). In: Proceedings 5. Workshop Fahrerassistenzsysteme, Walting

Nirschl G, Kopf M (1997) Untersuchung des Zusammenwirkens zwischen dem Fahrer und einem ACC-System in Grenzsituationen (Investigation of the interaction between the driver and an ACC system in borderline situations). In: Conference Der Mensch im Straßenverkehr, Berlin 1997, VDI Bericht 1317, VDI-FVT, VDI-Verlag, Düsseldorf

Nirschl G, Blum E-J, Kopf M (1999) Untersuchungen zur Benutzbarkeit und Akzeptanz eines ACC-Fahrerassistenzsystems (Investigation of usability and acceptance of an ACC driver assistance system). In: Fraunhofer Institut für Informations- und Datenverarbeitung IITB Mitteilungen, Karlsruhe

Pasenau T, Sauer T, Ebeling J (2007) Aktive Geschwindigkeitsregelung mit Stop&Go-Funktion im BMW 5er und 6er (Active cruise control with Stop&Go function in the BMW 5- and 6-series). In: ATZ 10/2007. Vieweg Verlag, Wiesbaden, pp 900–908

Schittenhelm HN (2013) Advanced brake assist – real world effectiveness of current implementations and next generation enlargements by Mercedes-Benz, 23. In: Technical conference enhanced safety of vehicles, proceedings, paper-no. 13–0194; 23. ESV-conference, Seoul

Steinle J, Hohmann S, Kopf M, Brandstäter M, Pfeiffer A, Farid N (2006) Keeping the focus on the driver: the BMW approach to driver assistance and active safety systems that interact with vehicle dynamics. In: Proceedings of FISITA World Automotive Congress 22–27 Oct 2006 FISITA F2006D185, Yokohama

Watanabe T, Kishimoto N, Hayafune K, Yamada K, Maede N (1995) Development of an intelligent cruise control system. In: Proceedings of the 2nd ITS World Congress in Yokohama, pp 1229–1235

Weinberger M (2001) Der Einfluss von Adaptive Cruise Control Systemen auf das Fahrverhalten (The impact of adaptive cruise control systems on driving behaviour), Dissertation TU München. In: Berichte aus der Ergonomie. Shaker-Verlag, Aachen

Weinberger M, Winner H, Bubb H (2001) Adaptive cruise control field operational test – the learning phase. In: JSAE review, vol 22. Elsevier, Amsterdam, p 487

Winner H (2003) Die lange Geschichte von ACC (The long history of ACC). In: Proceedings Workshop Fahrerassistenzsysteme, Leinsweiler

Winner H (2005) Die Aufklärung des Rätsels der ACC-Tagesform und daraus abgeleitete Schlussfolgerungen für die Entwicklerpraxis (Unravelling the mystery of the ACC form-of-the-day and related conclusions for engineering practice). In: Proceedings Fahrerassistenz-workshop, Walting

Winner H, Luh S (2007) Fahrversuche zur Bewertung von ACC – Eine Zwischenbilanz (Driving tests for assessment of ACC – an interim conclusion). In: Bruder R, Winner H (eds) Darmstädter Kolloquium Mensch & Fahrzeug – Wie objektiv sind Fahrversuche? Ergonomia-Verlag, Stuttgart

Winner H, Olbrich H (1998) Major design parameters of adaptive cruise control. AVEC’98 paper 130, Nagoya

Winner H et al (2003) Fahrversuche mit Probanden zur Funktionsbewertung von aktuellen und zukünftigen Fahrerassistenzsystemen (Driving test with test persons for function assessment of present and future driver assistance systems). In: Landau, Winner (ed) Fahrversuche mit Probanden – Nutzwert und Risiko, Darmstädter Kolloquium Mensch & Fahrzeug, 3./4.April 2003, TU Darmstadt, VDI Fortschritt-Berichte Reihe 12, Nr. 557. VDI-Verlag, Düsseldorf

Witte S (1996) Simulationsuntersuchungen zum Einfluss von Fahrerverhalten und technischen Abstandsregelsystemen auf den Kolonnenverkehr (Simulation investigations into the impact of driver behaviour and technical distance control systems on traffic queues), Dissertation Universität Karlsruhe, Karlsruhe, p 23

Fundamentals of Collision Protection Systems

46

Hermann Winner

Contents

Abstract

For the high portion of forward collisions within the figure of accidents and related damages, basic counteracting strategies are derived. Accident prevention, reaction assistance, and emergency maneuver are the fundamental strategies for protection and are discussed here. With the given physical and empirical basis for control, strategies of countermeasures in different constellations enable the estimation of safety effect of different strategies and definitions of requirements for sensors and actuators.

H. Winner (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_47

1 Problem Definition

Accidents that occur when vehicles are traveling in a moving line of traffic represent the largest group of accidents and the second largest group of accidents involving fatalities and serious injuries. Systems to protect against this type of accident therefore offer a lot of potential (▶ Chap. 4, "Driver Assistance and Road Safety"). Figure 1 shows how preventive measures can be derived.

A priori there is no direct or immediate relationship between the accident and the presence of a disturbance. Starting from a preexisting latent hazard level, the disturbance raises this level, but, at first, there is a considerable margin before an accident actually occurs. This disturbance will only result in an accident, if time passes without any response being forthcoming or if the response is inappropriate. In contrast, a prompt and correct intervention by the driver can defuse the situation so that the critical situation only results in a near accident. From this structuralization of the accident scenario, it is now possible to derive three accident prevention strategies, which are outlined briefly below and then explained in detail in terms of their application for forward collision avoidance.

1. **Preventive assistance:** Preventive accident avoidance by reducing the latent risk and thereby reducing the probability of getting into a critical situation or at least providing an effectively longer period to react in the event of a disturbance.

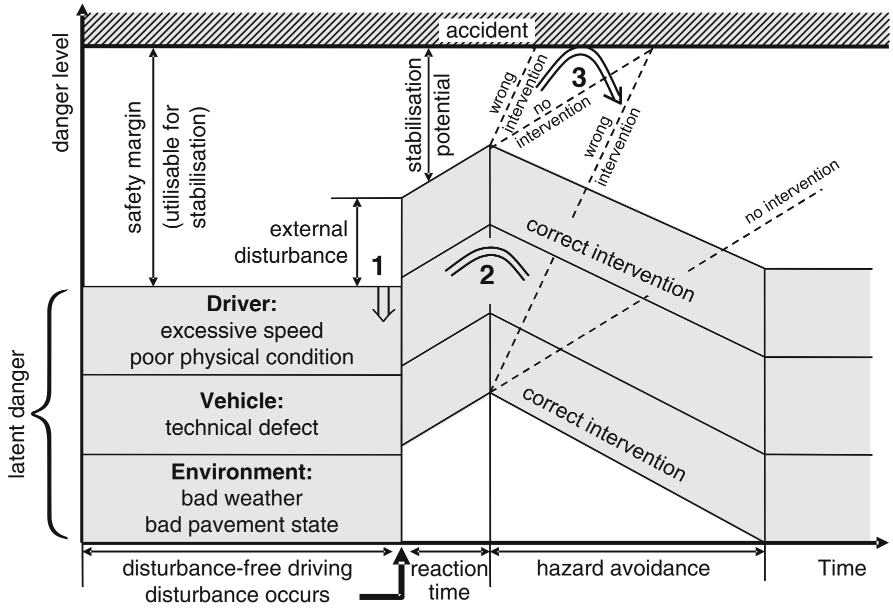


Fig. 1 Sequence of a critical road situation for deriving assistance strategies and action sequences to deal with a critical driving situation (as in Braun and Ihme (1983)). Numerals: areas of application of driver assistance systems (see text)

2. **Reaction assistance:** Accident prevention by means of assistance in critical situations so that the driver responds promptly and correctly. In practice, there are only two possible strategies for avoiding accidents when traveling in a line of traffic: braking or evasion. In this regard, driver assistance systems that stabilize the vehicle, such as ABS and ESP, have created the first prerequisite for the driver to be "willing" to take such action, because hazardous consequences in terms of vehicle stability, such as skidding, have already been taken care of and the driver is able to fully exploit the physical limits from the outset. Nevertheless, accident analyses (see ▶ Chap. 4, "Driver Assistance and Road Safety") (Kopischke 2000) and tests with volunteers (Bender 2008) have shown that this potential is not adequately exploited, if at all.
3. **Emergency maneuvers:** "Strong interventions" during the last few seconds before the accident, which avoid the accident by means of an emergency maneuver, if the driver has failed to respond promptly and appropriately, or which contribute to collision mitigation. Because of the cautious interpretation of the legal framework, in particular the Vienna Convention that applies in Germany (see ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems"), there was initially great reluctance, because of the legal implications, to launch accident avoidance systems that act in the last few seconds, if they can no longer be overridden by drivers. If active braking is not applied until "such time as evasion is objectively impossible" (see (Seeck and Gasser 2006)), there are no longer any legal reservations. However, since there is now greater acceptance of systems that prevent collisions by braking, and such systems are in demand according to consumer tests (▶ Chap. 11, "Test Methods for Consumer Protection and Legislation for ADAS"), systems have been developed that no longer wait until an accident is inevitable before they initiate strong braking.

2 Preventive Assistance

There are two main strategies for reducing the latent risk: increasing the available reaction time in terms of vehicle dynamics and increasing the driver's ability to overcome a disruptive situation. The latter is essentially a function of the driver's constitution and driving skills. Driving skills cannot be improved by driver assistance systems, but only by training, e.g., on a driving practice ground. The drivers' constitution, on the other hand, can be improved by relieving them of other demanding driving tasks such as, e.g., taking over control of the distance from the preceding vehicle by using ACC. The driver's constitution can be improved physiologically (less stress on the eyes) and also psychologically (more relaxed road awareness) (Weinberger 2001).

Of course, ACC is also a very good way of increasing the amount of time available for reacting. In the known investigations (see ▶ Chap. 45, "Adaptive Cruise Control"), ACC users choose longer time intervals than if they were setting the distance themselves. On the other hand, it is not clear whether reliance on ACC leads to later driver intervention or whether the response time is shortened by the early instigation of vehicle braking. Although early ACC versions were not really suitable for use in urban traffic, full speed range ACC systems are suitable also for this application, but, so far, there are no studies into their use and possible safety potential.

Like ACC, an active accelerator pedal (also known as a force feedback pedal) is similarly suitable for maintaining distance. This keeps the driver directly in the control loop. If the driver maintains a more or less constant force on the pedal, the accelerator angle changes in such a way that the time gap remains constant, thereby allowing distance to be maintained as with ACC, but without active braking.

3 Driver Reaction Assistance

Assisted response comprises the steps: attracting attention, clarifying the situation, and assisting with the intervention (see also ▶ Chap. 36, “Driver Warning Elements”). Since a critical situation during in-line driving is usually preceded by a lapse in attention (LeBlanc et al. 2001), changing this state is an essential prerequisite for taking the necessary subsequent action. The driver’s attention is usually explicitly attracted by warning elements but can also be implicitly attracted by unexpected control reactions on the part of the ACC. As already outlined in ▶ Chap. 36, “Driver Warning Elements,” warning strategies differ in terms of the amount of information they provide. A simple auditory warning is good at attracting attention but does not provide any information about the situation or the required response. This can be achieved by supplementary visual information or an auditory icon. A detailed description of the warning options, including the assessment of “forgiveness” of false warnings, is given in ▶ Chap. 36, “Driver Warning Elements,” and therefore we shall not discuss the different options further at this point.

If a critical situation arises during in-line driving, there are essentially two possible strategies for avoiding an accident: driving around the obstacle or stopping short of it. Based on certain initial parameters, the necessary intervention times for avoiding a collision are calculated in Sect. 6. In all practical situations there is a velocity above which evasion is calculated as the latest possible maneuver, whereas, below this velocity, the accident can still be successfully avoided by braking after the last possible evasion distance has been passed. However, consideration of both cases is based on an optimum maneuver. Whereas, so far, only one model of car (Lexus LS, since 2006) has provided an evasion assistance system, brake assist systems have been on the market since 1997 and are fitted in nearly all new cars in the form of the simple basic brake-pedal-triggered version. This basic function and additional functional enhancements are described in Sect. 5. What they all have in common is that the brake assist function can only be effective if the foot brake is actually pressed. However, since no braking action was taken in approximately one-third of the cases investigated in Kopischke (2000) and Bender (2008) and as many as half of those investigated in Wiacek and Najm (1999), in such cases the brake assist function is ineffective on its own. As was demonstrated in the studies with test persons (Bender 2008; Färber and Maurer 2005), automatic braking elicits a braking response from the driver. Thus, an automatic braking intervention can be used to stimulate the driver's decision. Possibilities for this are a brief jerk of the brakes and continuous partial braking. If the driver then initiates the brake assist function by pressing the foot brake, this produces braking with maximum deceleration or "target braking," providing the environment sensors allow such a function (see Sect. 5.2).

4 Emergency Maneuvers

If all the previously outlined warning stages have still failed to bring about an evasive or braking response by the driver, automatic emergency maneuvers with "strong" interventions within the last second before a predicted impact can prevent or mitigate the damage. At high differential velocities, automatic evasive maneuvers are more effective in avoiding accidents than braking maneuvers (as long as the evasive maneuver does not cause a more serious consequential accident). In the PRORETA project (see ▶ Chap. 57, "Anti Collision System PRORETA"), it was demonstrated how evasive maneuvers could be implemented, notably with an automatic steering impulse that is added to the existing steering angle and is dimensioned to allow the obstacle to be circumvented. According to the results of the study, this form of evasive maneuver can be regarded as acceptable. But the capabilities of current environment sensing systems are nowhere near what would be required to initiate an automatic emergency evasion.

In contrast, automatic emergency braking is already a commercially available technology. However, because of the legal constraints mentioned above, this is only activated when evasion is no longer a feasible option. In all known applications, initiation of automatic emergency braking is the last step in an action chain and is only triggered if the previous warning stages have failed to bring about a braking or steering response or if the mechanical system detects that the emergency situation does not allow sufficient time for a driver to respond.

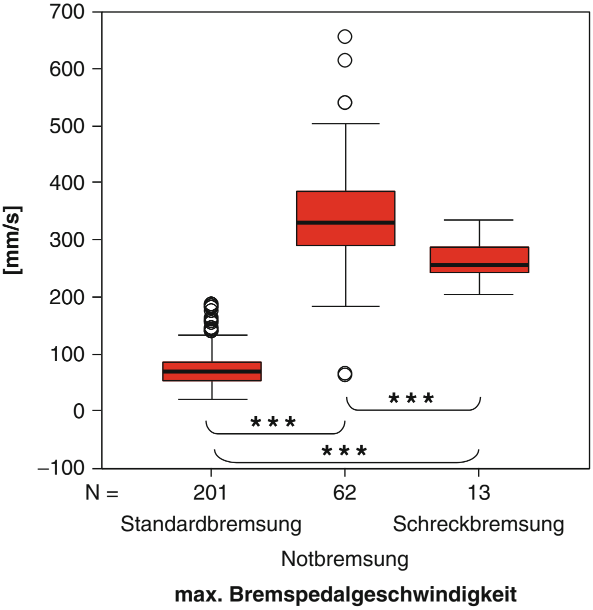
The effectiveness of the different designs of automatic emergency braking systems is theoretically derived and experimentally proven in Sect. 6.3.

5 Braking Assistance

5.1 Basic Function

Accident analyses (Kopischke 2000) and tests in a driving simulator (Kiesewetter et al. 1997) or on a test ground (Weiße 2003) show that many emergency braking maneuvers are not ideal for ensuring a short stopping distance. After a phase of steep pressure increase, subsequent pressure buildup is often hesitant. The function of the brake assist system is derived from this: As soon as the brake assist system

Fig. 2 Box plot of brake pedal velocities for standard ("Standardbremsung"), emergency ("Notbremsung"), and startle braking ("Schreckbremsung") (Weiße 2003)



(BAS) detects the intention to apply emergency braking, it builds up the maximum deceleration as quickly as possible and maintains it until it detects that emergency braking is no longer desired. Weiße (2003) quantifies the potential achievable with such a system as an average reduction in stopping distance of 8 m, or 20 %, when emergency braking from a velocity of 100 km/h. The relative percentage is even higher at lower speeds, which also explains why it is particularly beneficial for unprotected road users (Busch 2005).

The initial velocity of the brake pedal and the subsequent rise in brake pressure are useful basic criteria for identifying an emergency braking intervention. Normal braking situations and emergency braking situations are differentiated on the basis of empirically determined values. The switching thresholds relate to brake pedal velocity and are measured directly via the brake booster diaphragm travel or indirectly via pressure sensors on the master brake cylinder. They vary as a function of traveling speed and master cylinder pressure or brake pedal travel. Although erroneous decisions cannot be ruled out, this criterion is far superior to all other criteria that can be derived from foot movement (Weiße 2003). The standard braking maneuvers measured under real road conditions came nowhere near the brake pedal velocities (see Fig. 2) that characterize emergency braking. Nevertheless, similar brake pedal velocities were reached in the startle-response braking interventions referred to by Weiße (2003), without full braking being necessary for this situation. These "side effects" can be managed by pulling back the brake pedal, without causing any great problems or having any effect upon following traffic, and are similar to the responses observed when a driver changes from a vehicle requiring higher brake application energy to another vehicle with a "more responsive" brake pedal.

The control systems necessary for the brake assist function are part of modern braking systems and operate with pneumatic or electrical auxiliary energy to provide the brake caliper clamping force required for maximum deceleration. ▶ Chaps. 29, "Hydraulic Brake Systems for Passenger Vehicles" and ▶ 30, "Electro-Mechanical Brake Systems" outline the necessary force buildup techniques that are used today. From a functional point of view, it is not really important how the additional clamping force is achieved. There are still differences in buildup dynamics.

Since the main initiating criterion is brake pedal velocity, the basic function of the BAS can also be performed with mechanical control by means of the brake booster. A much greater brake boost is initiated when the piston rods are moving faster, so that more braking force is developed with the same pedal position than with "slow" brake application. This function is enhanced by the options of building up pressure via the ESC hydraulic pump (▶ Chap. 39, "Brake-Based Assistance Functions").

5.2 Further Developments

In the dissertation by Weiße (2003), which we have frequently quoted here, you will also find other approaches for triggering the brake assist function, in particular, to initiate it at an earlier stage. However, no criterion was found that offered anywhere near the same decision-making quality as brake pedal velocity. With a combination of criteria in the sense of an OR operation, it should be possible to increase the potential from 8 to 11 m. The same should be possible with a graduated function, operating in three stages – preconditioning, pre-braking (at 3 m/s^2), and full braking. However, for pre-braking the foot movement needs to be measured longitudinally to the vehicle, and this is not easy to do. Lowering of the trigger threshold when exceeding an accelerator pedal velocity threshold and of full braking triggered by brake pedal velocity produces a comparatively small gain of 0.6 m, to give a total of 8.6 m.

Any further improvement in the brake assist function would require bringing the start of braking forward. There are two possible strategies for doing this:

- Shortening the changeover time from accelerator to brake pedal. In many experiments (Morrison et al. 1986; Weiße 2003; Kopischke 2000; Hoffmann 2008) a consistent changeover time of approximately 0.2 s was achieved. This changeover time could only be reduced further by using an alternative operating concept, as has been demonstrated in Eckstein (2001) with a longitudinally isometric sidestick.
- Lowering the thresholds on detection of an emergency braking situation based on surroundings sensors.

However, if a situation has already been detected, the obvious solution is to make the degree of assistance dependent upon the distance still available, therefore

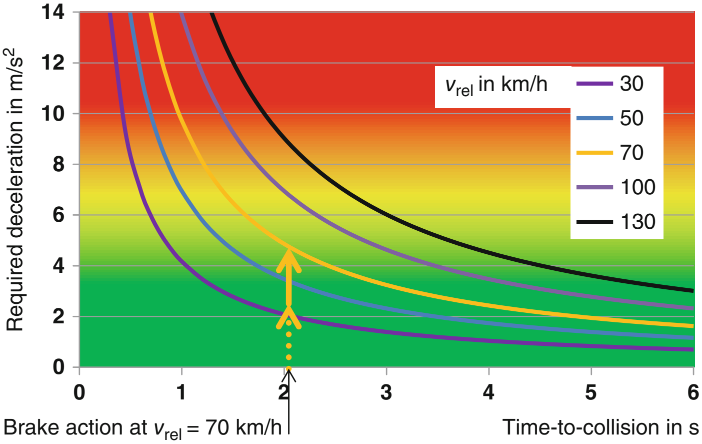


Fig. 3 Average deceleration required as a function of time-to-collision for different initial differential velocities

only to generate enough additional deceleration to ensure that the vehicle slows down in time. The required deceleration $D_{\text{req}} = v_{\text{diff}}^2/2d$ as a function of time-to-collision $t_{\text{tc}} = d/v_{\text{diff}}$, which in turn is formed from the distance d and the differential velocity v_{diff} between the subject vehicle and the obstacle, is shown in Fig. 3 for different initial differential velocities. Of course, the reduction in deceleration does not shorten the stopping distance, but this assistance can be useful in the run-up to a critical situation, provided that the driver assesses the situation as being less critical than it actually is and does not brake sufficiently, thereby reducing the margin available for slowing down in time. For example, it is assumed in the example shown in Fig. 3 that, at an initial differential velocity of 70 km/h and a t_{tc} of barely 2 s, the driver only applies braking of 3 m/s^2. Deceleration is now increased by the adaptive brake assist function to at least 5 m/s^2, so that, at this deceleration, the remaining distance can be used for steady deceleration.

6 Warning and Intervention Times

In the subsections below we derive warning and intervention times based on vehicle dynamics and driver behavior, which are suitable for different forward collision prevention measures. Essentially there are two ways of approaching the initiation of a preventive measure:

- Time criteria (time-to-collision, time-threshold-evasion, time-to-stop, time-threshold-brake; time margin required to avoid a collision)
- Acceleration criteria (longitudinal deceleration or lateral acceleration required to avoid a collision)

The current distance, velocity, and acceleration values are used for the time criteria, and then compared with threshold values derived from assumptions about the maximum possible decelerations and lateral accelerations, these threshold values are being increased by the assumed response time when warnings are given. In terms of vehicle dynamics, the acceleration criteria are extremely simple, since it is sufficient to compare them with the assumed maximum acceleration values. The advantage is lost if a response time, as is required for warning strategies, or a system downtime is included in the calculation. Since both of these "criteria sets" have advantages and disadvantages when it comes to representation, the relevant equations and threshold values are presented below for both approaches, even if ultimately both approaches can be converted into each other. Time analysis produces simpler terms for calculating evasion, and acceleration criteria are simpler to model for braking. The results for the individual criteria are referenced in a summary table at the end of this section.

6.1 Vehicle Dynamics Analyses

When analyzing vehicle dynamics, a distinction is made between three different cases. The simplest case assumes an un-accelerated obstacle moving at constant velocity (incl. the special case of a stationary obstacle). Criteria are then derived for a vehicle moving at a constant relative acceleration to the subject vehicle. In the third special case, the deceleration of the obstructing vehicle brings it to a standstill before the subject vehicle reaches it. A further approach looks at the opportunity for evasion with combined longitudinal and lateral motion of the subject vehicle.

6.1.1 Calculations for an Un-accelerated Obstacle

Braking Maneuvers

Even if the obstacle is moving at constant velocity, all calculations can relate to a system relative to this object. Thus, all the results relate to the case of a stationary obstacle, the negative relative velocity $-v_{\mathrm{rel}}$ replacing the absolute velocity $v_{\mathrm{x,v}}$ as velocity difference v_{diff} and the distance d replacing the absolute travel s. The distance to standstill $s_{\mathrm{B}}(v_{\mathrm{xv,0}})$ from initial velocity $v_{\mathrm{xv,0}}$ can therefore be equated to the stopping distance $d_{\mathrm{B}}(v_{\mathrm{diff}})$ necessary to offset the equal velocity difference v_{diff}. The subsequently introduced time-to-stop relates accordingly to the time required to equalize the velocities.

With a good approximation for modern braking systems, the stopping distance is as follows:

$$d_{\mathrm{B}}(v_{\mathrm{diff}}) = v_{\mathrm{diff}} \cdot \tau_{\mathrm{B}} + \frac{v_{\mathrm{diff}}^2}{2D_{\mathrm{max}}} \tag{1}$$

The time delay of the brake τ_B takes account of the effective time lost during buildup of the deceleration. If deceleration is assumed to ramp up linearly within a time τ_S up to mean full braking D_{max}, the time delay can be approximated by half the ramp-up time (i.e., $\tau_S/2$), without the error in calculated stopping distance exceeding more than a few centimeters. The brake response time, which is in the range of 50 ms, is added to the much higher driver response time τ_R, which is itself made up of the time-to-look, time-to-respond, and changeover time and is estimated to be between 0.5 and 1.5 s. For the example calculations, this value is set at $\tau_R = 1\text{s}$, which is far too high for situations that are clear to the driver (cf. Kopischke 2000; Hoffmann 2008).

The warning distance d_{warn} for timely emergency braking adds the driver response time τ_R to the time delay of the brake in Eq. 1:

$$d_{warn}(v_{diff}) = v_{diff} \cdot (\tau_B + \tau_R) + \frac{v_{diff}^2}{2D_{max}} \tag{2}$$

By relating the current distances d to the differential velocity v_{diff}, we obtain the time-to-collision (TTC) variable

$$t_{tc} = \frac{d}{v_{diff}}; d, v_{diff} > 0 \tag{3}$$

and Eqs. 1 and 2 can be simplified as follows:

$$t_{tB}(v_{diff}) = \tau_B + \frac{t_{ts}(v_{diff}, D_{max})}{2} \tag{4}$$

with time-threshold-brake t_{tB} and time-to-stop

$$t_{ts}(v_{diff}, D_{max}) = \frac{v_{diff}}{D_{max}} \tag{5}$$

and accordingly

$$t_{warn}(v_{diff}) = \tau_R + t_{tB}(v_{diff}) \tag{6}$$

The time interval required for full braking, or time-to-stop, t_{ts} is double the TTC at the start of braking. This principle also applies for the subsequent analyses, as long as constant and positive relative deceleration can be assumed.

Apart from considering time intervals and spatial distances, it is also possible to determine the currently required deceleration D_{req} and use it as a threshold. For the simple case of an obstacle moving at constant velocity, this is determined as follows:

$$D_{req,v} = \frac{v_{diff}^2}{2d} \tag{7}$$

Evasive Maneuvers

The evasion distance d_{eva} is calculated from the product of velocity difference and the time interval required for evasion t_{eva}, which in turn can be approximated from the deviation offset necessary for evasion y_{eva}, the average maximum lateral acceleration $a_{y,\max}$, and the time delay of steering τ_{S}, which, like the time delay of the brake, is of the order of 0.1 s:

$$t_{\text{eva}} = \sqrt{\frac{2y_{\text{eva}}}{a_{y,\max}}} + \tau_{\text{S}} \tag{8}$$

$$d_{\text{eva}} = v_{\text{diff}} \cdot t_{\text{eva}} \tag{9}$$

Depending on the type of tires, the maximum lateral acceleration $a_{y,\max}$ is between 80 % and 100 % of the maximum deceleration $D_{\max}$, which in turn is approximately 10 m/s^2 on dry road surfaces (hereafter a ratio of $a_{y,\max}/D_{\max} = 90\,\%$ is assumed). For narrow obstacles, an evasive offset of 1 m can be assumed and, for larger ones such as trucks, 1.8 m. Thus we obtain values of 0.55–0.7 s for t_{eva}. Below we have assumed a value $t_{\text{eva,phys}} = 0.6\,\text{s}$ as being representative of vehicle dynamics considerations. Of course, this is too high if a much smaller evasive offset is required, for example, when the obstacle is situated to one side of the direction of travel.

In the same way as the required deceleration, the required lateral acceleration $a_{y,\text{req}}$ can be calculated, this criterion now being codetermined by the evasive deflection, as well as distance and differential velocity:

$$a_{\text{y,req}} = 2y_{\text{req}}t_{\text{tc}}^{-2} = \frac{2y_{\text{eva}}v_{\text{diff}}^2}{d^2} \tag{10}$$

6.1.2 Calculations for a Constantly Decelerating Obstacle

Braking Maneuvers

For an obstacle moving with constant deceleration D_{obs}, the TTC depends upon the relative deceleration $D_{\text{rel}} = D_{\text{obs}} - D_{\text{sub}}$ relative to the vehicle behind:

$$t_{\text{tc}}(D_{\text{rel}}) = \frac{\sqrt{v_{\text{diff}}^2 + 2D_{\text{rel}}d} - v_{\text{diff}}}{D_{\text{rel}}};\; v_{\text{diff}}^2 > 2D_{\text{rel}}d \tag{11}$$

$t_{\text{tc}}(D_{\text{rel}})$ is also referred to as the enhanced time-to-collision (ETTC). With diminishing relative deceleration, Eq. 11 becomes a limit-value analysis, as in Eq. 3.

The ETTC required for timely braking behind a similarly decelerating obstacle ($D_{\text{obs}} > 0$) is calculated from the maximum relative deceleration:

$$D_{\max,\mathrm{rel}} = D_{\max} - D_{\mathrm{obs}} \tag{12}$$

$$t_{\mathrm{tB}}\left(v_{\mathrm{diff}}, D_{\mathrm{rel}}\right) = \tau_{\mathrm{B}} + \frac{t_{\mathrm{ts}}\left(v_{\mathrm{diff}} + D_{\mathrm{rel}} \cdot \tau_{\mathrm{B}}, D_{\max,\mathrm{rel}}\right)}{2} \tag{13}$$

$$t_{\mathrm{warn}}\left(v_{\mathrm{diff}}, D_{\mathrm{rel}}\right) = \tau_{\mathrm{R}} + t_{\mathrm{tB}}\left(v_{\mathrm{diff}}, D_{\mathrm{rel}}\right) \tag{14}$$

$$d_{\mathrm{B}}\left(v_{\mathrm{diff}}, a_{\mathrm{rel}}\right) = \left(v_{\mathrm{diff}} + D_{\mathrm{rel}}\frac{\tau_{\mathrm{B}}}{2}\right) \cdot \tau_{\mathrm{B}} + \frac{\left(v_{\mathrm{diff}} + D_{\mathrm{rel}} \cdot \tau_{\mathrm{B}}\right)^2}{2D_{\max,\mathrm{rel}}} \tag{15}$$

$$\begin{aligned} d_{\mathrm{warn}}\left(v_{\mathrm{diff}}, a_{\mathrm{rel}}\right) = {} & \left(v_{\mathrm{diff}} + D_{\mathrm{rel}}\frac{\tau_{\mathrm{B}} + \tau_{\mathrm{R}}}{2}\right) \cdot \left(\tau_{\mathrm{B}} + \tau_{\mathrm{R}}\right) \\ & + \frac{\left(v_{\mathrm{diff}} + D_{\mathrm{rel}} \cdot \left(\tau_{\mathrm{B}} + \tau_{\mathrm{R}}\right)\right)^2}{2D_{\max,\mathrm{rel}}} \end{aligned} \tag{16}$$

As regards the stopping distance, in practice the deceleration of the obstacle has the effect of reducing the maximum deceleration of the subject vehicle. For the warning threshold, the relative velocity can still increase considerably by $D_{\mathrm{rel}} \cdot \tau_{\mathrm{R}}$ within the response time.

The relative deceleration is not included in the criterion of required deceleration, but rather, in addition to Eq. 7, only the absolute deceleration of the obstacle D_{obs}:

$$D_{\mathrm{req,D}} = D_{\mathrm{obs}} + \frac{v_{\mathrm{diff}}^2}{2d} \tag{17}$$

Compared to the previous simple case where $v_{\mathrm{obs}} = \mathrm{const.}$, braking must be instigated sooner, i.e., at greater distances, if the obstructing vehicle slows down, because relative braking capability is reduced by this deceleration (cf. Eq. 12).

Evasive Maneuvers

The time required for evasion t_{eva} does not change, even with a relative acceleration, but the required distance does, and this must now be greater by $\left(D_{\mathrm{rel}} \cdot t_{\mathrm{eva}}^2/2\right)$, if the obstructing object is decelerating more than the subject vehicle:

$$d_{\mathrm{eva}}\left(v_{\mathrm{diff}}, D_{\mathrm{rel}}\right) = v_{\mathrm{diff}} \cdot t_{\mathrm{eva}} + D_{\mathrm{rel}}\frac{t_{\mathrm{eva}}^2}{2} \tag{18}$$

Equally, the required lateral acceleration increases relative to the un-accelerated case. However, the variable $t_{\mathrm{tc}}(D_{\mathrm{rel}})$ is needed to calculate the required lateral acceleration for the deviation, this variable describing the time interval available for a lateral movement:

$$a_{\mathrm{y,req,D}} = 2y_{\mathrm{eva}}t_{\mathrm{tc}}^{-2}(D_{\mathrm{rel}}) = \frac{2y_{\mathrm{eva}}D_{\mathrm{rel}}^2}{\left(\sqrt{v_{\mathrm{diff}}^2 + 2D_{\mathrm{rel}}d} - v_{\mathrm{diff}}\right)^2} \tag{19}$$

6.1.3 Calculations for an Obstacle Braking to Standstill

The case considered here falls somewhere between the two scenarios described above. Accordingly, the results also fall between the results for those scenarios.

Braking Maneuvers

If the obstacle (velocity v_{obs}) comes to a standstill before it is reached (at ETTC), and therefore if $(v_{\text{obs}}/D_{\text{obs}}) < t_{\text{tc}}(D_{\text{rel}})$, the TTC increases and the distances required to stop and to evade are shorter than in Eqs. 11, 12, 13, 14, 15, 16, 17, and 18:

$$t_{\text{tc,stop}} = \frac{v_{\text{sub}} - \sqrt{v_{\text{sub}}^2 - 2D_{\text{sub}} \cdot d - v_{\text{obs}}^2 \dfrac{D_{\text{sub}}}{D_{\text{obs}}}}}{D_{\text{sub}}};\ \left(v_{\text{sub}}^2 - 2D_{\text{sub}} \cdot d - v_{\text{obs}}^2 \frac{D_{\text{sub}}}{D_{\text{obs}}}\right) > 0 \tag{20}$$

The required stopping distance is therefore obtained from the difference between the stopping distances of the subject vehicle and the obstructing vehicle plus the distance associated with delay time of the brake:

$$d_{\text{B,stop}} = \frac{v_{\text{sub}}^2}{2D_{\text{max}}} - \frac{v_{\text{obs}}^2}{2D_{\text{obs}}} + v_{\text{sub}} \cdot \tau_{\text{B}} \tag{21}$$

A time criterion t_{tB} for this case does not simplify the representation of Eq. 21 as we have already seen in Eq. 20. For the warning distance, Eq. 21 must be extended to include the response time multiplied by the average velocity difference during the response interval:

$$d_{\text{warn, stop}} = \frac{v_{\text{sub}}^2}{2D_{\text{max}}} - \frac{v_{\text{obs}}^2}{2D_{\text{obs}}} + v_{\text{sub}} \cdot \tau_{\text{B}} + \left(v_{\text{diff}} + D_{\text{rel}} \cdot \frac{\tau_{\text{R}}}{2}\right) \cdot \tau_{\text{R}} \tag{22}$$

The required deceleration $D_{\text{req,stop}}$ is calculated from the sum of the current distance d and the braking distance $v_{\text{obs}}^2/2D_{\text{obs}}$ of the preceding vehicle:

$$D_{\text{req,stop}} = \frac{v_{\text{sub}}^2}{2\left(d + \dfrac{v_{\text{obs}}^2}{2D_{\text{obs}}}\right)} \tag{23}$$

The result always lies between the values for constant velocity and for constant deceleration:

$$D_{\text{req,v}} \leq D_{\text{req,stop}} \leq D_{\text{req,D}} \tag{24}$$

$D_{\text{req,stop}}$ is close to the value $D_{\text{req,v}}$ for an un-accelerated obstacle (Eq. 7), if $v_{\text{obs}}/D_{\text{obs}}$ is low, and close to the value for a constantly decelerating obstacle (Eq. 17), if $v_{\text{obs}}/D_{\text{obs}}$ is high.

Evasive Maneuvers

Even if the obstacle is slowing down to standstill, the time-threshold-evasion remains as in Eq. 8, but, in order to calculate the necessary evasion distance, this time it must be multiplied by the average velocity of the subject vehicle ($v_{\text{diff}} - D_{\text{sub}} \cdot t_{\text{eva}}/2$). The current distance d and the deceleration distance $v_{\text{obs}}^2/2D_{\text{obs}}$ of the obstructing object are already available:

$$d_{\text{eva}}(v_{\text{obs}}, D_{\text{obs}}) = \left(v_{\text{diff}} - D_{\text{sub}} \cdot \frac{t_{\text{eva}}}{2}\right) \cdot \tau_{\text{eva}} - \frac{v_{\text{obs}}^2}{2D_{\text{obs}}} \tag{25}$$

Like the required deceleration, the required lateral acceleration

$$a_{y,\text{req}} = 2y_{\text{eva}} t_{\text{tc}}^{-2}(v_{\text{obs}}, D_{\text{obs}}) = \frac{2y_{\text{eva}} D_{\text{sub}}^2}{\left(v_{\text{sub}} - \sqrt{v_{\text{sub}}^2 - 2D_{\text{sub}} \cdot d - v_{\text{obs}}^2 \cdot \frac{D_{\text{sub}}}{D_{\text{obs}}}}\right)^2} \tag{26}$$

lies between the values for an un-accelerated obstacle (Eq. 10) and those for a constantly decelerating obstacle (Eq. 19). The root then only remains real if continuation of the movement would actually result in a collision. Otherwise, the subject vehicle stops at a finite distance away from the obstacle and thus discontinues the evasive action so that it is impossible to calculate a meaningful solution for required lateral acceleration.

Table 1 can be used to find the corresponding results for all the specified criteria and for the three different cases discussed here.

Calculations for Simultaneous Braking and Steering

If a tire is subject to longitudinal forces, it can no longer provide maximum transverse forces. This relationship is described – in simplified form – by Kamm's circle (see Fig. 4), which, in accordance with the equation

$$a_{y,\max}(D_x) = a_{y,\max}\sqrt{1 - \frac{D^2}{D_{\max}^2}}; 0 \le D \le D_{\max} \tag{27}$$

can also be used for an ellipse with different maximum friction coefficients for the longitudinal and transverse directions. In a combined braking and steering maneuver, the evasion time t_{eva} is extended, because now the reduced maximum lateral acceleration $a_{y,\max}(D_x)$ is inserted into Eq. 8.

On the other hand, the time interval to reach the obstacle is extended due to the positive relative acceleration achieved by braking, in accordance with Eq. 11.

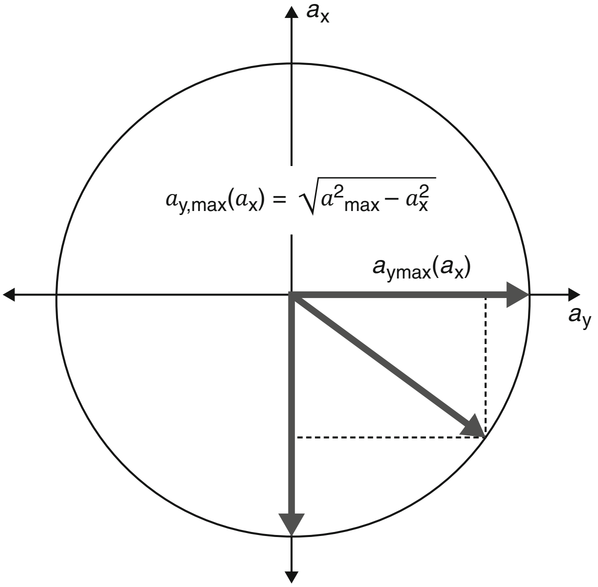
This effect dominates at low ratios of $D/D_{\max}$. In an adapted representation (Schmidt et al. 2005), the effect of combined braking and steering interventions can be illustrated via the movement of an ellipse that increases in size over time, its midpoint moving with the initial velocity (see Fig. 5).

This representation shows that braking up to a degree that is a function of the initial conditions results in greater localized evasion capability. Related to the

Table 1 Triggering thresholds with references to the calculations in Sect. 6

Warning and intervention thresholds			
	Vehicle dynamics scenarios		
	Un-accelerated obstacle	Obstacle relatively decelerating at D_{rel}	Obstacle braking to standstill
Time-threshold-evasion (t_{eva})		(Eq. 8) 0.55 … 0.7 s	
Evasion distance d_{eva}	Eq. 9 $v_{\mathrm{diff}} \cdot t_{\mathrm{eva}}$	Eq. 18 $v_{\mathrm{diff}} \cdot t_{\mathrm{eva}} + D_{\mathrm{rel}} \frac{t_{\mathrm{eva}}^2}{2}$	Eq. 25
Required lateral acceleration $a_{y,\mathrm{req}}$	Eq. 10 $\frac{2y_{\mathrm{eva}} v_{\mathrm{diff}}^2}{d^2}$	Eq. 19	Eq. 26
Time-threshold-brake (t_{tB})	Eqs. 4 and 5 $\tau_{\mathrm{B}} + \frac{v_{\mathrm{diff}}}{2D_{\mathrm{max}}}$	Eq. 13	–
Braking distance d_{B}	Eq. 1 $v_{\mathrm{diff}} \cdot \tau_{\mathrm{B}} + \frac{v_{\mathrm{diff}}^2}{2D_{\mathrm{max}}}$	Eq. 15	Eq. 21
Warning time t_{warn}	Eq. 6 $\tau_{\mathrm{R}} + t_{\mathrm{tB}}$	Eq. 14	–
Warning distance d_{warn}	Eq. 2	Eq. 16	Eq. 22
Required deceleration D_{req}	Eq. 7 $\frac{v_{\mathrm{diff}}^2}{2d}$	Eq. 17 $D_{\mathrm{obs}} + \frac{v_{\mathrm{diff}}^2}{2d}$	Eq. 23
	Driver behavior		
	Driver's limit	**Comfort limit**	
Time-threshold-evasion	1 s	1.6 s	

Fig. 4 Distribution of longitudinal and lateral forces on the tire (Kamm's circle)



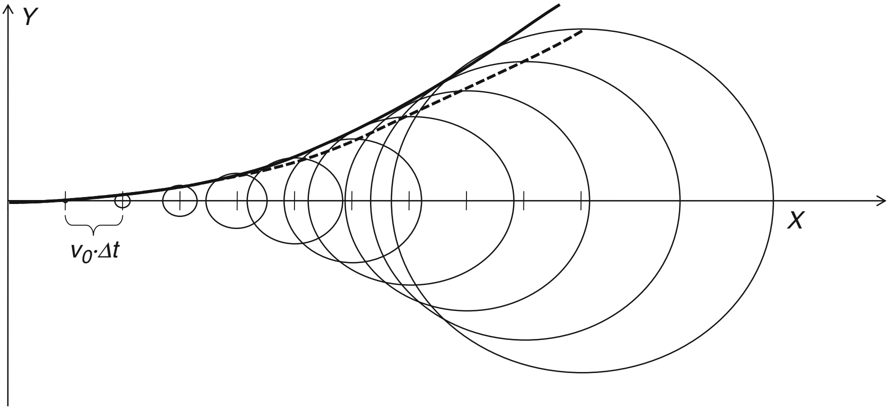


Fig. 5 Possible safe driving zones with combined longitudinal and lateral acceleration. The ellipses move with the initial velocity v_0 and increase in size with the square of time. The solid line corresponds to the ideal lateral distance at each point, and the dotted line to the trajectory with lateral acceleration alone (Representation as in Schmidt et al. (2005))

distance at the start of evasion, the gain due to an ideal braking/steering maneuver, as opposed to a steering maneuver alone, is only usually a few centimeters (Schmidt et al. 2005) and therefore lies within the range of miscellaneous inaccuracies. The equations regarding evasion criteria can therefore continue to be used, especially the basic equation for the time criterion, Eq. 8. However, the distance $\Delta d = D_{\text{obs}} t_{\text{eva}}^2/2$ must be added for a decelerating obstacle ($D_{\text{obs}} > 0$), and the distance $\Delta d = v_{\text{obs}}^2/2D_{\text{obs}}$ for an obstacle braking to standstill. Related to the TTC with un-accelerated initial conditions, the time requirement therefore increases by $\Delta d/v_{\text{sub}}$, where the smaller of the abovementioned Δd values is to be used. With the previously mentioned representative value $t_{\text{eva,phys}} = 0.6$ s for the maximum friction coefficient and average evasive deflection, an additional distance of 1.8 m must be considered, which consequently results in an increase in the time limit by a maximum of 0.3 s, but more typically by 0.15 s.

Even if the description of the safe driving zone is very easy to construct, as in Fig. 5, caution is advised if time aspects are to be considered as well as spatial aspects (▶ Chap. 50, "Intersection Assistance"). In the case of lateral acceleration alone, the trajectory does not run along the parabola shown in Fig. 5, but rather, as we know, in a circle. The main difference between these two courses does not initially relate to the transverse direction, but the longitudinal direction, as the equation below shows.

The projection of the circular motion with constant circular velocity on the original longitudinal axis deviates from the course shown in Fig. 5 by

$$\Delta x = R(\vartheta - \sin\vartheta); \; \vartheta = v_0 t/R; R = v_0^2/a_y \tag{28}$$

where by the lateral path

$$y = R(1 - \cos\vartheta);\ \vartheta = \arccos(1 - y/R) \tag{29}$$

is traveled. The angle ϑ can therefore be replaced by a term that is dependent upon y and R:

$$\Delta x/R = (\arccos(1 - y/R) - \sin[\arccos(1 - y/R)]) \approx (2y/R)^{3/2}/6 \tag{30}$$

The approximation is obtained by series expansion of the trigonometric functions, cut off after the cubic term. The smaller the radius R (implicitly at lower velocity) and the greater the offset, the more weight this difference carries. The time difference depends firstly upon the square of the ratio a_y/v_0 and secondly upon the third power of ($2y/a_y$), the time to lateral displacement y (cf. Eq. 8, first term):

$$\Delta x/v_0 \approx \left(a_y/v_0\right)^2 \left(2y/a_y\right)^{3/2}/6 \tag{31}$$

This difference is only relevant if v_0 is small and y is of the order of a lane width (e.g., $\Delta x/v_0 \approx 100$ ms at $v_0 = 10\,\frac{\mathrm{m}}{\mathrm{s}}$, $a_y = 10\,\frac{\mathrm{m}}{\mathrm{s}^2}$, $y = 3.5$ m).

Drivers' Evasive Behavior

In the previous section we only looked at the physical aspects of driving. However, only exceptionally skilled drivers drive to the physical limits of the vehicle. In a study conducted by Honda (Kodaka et al. 2003) (Fig. 6), evasive maneuvers were rated in three danger categories. The lower limit of the medium rating ("feel somewhat dangerous") can clearly be identified as being in the range of TTC $>$ 1.6 s. The lower limit of the evasive actions rated as "feel no danger" equates to a TTC of 2.5 s. We can conclude from these two values that, although an evasive maneuver within one-second TTC is physically possible, it will not be deliberately undertaken, because of the rating, even if the drivers are risk-takers. This threshold is referred to below as the driver's limit. However, drivers perceive that the safe range is left even earlier, at a TTC of around 1.6 s, so that even this can no longer be regarded as a normal evasive situation, and a forward collision countermeasure seems to be justified. This threshold is referred to below as the comfort limit.

Together with the physical limit, there are now three representative thresholds for forward collision countermeasures:

- At t_{eva} (approx. 0.6 s) evasion is no longer physically possible.
- At t_{driver} (approx. 1 s) a driver is no longer capable of evasive action in practice.
- At t_{comfort} (approx. 1.6 s) evasive action is regarded as dangerous.

However, even at the earlier threshold t_{comfort}, warnings to brake are still not soon enough. If we assume a response time of 1 s (plus delay time of the brake of 0.1 s), this can only compensate a velocity difference of $2D_{\max} \cdot 0.6\ \mathrm{s} \approx 12$ m/s. A particularly effective warning with a response time of 0.5 s would produce values of

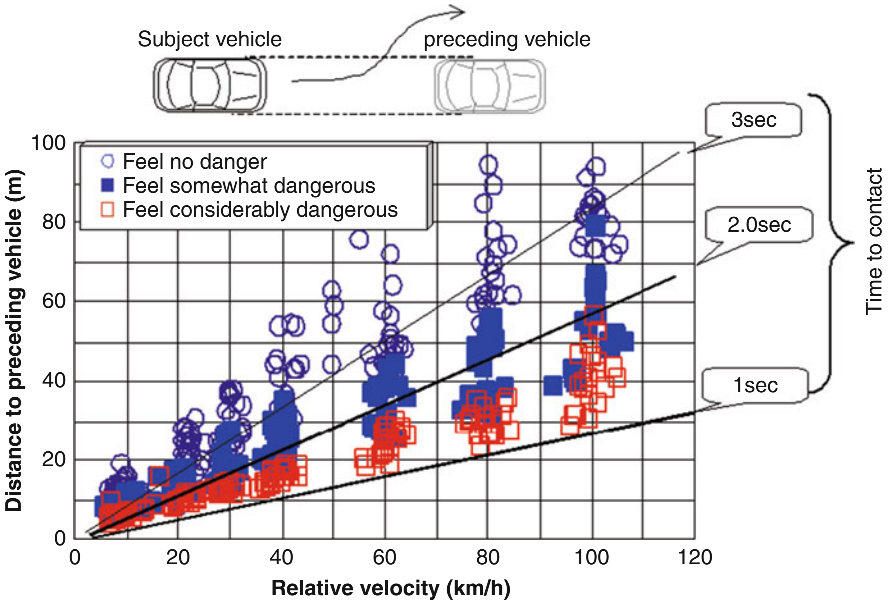


Fig. 6 Subjective assessment of evasive maneuvers (Source: Honda (Kodaka et al. 2003))

22 m/s. However these examples assume full emergency braking under good road surface conditions. A warning that is given approximately one second earlier gives the driver time for a moderate response and, if emergency braking is applied, can even offset velocity differences of 30–40 m/s, which would cover the majority of cases in practice. In line with this thinking, an additional threshold is introduced:

- The warning threshold t_{warn} with values of between 2.5 and 3 s

For these implementation thresholds that describe normal driving applications, the values could be modified by additional criteria. One strategy is observing drivers to see whether they are paying attention. In a commercially available example (Lexus LS, APCS or advanced pre-collision system), a driver monitor mounted on the steering console checks whether the driver is looking to one side. If it identifies a longer visual distraction time, an earlier warning is given and there is an earlier intervention. Other criteria that can alter the threshold values are the friction coefficient μ and the visual range. A lower friction coefficient increases the derived physical limits for driving, notably the time-threshold-evasion by $1/\sqrt{\mu}$ and the braking time criteria by $1/\mu$. If acceleration thresholds are used for triggering, the friction coefficient, if it is determined in some way, can be used directly for the comparison between required deceleration and lateral acceleration, since the maximum possible deceleration/lateral acceleration can be estimated via $\mu \cdot g$. However, there are currently no known methods for determining the friction coefficient in advance, to allow corresponding modification of the thresholds.

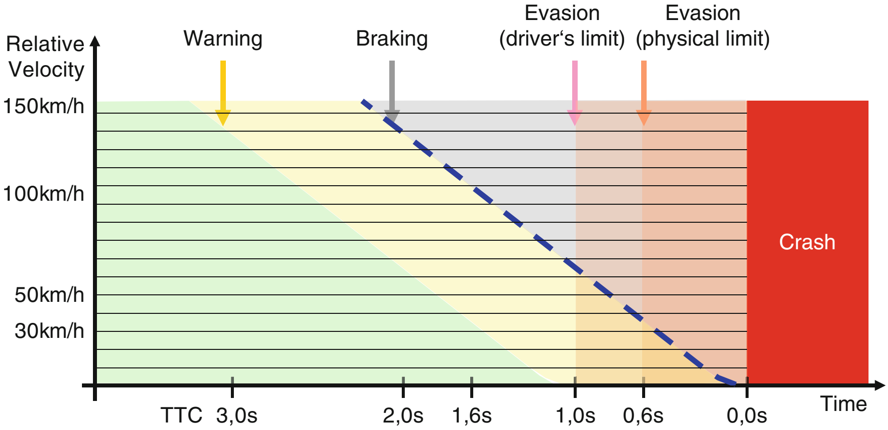


Fig. 7 Illustration of warning and intervention times for the case of an un-accelerated moving obstacle and subject vehicle (TTC = distance/relative velocity)

As far as the visual range is concerned, there are optical methods for measuring the backscatter with LIDAR or LIDAR-like sensors or with cameras. Another possible indicator of a short visual range and also a diminished friction coefficient is rapid operation of the windshield wipers. The fog light being switched on can also be used as a reason for modification, especially in the case of warnings at a greater distance from the obstacle.

Summary

The simplified diagram in Fig. 7 shows all the interventions derived in the previous sections. Once again, we find the value representing the physical limit t_{eva} (approx. 0.6 s), the driver's limit is represented by the value t_{driver} (approx. 1 s), while, because the braking distance increases as the square of the velocity, the warning threshold is a linear function of the relative velocity. Full braking applied before the broken line can prevent the crash. With the assumptions made here, at differential velocities in excess of 10 m/s (36 km/h), collisions can be avoided by an evasive maneuver at a later stage than they can be avoided by braking.

The illustration chosen in Fig. 7 can be used for a simple and correct analysis of the case where constant velocities are assumed for both the obstacle and the subject vehicle. On the other hand, if the subject vehicle decelerates (relative to the object), the TTC axis is not identical to the time axis. Thus, full braking starting along the dotted line lasts approximately twice as long as the given TTC value, because, due to the deceleration, the (time-related) average relative velocity is only approximately half the initial velocity.

The illustration chosen here represents a favorable case. With lower friction values or decelerating obstructing objects, the thresholds move toward higher TTC values; the evasion values largely move toward higher TTC values, irrespective of the relative velocity; while the braking limit, and therefore also the warning limit, is steeper, that is to say, it is displaced toward higher TTC values proportionate to the relative velocity. Only if the vehicle is traveling in a curved trajectory could evasion to the outside of the curve allow a successful evasive maneuver to be initiated even later (where success relates to the obstacle being avoided and does not take account of whether the escape corridor is safe).

6.2 Forward Collision Countermeasures

In addition to reaction assistance measures, advisory and automatically acting countermeasures can also be taken. Whereas the informative measures aim to prevent the accident, the interventional ones aim, additionally, to mitigate its severity. The measures discussed below can either be implemented as a whole package or in part, whereby a measure that is triggered at a later stage should essentially indicate a stronger braking intervention.

Standards exist for two families of systems. Firstly, ISO 15623 "Road vehicles – Forward vehicle collision warning systems – Performance requirements and tests procedures" (ISO 15623 2013), which was introduced in 2002 and revised in 2013, describes the minimum requirements for forward collision warning systems. Secondly, ISO 22839 "Intelligent transport system – Forward vehicle collision mitigation systems – Operation, performance, and verification requirements" (ISO 22839 2013) was introduced in 2013 and describes the requirements of speed reduction braking (SRB) systems and/or mitigation braking (MB) systems.

6.2.1 Collision Warning

Since the warning elements are described in detail in ▶ Chap. 36, "Driver Warning Elements," here we merely provide a list of the collision warning elements currently in use. Particularly common are auditory warnings, most of which have a short warning tone, supplemented by an additional blinking optical display featuring a warning symbol. In vehicles fitted with a reversible seat belt tensioner, this can also be used to provide a warning. An active feedback pedal can also initiate a haptic warning, in that the movable spring mount is suddenly jerked against the driver's foot. However, if this is to be successful in attracting attention, the driver's foot must be on the pedal in the first place, and this is not always the case in every critical situation.

The application range of the warning is between the two earliest values t_{warn} and t_{comfort}. Trade-offs between effectiveness and forgiveness can mean that more forgivable warnings can be implemented earlier on, whereas less forgivable, but very effective, warnings can be implemented later, that is, at shorter TTC (▶ Chap. 36, "Driver Warning Elements").

Standard ISO 15623 "Road vehicles – Forward vehicle collision warning systems – Performance requirements and tests procedures" specifies implementation times derived from the required deceleration and a reaction time, in line with the considerations in Sect. 6.1. $D_{\mathrm{req}} = 6.67\ \mathrm{m/s2}$ and $\tau_{\mathrm{R}} = 0.8$ s are specified as the maximum values.

6.2.2 Conditioning for an Emergency Maneuver

At the same time as the warning or slightly thereafter, measures can be taken to assist the emergency maneuver performed by the driver. Brake pre-fill is already initiated virtually as standard. This means applying a small clamping force to the brake (in the case of hydraulic brakes, approximately 1–5 bar brake pressure). The deceleration this brings about is not noticeable but makes the brake more responsive. Another measure that is included in forward collision avoidance systems is reducing the triggering threshold of the brake assist system when there is a threat of forward collision. If it is possible to change the suspension setting of the vehicle, this can enhance emergency braking and also emergency evasion. In this way, handling can be improved with a short-term adjustment, at the expense of comfort. In vehicles with superimposed and/or electrically assisted steering, emergency evasive maneuvers can also be preconditioned by altered characteristics. Examples of extensive preconditioning are PRE-SAFE from Mercedes-Benz (albeit without evasion conditioning) and the Advanced Pre-Collision System (APCS) from Lexus.

6.2.3 Speed Reduction Braking

An automatic, dynamically effective collision countermeasure can be implemented at the comfort threshold between 1.5 and 2.0 s. There are two potential options:

1. Jerk (warning braking)
   The main effect of warning braking is that it acts as a haptic alert, clearly prompting the driver to brake. An example of warning braking with a typical deceleration amplitude 4 m/s^2, buildup and reduction periods of 0.2 s, and a duration of typically 0.3 s produces a speed reduction of around 2 m/s and therefore, at a velocity difference of 20 m/s, a 20 % reduction in kinetic energy or braking distance, without critically changing the driving situation, if we disregard an overtaking maneuver that has already been embarked upon.
2. Speed reduction braking (SRB)
   Partial braking of 30–40 % of the maximum deceleration combines a strong warning effect with a considerable reduction in kinetic energy. Thus, partial braking commencing at a TTC of 1.5 s and a braking action of 4 m/s^2 can reduce the differential velocity relative to a non-decelerating obstacle by around 12 m/s. However, with such an early initiation, it must be ensured that it can be overridden by the driver. Especially if an evasive maneuver is detected, the braking intervention would have to be canceled. The same applies if accelerator operation indicates another intent on the part of the driver. However, it is first of all necessary to exclude the possibility that the accelerator movement is actually due to warning braking, as is documented in the case of full braking (Bender 2008). Another measure for reducing the potential damage caused by such a

mistaken intervention is to limit the duration of the intervention to double the intervention TTC: because, depending upon how this limit is set, the duration will either be sufficient to avoid the collision or it must have taken place within this time period. A reason for reducing it further is that driver behavior observed in experiments (Hoffmann 2008) in controlled driving tests showed that automatic braking speeded up the driver's braking response, so it can be assumed that the driver will respond within a maximum of 1.3 s if a dangerous situation threatens.

6.2.4 Mitigation Braking

A strong application of the brakes (defined here as at least 50 % of maximum deceleration capacity) can take place when there is no possibility of evasion. However, some of the parameters that need to be used for making this decision, e.g., assumed evasive offset, cannot be determined with sufficient accuracy. So far it has been practiced in passenger car applications to implement mitigation braking with a deceleration of approximately 6 m/s^2 from 1 s TTC. With an ideally fast-acting brake, this would be sufficient to slow the vehicle down by between 6 and 12 m/s. Again, it is known that most drivers brake with double the time, therefore 2 s, so that, even if the system were falsely triggered, only 12 m/s would be the maximum speed reduction.

In principle, even stronger braking would be more beneficial. However, tests with drivers (Hoffmann 2008) show that this benefit can only be achieved if the buildup is very rapid. If the deceleration is built up slowly (approx. 10. . .20 m/s^3), even braking of 6 m/s^2 would be just as effective as automatic full braking, because of the subsequent brake operation by the driver. Moreover, strong collision mitigation braking can be counterproductive if the brake jerk moves the occupants' bodies and heads out of the most favorable position for deployment of the retaining systems (airbag, seatbelt tensioning). This side effect can be considerably reduced if reversible seatbelt tensioners restrain the occupants before or at the same time as mitigation braking is implemented.

6.3 Potential Benefits of Forward Collision Countermeasures

The benefit of forward collision countermeasures is that they avoid forward collisions and mitigate the damage caused by the crash if it is no longer possible to prevent it. As the calculations so far have shown, whether a collision can be avoided and, if not, how great the damage mitigation will be depends to a large extent upon the initial situation. The effective reduction in velocity can be determined to serve as a parameter of benefit that is largely independent of the initial conditions. However – even with an idealized system – this effective reduction in velocity is still a function of the initial relative velocity, as the following example should illustrate, based on the simplest case – that of a stationary obstacle.

Idealized emergency braking is assumed, this being initiated at a time t_{tB} and decelerating immediately with D_0. At an initial velocity difference of $v_0 = 2t_{\mathrm{tB}} \cdot D_0$,

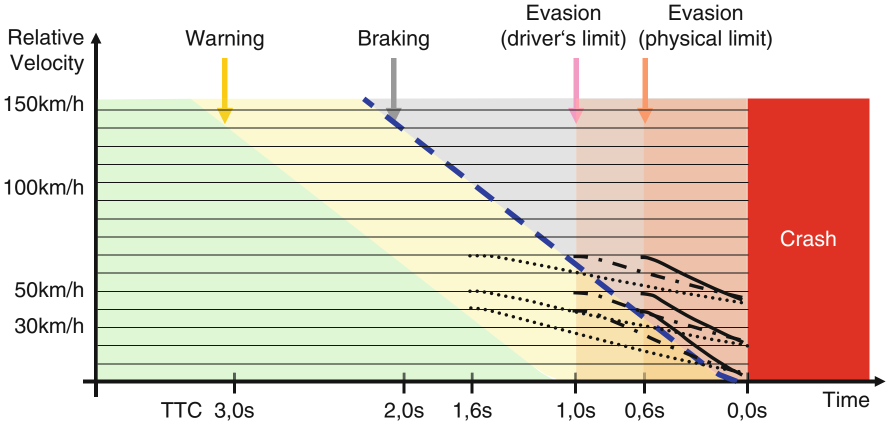


Fig. 8 Three equally effective intervention strategies ($\Delta v_{\text{CM}} = 5$ m/s for three different initial relative velocities (40, 50, 70 km/h)

a collision can still just be avoided, and a velocity of $\Delta v = v_0 = 2t_{\text{tB}} \cdot D_0$ can be reduced. On the other hand, if the initial velocity is very high ($v_1 \gg 2t_{\text{tB}} \cdot D_0$), only half as much is reduced. To obtain a transferable parameter, we therefore only consider the difference from the initial velocity, which lies within the period corresponding to the TTC prevailing at the time of initiation, that is, $\Delta v_{\text{CM}} = v_{\text{sub}}(t_{\text{i}}) - v_{\text{sub}}(t_{\text{i}} + t_{\text{tB}})$.

This definition of effectiveness not only allows simple assessment of idealized collision countermeasures but also permits an objective evaluation of measures that is not dependent upon the technical solution. It is even possible to make a comparative assessment of automatically acting systems and driver assistance systems (e.g., via warnings or jerk), as documented in ▶ Chap. 13, "Evaluation Concept EVITA." Figure 8 shows the course of three strategies with a general effectiveness of $\Delta v_{\text{CM}} = 5$ m/s. All three could effectively still prevent a collision from an initial velocity $v_0 \leq 2\Delta v_{\text{CM}} = 10$ m/s $= 36$ km/h. At 40 km/h they reduce the damage to merely "parking bumps" and reduce the kinetic energy proportionate to the initial velocity by $m \cdot v_{\text{sub}}(t_i) \cdot \Delta v_{\text{CM}}$, the velocity reduction diminishing at higher velocities to $\Delta v_{\text{CM}} = 5$ m/s $= 18$ km/h.

An effectiveness of $\Delta v_{\text{CM}} = 5$ m/s is representative of individual interventions, such as are found, for example, in Honda CMBS with a braking intervention of 6 m/s^2, triggered at $t_{\text{tc}} = 1$ s and with approx. $t_{\text{B}} = 0.2$ s delay time or alternatively with gentle but early partial braking at 1.6 s with 3.3 m/s^2 and $t_{\text{B}} = 0.1$s. Full emergency braking, which can only be activated once the possibility of evasion has been ruled out, that is, at $t_{\text{tc}} = 0.6$ s, can achieve this type of effectiveness with a maximum deceleration (10 m/s^2) and very fast braking buildup dynamics ($t_{\text{B}} = 0.1$ s).

However, even greater potential can be achieved using a multistage process. The initiation of gentle braking (speed reduction braking) after reaching the comfort limit

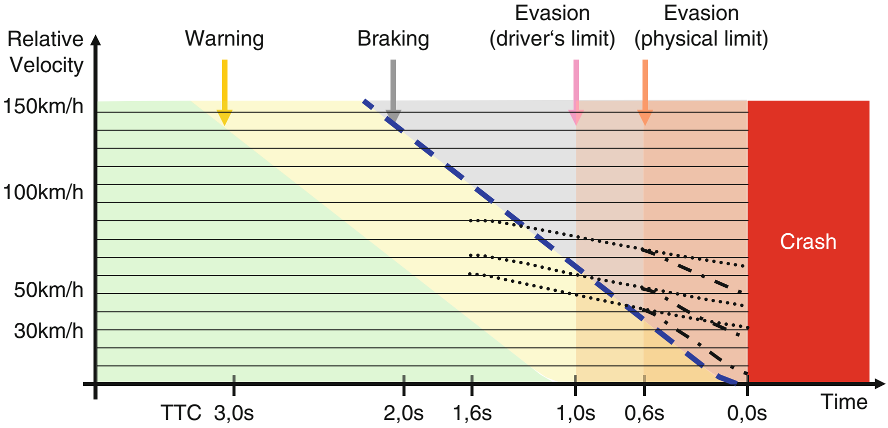


Fig. 9 Two-stage intervention strategy ($t_{\mathrm{tc},1} = 1.6$ s, $D_1 = 4\mathrm{m/s^2}$, $t_{\mathrm{tc},2} = 0.6$ s, $D_2 = 10\ \mathrm{m/s^2}$, delay time 0.1 s in each case) with an effectiveness of $\Delta v_{\mathrm{CM}} = 9$ m/s ($= 1\ \mathrm{s} \cdot 4\ \mathrm{m/s^2} + 0.5\ \mathrm{s} \cdot 10\ \mathrm{m/s^2}$) for three different initial relative velocities (60, 70, 90 km/h)

for evasion ($t_{\mathrm{tc}} = 1.6$ s) and full braking following one second on TTC scale later is used as an example. This gives an effectiveness of $\Delta v_{\mathrm{CM}} = 9\ \mathrm{m/s}$. However, in contrast to the single-stage process, since deceleration is not constant and is initially smaller, less than double the velocity reduction is achieved (see Fig. 9). Nevertheless, a speed reduction of around 60 km/h is possible with such an arrangement. Even the crash test conducted at the highest velocity of 64 km/h was downgraded to a “parking bump,” but it should not be forgotten that the analyses correspond to a best case and this protective effect could in no way be expected in every crash situation. Nevertheless, the example shows what can be achieved.

Instead of extending functionality, in 2008 Volvo started to have a strong market impact with its “bottom-up” approach with the City Safety concept. This concept uses a very reasonably priced LIDAR sensor with only a small range to prevent rear-end collisions at low velocities of up to 30 km/h. The concept was also adopted by other vehicle manufacturers or integrated as a partial function into other automatic emergency braking systems.

6.4 Surroundings Sensing Requirements

Requirements of surroundings sensing are determined by the vehicle responses that are triggered. For example, the requirements to be met by hazard detection systems that trigger warnings are less stringent in terms of the false alarm rate than those that trigger strong braking. However, the degree of difficulty also varies accordingly, i.e., warning detections take place at greater distances. With warning thresholds t_{warn} of 2.5–3.0 s, we obtain distances of $v_{\mathrm{diff}} \cdot t_{\mathrm{warn}} \approx 50\ \mathrm{m}$ for un-accelerated obstacles at velocity differences of 60–70 km/h. The derived $d_{\max}$, given in standard ISO 15623 (ISO 15623 2013) for forward vehicle collision warning systems (FVCWS) as $d_{\max} = v_{\text{rel},\max} \cdot \tau_{\max} + v^2_{\text{rel},\max}/2D_{\max}$ with $v_{\text{rel,max}}$ as the upper relative velocity limit of the system, response time $\tau_{\max} \geq 0.8$ s, and $D_{\max} = 6.67$ m/s2, results in minimum warning distances of $d_{\max} \geq$ 46 m if it is designed with $v_{\text{rel,max}} =$ 20 m/s. This value increases for decelerating preceding vehicles and, of course, for the margin for target plausibility checking.

In contrast, the requirements for the precision of distance measurement with MAX(±2 m, ± 1 % ·*d*) formulated in standard ISO 15623 can easily be achieved with point target test reflectors. The azimuthal (lateral) sensing area is dependent upon the curve capability class. The criterion to be fulfilled is the detection of a point target, located at a distance defined as d_2 from the vehicle along the extension of the lateral line of the vehicle. For a 1.80 m wide vehicle, a single sensor mounted in the center must cover an azimuth angle range of ±9° (Curve Capability Class I, $d_2 = 10$ m) to ±18° (Class III, $d_2 = 5$ m). In contrast, the elevation angle range (vertical visual range) must be large enough to detect a point target located centrally at d_2 at a height of 0.2 and 1.1 m. If it is suitably aligned, the elevation range must consequently be exactly half that required for the azimuth. A new requirement imposed by the standard is that vehicles should be able to drive under bridges with a clearance of at least 4.5 m without an alert. Many of the sensor requirements arising from ISO 15623 are also to be found in ISO 22839. However, no explicit values are specified for the intervention thresholds. Only the velocity range of the subject and object vehicles in which a braking response must take place (at a not directly defined distance) is given. However, this response is defined via the intensity and effectiveness of the intervention (velocity reduction).

In practice, ranges of 60–80 m are required for the functions discussed here (Randler and Schneider 2006). This allows a TTC of more than 2 s to be achieved relative to stationary obstacles, even at 100 km/h, thereby providing a value that would still allow a collision to be successfully prevented by means of safe evasion or, with a fast response, by braking. The strong braking required to mitigate a collision is only triggered at a TTC of around 1 s, which means that a response distance of less than 15 m is required for collisions with differential velocities of up to 50 km/h. Assuming a signal plausibility checking time of 0.3 s, the required distance is increased to approximately 20 m.

The greatest challenge is not so much detecting relevant objects, but rather selecting those that actually constitute a threat. Sign gantries, manhole covers, and conflated objects should not cause triggering. Such spurious objects can be successfully filtered out via longer observation (Jordan et al. 2006) of the reflection intensity and constancy of the angle values as a function of distance. Only objects with such plausible behavior are used in evaluation for the purposes of collision countermeasures.

In order to derive the requirement for robustness, we can analyze the number of deployments. For this purpose, the figures for Germany for accidents involving personal injury (more than 300,000) are related to the mileage of the car ($6{\cdot}10^{11}$ km). The accident rate therefore equates to approximately one accident involving personal injury per two million kilometers. Only some of these accidents are forward collisions so that we can assume a deployment rate of one time per five million kilometers, which can still be regarded as optimistic. Since not every instance of erroneous triggering necessarily results in a serious accident, a rate that is similar to that of strong interventions should be considered acceptable. However, that also means that there would be an average mileage of five million kilometers between two erroneous triggerings or, viewed from another perspective, only one within a total of 25 vehicle lifetimes.

The determined TTC and the required deceleration are used to represent the precision of the object data. An inaccuracy of 10 % TTC and 0.5 m/s^2 D_{req} should not be exceeded, and the filter delay times should not give rise to a total time delay of more than 200 ms.

Restriction to low speeds, as discussed above, and therefore restriction to the inner city area, reduces this requirement considerably. If the sensors are suitably located, e.g., behind the windshield, and overall only leave small distance ranges of approximately 10 m, it is possible to eliminate ground reflection or other constellations that generate false measurements. Therefore, an object detected within this range is essentially assessed as a relevant obstacle, and consequently speed reduction braking can be initiated if there is a threat of collision. However, since a function that is only suitable for low speeds is no longer what the market demands, sensor functions need to be extended or additional sensors are required.

7 Outlook

Although people have been working on forward collision avoidance systems for more than 50 years, only the recent developments in environment sensors, which were stimulated by comfort functions such as ACC or FSRA, have led to applications suitable for series application. This opens the way, technologically speaking, to allow environment sensors to be used directly for collision avoidance. Partly for reasons of strategic market placement and partly driven by pressure from competitors, the available range has expanded rapidly, so that buyers of new cars can choose from a large number of models that are equipped with forward collision avoidance systems. Inclusion in the EURO-NCAP catalog reflects this development and, at the same time, drives it forward, so that we can expect it to be fitted as standard in nearly all classes of vehicles.

The function wish list includes enhancements to the protection of unprotected road users, pedestrians, and cyclists and extension to cover more complex scenarios such as collision avoidance and mitigation in side-on collision scenarios at intersections (▶ Chap. 50, “Intersection Assistance”).

References

Bender E (2008) Handlungen und Subjektivurteile von Kraftfahrzeugführern bei automatischen Brems- und Lenkeingriffen eines Unterstützungssystems zur Kollisionsvermeidung (Driver actions and subjective assessment of automatic braking and steering interventions by a collision-avoidance assistance system). Dissertation TU Darmstadt, Ergonomia-Verlag, Stuttgart

Braun H, Ihme J (1983) Definition kritischer Situationen im Kraftfahrzeugverkehr – Eine Pilotstudie (Definition of critical situations in road traffic – a pilot study). Automobil-Industrie 3/1983, Vogel-Verlag, Würzburg, p. 367–375

Busch S (2005) Entwicklung einer Bewertungsmethodik zur Prognose des Sicherheitsgewinns ausgewählter Fahrerassistenzsysteme (Development of an evaluation methodology to forecast the safety benefits of selected driver assistance systems). VDI Fortschritt-Berichte Series, 12. Dissertation TU Dresden, VDI-Verlag, Düsseldorf

Eckstein L (2001) Entwicklung und Überprüfung eines Bedienkonzepts und von Algorithmen zum Fahren eines Kraftfahrzeuges mit aktiven Sidesticks (Development and verification of an operating concept and of algorithms for driving a motor vehicle with active sidesticks). Fortschritt-Berichte VDI-Series, 12, vol 471. Dissertation TU Munich, VDI-Verlag, Düsseldorf

Färber B, Maurer M (2005) Nutzer- und Nutzen-Parameter von Collision Warning und Collision Mitigation Systemen im Verkehr (User and use parameters of collision warning and collision mitigation systems in road traffic). In: Tagungsband 3. Workshop Fahrerassistenz FAS2005, Walting

Hoffmann J (2008) Das Darmstädter Verfahren (EVITA) zum Testen und Bewerten von Frontalkollisionsgegenmaßnahmen (The Darmstadt Method (EVITA) to test and evaluate forward collision countermeasures). Dissertation TU Darmstadt Fortschritt-Berichte VDI Reihe 12 Nr. 693, Düsseldorf

ISO 15623 Standard (2013) Transport information and control systems – forward vehicle collision warning systems – performance requirements and test procedures

ISO 22839 Standard (2013) Intelligent transport system – forward vehicle collision mitigation systems – operation, performance, and verification requirements

Jordan R, Ahlrichs U, Leineweber T, Lucas B, Knoll P (2006) Hindernisklassifikation von stationären Objekten auf Basis eines nichtwinkeltrennfähigen Long-Range-Radar Sensors (Obstacle classification for stationary objects on the basis of a long-range radar sensor without angle separation capability), 4. Workshop Fahrerassistenzsysteme, 4–6 Oct 2006, Löwenstein/Hößlinsülz, p 153

Kiesewetter W, Klinkner W, Reichelt W, Steiner M (1997) Der neue Brake-Assist von Mercedes-Benz – aktive Fahrerunterstützung in Notsituationen The new brake-assist from Mercedes-Benz – active driver support in emergency situations. ATZ Automob Z 99(6):330–339

Kodaka K, Otabe M, Urai Y, Koike H (2003) Rear-end collision velocity reduction system, SAE paper 2003–01-0503, Warrendale

Kopischke S (2000) Entwicklung einer Notbremsfunktion mit Rapid Prototyping Methoden (Development of an emergency braking function with rapid prototyping methods). Report, Institut für Elektrische Messtechnik und Grundlagen der Elektrotechnik der Technischen Universität Braunschweig, vol 10. Dissertation TU-Braunschweig, Aachen/Mainz

LeBlanc DJ, Kiefer RJ, Deering RK, Shulman MA, Palmer MD, Salinger J (2001) Forward collision warning: preliminary requirements for crash alert timing, SAE 2001– 01-0462, Warrendale

Morrison RW, Swope JG, Halcomb CG (1986) Movement times and brake pedal placement. J Hum Factors Ergon Soc 28(2):241–246

Randler M, Schneider K (2006) Realisierung von aktiven Komfort- und Sicherheitsfunktionen mit Lidarsensorik (Implementing active comfort and safety functions using lidar sensorics). In: Fachforum Sensorik für Fahrerassistenzsysteme. FH Heilbronn, Heilbronn

Schmidt C, Oechsle F, Branz W (2005) Untersuchungen zu letztmöglichen Ausweichmanövern für stehende und bewegte Hindernisse (Investigations into the last possible maneuver to avoid stationary and moving obstacles), 3. In: FAS-Workshop, Walting

Seeck A, Gasser TM (2006) Klassifizierung und Würdigung der rechtlichen Rahmenbedingungen im Zusammenhang mit der Einführung moderner FAS. Tagung "Aktive Sicherheit durch Fahrerassistenzsysteme" (Classification and evaluation of legal framework conditions relating to the introduction of modern driver assistance systems). In: Conference on active safety through driving assistance, Munich

Weinberger M (2001) Der Einfluss von Adaptive Cruise Control Systemen auf das Fahrerverhalten. Berichte aus der Ergonomie (The influence of adaptive cruise control systems on driver behavior. Reports from ergonomics). Dissertation Technische Universität Munich, Shaker/Aachen

Weiße J (2003) Beitrag zur Entwicklung eines optimierten Bremsassistenten (On the development of an optimated brake assistant). Dissertation TU Darmstadt, Ergonomia Verlag, Stuttgart

Wiacek CJ, Najm WG(1999) Driver/vehicle characteristics in rear-end precrash scenarios based on the general estimates system (GES), SAE-1999–01-0817, Warrendale

Development Process of Forward Collision Prevention Systems

47

Jens Rieken, Andreas Reschka, and Markus Maurer

Contents

An earlier version of this chapter has been published in Eskandarian A (2012) Handbook of Intelligent Vehicles. Springer London

J. Rieken (✉) • A. Reschka • M. Maurer
Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany
e-mail: rieken@ifr.ing.tu-bs.de; reschka@ifr.ing.tu-bs.de; maurer@ifr.ing.tu-bs.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_48

Abstract

Forward collisions represent a significant portion of all severe accidents. This is why appropriate warning and collision avoidance systems are of great importance to increase traffic safety. Different system specifications are subsumed under the so-called FVCX-systems; they differ in their way of affecting the overall system driver–vehicle–environment as forward collision conditioning, forward collision mitigation, forward collision warning, and forward collision avoidance systems. A more specific definition of FVCX-systems is derived by distinguishing them from other related systems such as adaptive cruise control and pedestrian safety systems which can also have an impact on forward collisions. The specifications of the actual systems already on the market can only be understood if the characteristics of machine perception are considered carefully. The progress in the field of machine perception enables the forward collision warning and avoidance systems. There are still limitations of state-of-the-art perception systems compared to attentive human drivers which must be considered when designing FVCX-functions. The state of the art in FVCX-systems is sketched out, highlighting realized examples of FVCX-systems of different car manufacturers. Another focus is on a systematic design process which is recommended for driver assistance systems. The motivation for driver assistance can always be derived from accident research. Already in the early conceptual phase functional safety, legal, ergonomic, and marketing aspects should be taken into consideration. Only if a consistent functional specification is found are further developments including package and architecture aspects justified. Concepts for testing and evaluation should be designed in an early development phase as well.

1 Introduction

1.1 Motivation and Early Research Approaches

Forward collisions represent a significant portion of all severe accidents. This is why systems for obstacle and collision warning have been included in the recommendations of the eSafety Support, one of the leading European initiatives (eSafety 2010). They list systems with high efficiency regarding traffic safety and a high impact on the annual fatalities on the road (Gelau et al. 2009).

Detailed analyses of accidents have revealed that many drivers either do not brake at all or do not make use of the full deceleration capacity of their braking systems. Figure 1 shows the percentage of the drivers who either applied the brake just comfortably or even missed braking altogether although an emergency braking would have been the adequate reaction. Their portion in percent is plotted over the severity of the accidents quantified by the so-called Maximum of the Abbreviated Injury Scale (MAIS), which takes the most severe injury the driver suffered as a measure.

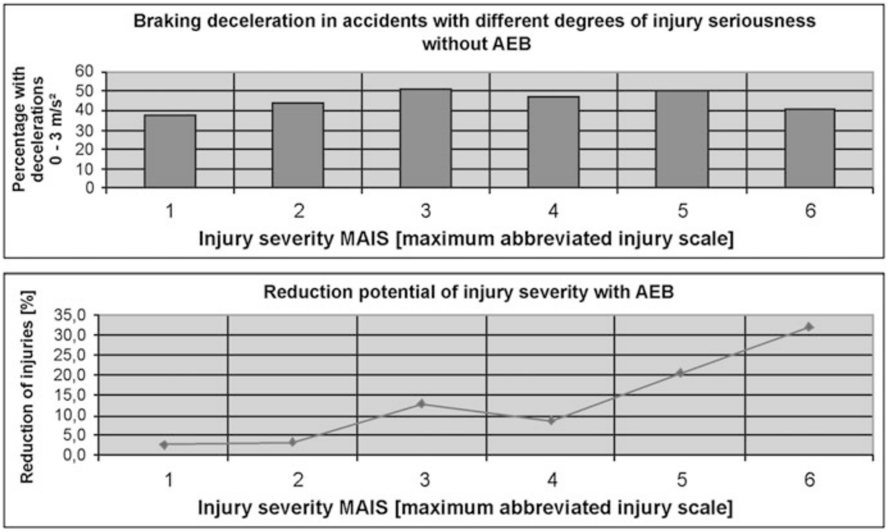


Fig. 1 Theoretical potential of an automatic emergency brake: decelerations at accidents with different degrees of severity of the injuries (nonassisted driving (Kopischke S (2000) personal communication); *AEB* automatic emergency brake, *MAIS* Maximum of the Abbreviated Injury Scale)

Early publications from the 1950s described prototypical systems designed to warn the driver of forward collisions. General Motors built up a prototype that perceived the relative velocity and the distance to the leading vehicle with an airborne RADAR system. Both state variables were indicated at the dashboard (Wiesbeck 2006). It took another 40 years to industrialize appropriate RADAR systems in such a way that they could be produced within economic constraints at least for small numbers of luxury cars. Meanwhile even compact cars are equipped with RADAR sensors, and several million sensors are sold and used in cars nowadays.

Electronic controllable braking actuators were made available in a high percentage of new vehicles by the introduction of vehicle dynamics control systems in the 1990s. For the first time, braking support was feasible without equipping cars with additional actuators.

These actuators were used by the so-called hydraulic brake assist (HBA). Depending on the velocity with which the driver hits the brake pedal, the drivers' intention was to be determined. If an emergency situation was recognized, additional braking force was to be triggered by the HBA system (see ▶ Chap. 46, "Fundamentals of Collision Protection Systems") (Kiesewetter et al. 1997). In practice, it turned out that the driver's intention can be assessed only in rare occasions if just the drivers' pedal velocity is taken into account. Sporty drivers hit the brake pedals so dynamically even in standard situations that they can hardly be distinguished from actions of average drivers in emergency situations. In order to avoid inadequate interventions of the HBA system, the triggering thresholds of these systems are adjusted conservatively nowadays. Therefore, average drivers can only be supported by HBA systems in selected emergency situations. Depending on the design principle, HBA cannot assist if the driver does not brake at all.

By introducing RADAR systems for adaptive cruise control (ACC, ▶ Chap. 45, "Adaptive Cruise Control"), the technological base for machine perception of the outside world has been integrated into modern vehicles. Based on these sensors, systems were proposed in the 1990s which automatically perceive the environment in front of the vehicle and trigger an emergency braking once an accident cannot be avoided any longer within the limits of handling (Kopischke 2000). 2003 Honda released a "Collision Mitigation Braking System" (CMBS) as a first OEM in the market (Sugimoto and Sauer 2005). The basics for collision mitigation systems are covered in ▶ Chap. 46, "Fundamentals of Collision Protection Systems".

1.2 Definitions and Abbreviations

Following the first ideas for obstacle avoidance and collision warning with machine perception, a variety of different systems has been derived. Each system was designed to protect the driver in special types of accidents. A few classifying concepts are defined below which separate the systems into different categories. In addition, a few abbreviations are introduced for easier reading. The abbreviations have been adapted to the ISO 15623 and ISO 22839 standards. ISO 15623 covers forward vehicle collision warning systems, and ISO 22839 covers forward vehicle collision mitigation systems (ISO 15623 2011; ISO 22839 2011).

Active safety:	Collision avoidance is called active safety as well (Naab and Reichart 1998)
FVC	*Forward vehicle collision:* collisions, which involve the equipped vehicle and a vehicle in front of the equipped vehicle
FVCX	*Forward vehicle collisions systems:* systems which react in an appropriate way to reduce the impact of an impending forward collision on the passengers of a vehicle
FVCW	*Forward vehicle collision warning:* systems which warn the driver of an impending forward vehicle collision
FVCC	*Forward vehicle collision conditioning:* systems which condition subsystems in the vehicle in a way that, once triggered, they react faster (brakes), more effectively (brakes, seatbelt pretensioners), or more gently (smart airbags)
FVCM	*Forward vehicle collision mitigation:* systems which reduce the severity of an accident by applying appropriate countermeasures
FVCA	*Forward vehicle collision avoidance:* systems which avoid an impending forward collision by influencing vehicle dynamics
Passive safety:	"Passive safety is the mitigation of the severity of accidents," translated from Naab and Reichart (1998); active systems for

passive safety are supposed to reduce the severity of accidents, e.g., airbags, seatbelt pretensioners, and automatic emergency brakes (AEB), if implemented as a FVCM-system

2 Machine Perception for Forward Vehicle Collision Warning and Avoidance

One reason for the emergence of the wide variety of current system collections is that in many aspects machine perception is inferior to the perception of attentive human drivers. It has been pointed out above that the systems for machine perception are the enabling technology for automotive driver assistance.

In the following, the characteristic features of machine perception systems are explained by a simple comparison: driver assistance systems with machine perception are contrasted with the so-called conventional driver assistance systems which rely on direct measures or model-based observations.

Conventional driver assistance systems support the driver in situations which are easy to measure or to estimate. Antilock braking systems intervene when a wheel is about to block. This can be determined by a conventional wheel speed sensor.

A vehicle dynamics control system brakes single wheels when the estimated sideslip angle exceeds an experimentally obtained threshold. In a sense, vehicle dynamics control systems already include tasks of machine perception: the estimation of the friction coefficient is a very challenging perception task especially if it is to be solved on time in order to adapt the current velocity of the vehicle to the road conditions.

A similar distinction between the two types of driver assistance systems is made in the code of practice for the so-called advanced driver assistance systems (ADAS): in contrast to conventional driver assistance systems, ADAS are equipped with sensors for the detection and interpretation of the environment of the vehicle (Donner et al. 2007).

To qualify as a driver assistance system with machine perception, support has to be given in situations recognized automatically. In an adaptive cruise control (ACC) system, RADAR reflections are interpreted as vehicles once these reflections fulfill certain temporal and spatial criteria. In a lane departure warning system, bright–dark transitions in the video image matching a specific gestalt are understood as a lane with its markings. In a sense, machine perception means the ability to interpret automatically. In the current state of the art in this field, machine perception enables unprecedented possibilities of interpretation but also unprecedented ways of misinterpretation.

When designing an innovative assistance system, the functional limitations of the state-of-the-art machine perception are to be taken into consideration from the very beginning. A successful strategy for the system design may put up with significant misinterpretations (e.g., one unjustified activation per 10,000 km in a safety system) if the system reaction is designed in a way that neither irritates nor endangers the driver. As an example, the design of an automatic warning jerk will be discussed later on (see Sect. 4.2).

As soon as an automatic intervention significantly influences the vehicle dynamics based upon machine perception, false automatic reactions are not accepted at all. In the automobile industry, there is currently no general agreement on how to define an appropriate false alarm rate given the impact of the machine intervention.

In order to increase the correctness of the interpretation process, machine perception systems are equipped with redundant sensors whose data are fused to an internal environmental representation as consistently as possible (see ▶ Chaps. 23, "Data Fusion of Environment-Perception Sensors for ADAS," and ▶ 24, "Representation of Fused Environment Data"). Interventions are only allowed in situations which can be specified formally so that erroneous automatic interpretations become very unlikely. In addition, the tracking of the traffic situation as it is developing over time is exploited in order to verify the machine interpretation. If in doubt, the assisting action is suppressed in order to avoid false reactions which could endanger any traffic participant. In the design of safety systems, this is also called a conservative systems design.

This demand on redundancy is supported by a legal line of arguments which attempts to assess new systems by looking for analogies. Lawyers could argue that also in vehicle dynamics systems, important state variables are perceived with redundant or at least functionally redundant sensor systems.

Given a special situation to be intervened in, the robustness of the machine perception can be further improved by tailoring both sensors and signal processing for this special task. Even though this principle is widely spread in nature where evolution supports the individuals best adapted to their ecological niche, a similar way of tailoring machine perception systems to special tasks is a major barrier for reusing sensors and systems in other assistance systems.

3 Separation from Other Systems and Chapters

In this section, a detailed separation from other systems and chapters in this book helps to define FVCX-systems.

3.1 Separation from ACC

The comfort system ACC (see ▶ Chap. 45, "Adaptive Cruise Control") and different FVCX-systems are basically discussed as two separate system groups, which are closely connected by their technology and their effects on accidents. As already shown in the introduction, the RADAR system of ACC serves as the technological base for the machine perception of FVCX-systems in series production cars.

There is an ongoing controversial discussion how ACC affects traffic safety in general and forward collisions in particular. Users reported that, by intervening automatically, ACC had warned them of dangerous situations or avoided accidents directly.

Within the euroFOT study, the behavior of vehicles with ACC has been investigated focusing on safety and efficiency. As a result, the study shows that safety is increased for passenger cars and trucks (Benmimoun et al. 2012). Considering single cases, ACC works as an FVCX-system in a sense. When using ACC, many drivers accept longer distances to the leading vehicles compared to their normal style of driving (Benmimoun et al. 2012). A reduction of forward collisions is expected as long as the driver supervises the system attentively. There is still a need of further statistical analyses quantifying the impact of ACC as an FVCX-system. As a result of the euroFOT study, there started a discussion among lawyers whether drivers should use ACC mandatorily if available (Vogt W (2010) personal communication).

Apart from the positive impact just sketched on traffic safety, system developers have made sure from the very beginning that the use of an ACC system would not affect the safety of the vehicle negatively. The fundamental principles of the concerns are derived in Bainbridge (1983) or from basic research in psychology (Yerkes-Dodson Law, Yerkes and Dodson 1908). In simple terms, the evident experience is stated that the mental workload should not be further reduced when the driver is already bored. As long as the driver is responsible for the driving task, it has to be ensured that he is sufficiently involved in the vehicle guidance task.

Buld et al. demonstrated in a driving simulator that the drivers' performance related to their driving task would rather decrease than improve with the perfection of the ACC system due to progresses in the technical development (Buld et al. 2005, p. 184). When driving with ACC, the driver may get tired faster than a nonassisted driver.

Therefore, it is a significant milestone in the development process of car manufacturers to test the usability of ACC intensively when developing any innovative system variant. If any doubt about the usability remains, the system variant will be modified to ensure that there are no negative impacts on traffic safety (Neukum et al. 2008). As another aspect, the usability of ACC during long-term operation was also analyzed and intensely documented for the first time by Weinberger (2001).

3.2 Separation from Proactive Pedestrian Protection

Formally, there is a big overlap of proactive pedestrian protection systems (see ▶ Chap. 22, "Camera Based Pedestrian Detection") and FVCX-systems; forward collisions with pedestrians are a very important type of accidents. The diversification of the systems is again caused by the limited possibilities of machine perception and the resulting highly specialized approaches to machine perception (see Sect. 2).

FVCX-systems are designed first of all to protect the passengers of the assisted car. Therefore, the recognition of other vehicles even far ahead of the vehicle is of major importance. The proactive pedestrian protection systems primarily guard the pedestrians outside the assisted vehicle. Therefore, the specialized machine perception has to recognize explicitly pedestrians and react appropriately to their needs. In this meaning, proactive pedestrian protection systems are specialized FVCW-, FVCM-, and FVCA-systems. Even though there is no explicit representation of pedestrians, FVCW-, FVCM-, and FVCA-systems can also guard pedestrians by recognizing them as relevant objects (but not explicitly as pedestrians!) and by reacting properly.

3.3 Separation from Integrated Safety Systems

Integrated safety systems coordinate several safety systems. A system coordinating several FVCX-systems is therefore a special case of an integrated safety system. Kompass and Huber (2009) investigate these further.

3.4 Separation from Evasion Assist

Evasion assist systems adapt the yaw rate of the ego-vehicle either in order to avoid a collision with an obstacle or at least to mitigate the severity of an accident. In addition, the systems can decelerate the vehicle. Evasion assist systems can be seen as a special class of FVCA-systems. In this book, they are discussed in another (▶ Chap. 46, "Fundamentals of Collision Protection Systems").

3.5 Separation from Conventional Assistance Systems for Longitudinal Guidance

FVCX-systems rely on machine perception systems for the outside world. That way they can be distinguished from conventional assistance systems for longitudinal control like a hydraulic brake assist system.

3.6 Summary

To sum up, FVCX-systems use sensors for machine perception of the environment mainly designed for other systems – like the RADAR sensor of the ACC system. With these sensors, they avoid forward collisions or at least mitigate the impact of the accidents. Pedestrians, however, are neither explicitly represented nor recognized in state-of-the-art FVCX-systems as their machine perception systems are not specialized for passenger protection and thus dedicated to the recognition of other vehicles.

4 Design Parameters and Current Realizations of FVCX-Systems

As FVCX-systems are to reduce the likelihood of a forward collision and as a result increase the safety of the passengers, they address situations in which a driver is potentially threatened with being involved in an accident. FVCX-systems intervene if there is an increased likelihood that unassisted driving will end in a collision.

A reliable automatic situation assessment is crucial for the adequate intervention of FVCX-systems. In this context, the item “situation” means that objects are not only described by spatial and temporal representations but also by their meaning for the ego-vehicle and its goals. For robust situation assessment, the FVCX-system needs reliable recognition of the relevant objects in the driving environment by using machine perception as well as an unambiguous perception of the drivers’ intention.

The intentions of both the driver to be assisted and the drivers of other vehicles are relevant for a proper situation assessment. In the general case, these requirements exceed state-of-the-art technology. Any machine perception system available nowadays has relevant system limitations causing wrong assessments also in series production cars unless their field of operation is strictly limited. It is also impossible with the state-of-the-art technology to recognize the drivers’ intention in all situations automatically. Again, this uncertainty is dealt with by strictly limiting the situations of interventions.

Two degrees of freedom help to design FVCX-systems ready for the market with today’s automatic situation assessment: the severity of the intervention and the – limiting – definition of situations of intervention.

4.1 CU-Criterion

The so-called CU-criterion is of special importance when limiting the FVCX-situations (CU: collision unavoidable). If an accident were unavoidable, even the best driver would not be able to escape from the collision. Legally, this driver is close to the “ideal driver” who drives as well as the best 2 % of all drivers. If the FVCX-systems trigger only when an accident is unavoidable, rash reactions can be excluded. In rare situations – also called pathological situations – these rash decisions could lead to fatal consequences.

A short reflection may illustrate the relevance of the CU-criterion: if an automatic emergency brake is triggered while the assisted car is overtaking and before a collision is unavoidable, then an accident could even be caused by the trigger in case the driver would have finished the overtaking safely without being decelerated by the intervention. In this example, it is assumed contrary to state-of-the-art systems that the system would react to oncoming traffic (see Fig. 2).

The CU-criterion influences the definition of many interventions essentially. It is still good practice to allow automatic emergency braking only if the CU-criterion is true: an automatic emergency brake, i.e., a braking intervention with maximum

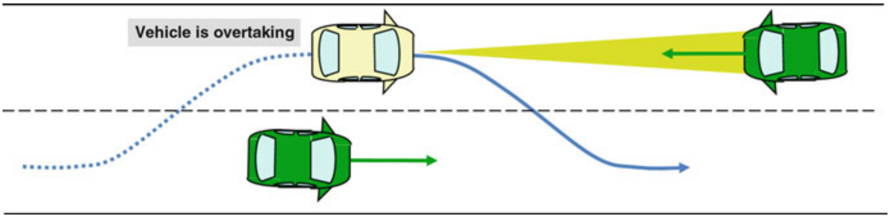


Fig. 2 "Pathological case" of an automatic emergency brake: braking during overtaking

deceleration, is triggered if the accident cannot be avoided by any means due to the limits of handling. So the driver is given any freedom; he is only overridden by the automatic system if he cannot avoid the collision assuming perfect driving capabilities and often even perfect weather conditions (Kopischke 2000).

A thorough analysis shows that the CU-criterion is especially important if emergency situations with high relative velocity occur. In this case, evading by steering is still possible while braking would avoid the accident no longer. At low relative velocities, the situation changes; even when evading is no longer possible, braking can still be an option to avoid the accident.

In Fig. 3, the minimal distance at which the avoiding maneuver is to be triggered is given as a function of the relative velocity. The solid black line therefore illustrates the braking distance depending on the relative velocity. The dotted line sketches the distance at which an obstacle with a width of 2 m can still be evaded by a steering maneuver. Left from the point of intersection of the two curves, the accident can be avoided by braking even if escaping by steering is no longer possible. This alters when the relative velocity increases which is shown on the right of the point of intersection.

Note that the CU-criterion discriminates whether an FVCX-system is a passive safety system (FVCM-system) or an active safety system (FVCA-system).

4.2 Fundamentals of Driver Warning

With respect to the limited reliability of machine perception systems currently available and to the risks of product liability linked to these limits, FVCW-systems are of great importance.

Regarding limited reliability and correctness of today's machine perception systems and strongly connected product liability issues, FVCW-systems demand a special consideration. The driver should be warned in time, so that he can still avoid the accident himself. In addition, the driver should not be irritated by false alarms. A more detailed analysis has shown that the time window for sensible warning is extremely short in the general case as the intention of the driver is not known.

A short example illustrates the challenge when interpreting the situation automatically. The driver to be assisted approaches a slowly driving commercial vehicle

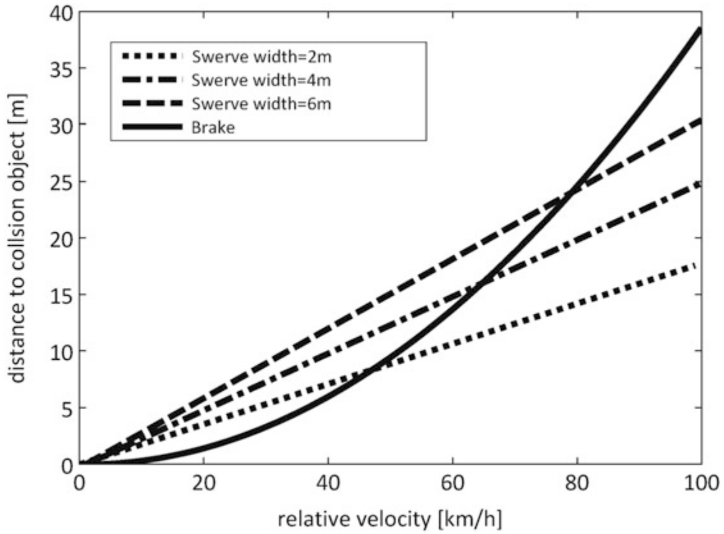


Fig. 3 Influence of the relative velocity upon the CU-criterion; “width = 2 m” means that the obstacle ahead is 2 m wide

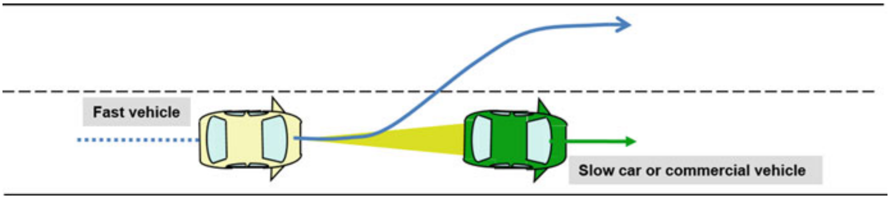


Fig. 4 Warning dilemma: approaching a slow commercial vehicle with a high relative velocity (Lucas B (2002) personal communication)

with a high relative velocity on the right lane of a highway with multiple lanes per direction. The lane left to the ego-vehicle (and the truck) is empty; a lane change toward this lane is both feasible and permitted. A warning in time to decelerate the vehicle behind the truck might be way too early for a sporty driver intending to change lanes shortly before contacting the commercial vehicle (see Fig. 4).

As this time window is so short, it is important that the driver is assisted by warnings easy to interpret and pointing intuitively to the danger ahead. Experiments show that haptic warnings by braking jerks or by a reversible belt pretensioner are very intuitive (Färber and Maurer 2005). At a braking jerk, brake pressure is shortly increased and directly afterward released again with a steep rise so that the jerk is noticeable for the passengers but the vehicle will not be slowed down significantly.

The results referred to show that the driver is made to look outside the front window but does not brake automatically. Similar reactions are reported in studies about jerking with a reversible belt pretensioner.

The example sketched above further underlines the immense importance of observing the driver's condition. Is he/she distracted by secondary tasks – because he/she is entering destinations into the navigation system or using a hands-free cellular device? Is he/she fatigued or is he just enjoying his/her dynamic style of driving being totally aware of the overall situation? Experience in the design of warning systems revealed that already relatively simple real-time warning models can significantly contribute to ease the warning dilemma (Mielich W (2005) personal communication).

In scientific publications, two kinds of warnings are distinguished: latency warning and prior warning (e.g., ▶ Chap. 46, "Fundamentals of Collision Protection Systems"). A latency warning can be appropriate if there is no danger at all as long as the situation continues stationary. But even a minor distortion could lead to unavoidable accidents. In textbooks, the classic examples of these latent threats are vehicles following the leading vehicles within very short distances. A prewarning is triggered when an accident can be predicted on the base of the current state variables.

4.3 Levels of Assistance in Dangerous Situations

Modern intervention strategies obey several partly contradicting principles:

- The vehicle should intervene in time so that the driver can avoid the accident.
- The level of assistance should be appropriate in the sense that the driver is supported but neither he nor the passengers and other traffic participants are irritated by exaggerated reactions.
- The impact of interventions at the wrong time is to be minimized so that traffic safety is not endangered.

These basic principles have led to different levels of the interventions in all FVCX-systems in the market. An important parameter for selecting the countermeasure is the time-to-collision (TTC) – time expected to pass until the vehicle crashes into an obstacle. Psychological experiments have revealed that also for humans the TTC is the decisive measure for situation assessment (Gibson 1950; Färber 1986).

In the following, a variety of single assistance functionalities is sketched. Each of them is part of a current realization of an FVCX-system placed by various car manufacturers in their models (see ▶ Chap. 46, "Fundamentals of Collision Protection Systems"):

4.3.1 FVCW-Systems

FVCW-systems warn the driver so that he/she is able to perceive the hazard and to prevent any accident. The severity of the warning depends on how much time is left for the driver to avoid the accident. To warn exactly in time, many systems analyze the actions of the driver either by directly observing them visually (e.g., Lexus) or by interpreting his/her style of driving based on current driving data. In ▶ Chap. 36, “Driver Warning Elements”, warning elements for the driver are investigated further.

- **Optical warning**: Warnings can be symbolic and text messages signaled by warning lamps or displays. Optical latency warnings may arise at very close distances to leading vehicles. Optical and acoustic prewarning support when the reaction time to the leading vehicles is shorter than a given threshold.
- **Acoustic warning**: Gongs, buzzers, or other sounds are meant to direct the driver’s attention to dangerous situations.
- **Braking jerk**: Significant short-period changes of the actual acceleration by introducing a short pressure pulse to the brake system are known to be a powerful warning means.
- **Reversive belt pretensioner**: Warning jerks from the reversible belt pretensioner have been proven as appropriate warning devices.
- **Active gas pedal**: As long as the driver hits the gas pedal, a growing resistance of this pedal can signal deceleration needs to the driver (e.g., Infiniti)

4.3.2 FVCC-Systems

FVCC-systems condition the vehicle in such a way that in an impending dangerous situation the likelihood of the driver to survive is optimized. Actuators contributing to active and passive safety can be preconditioned accordingly.

- **Prefill**: In case of an impending FVC, a light pressure is established in the brake booster – experts even speak about prefilling the brakes. Thereby, the delay time is reduced as soon as the driver hits the brake pedal or another FVCX-system requires brake pressure electronically (e.g., Audi A8).
- **Adaptive brake assist**: As soon as a hazardous situation is recognized by the machine perception system, the threshold defining the trigger point of the hydraulic brake assist is reduced (Zanten van and Kost 2009).
- **Adaptive dampers**: At the same time, the damping parameters are adapted to reduce the stopping distance of the vehicle (e.g., Audi A8).
- **Reversible seat belt pretensioner**: The slack of the seat belt is reduced by a reversible seat belt pretensioner after fastening the belt. A further reduction of the slack is realized shortly before the crash (e.g., Audi A8).
- **Preset of the airbag**: Preset functions help to speed up the decision-making process between crash and no-crash situations based upon machine perception. The additional information is taken into account at the time of collision (Mäkinen et al. 2007), e.g., Audi A8.

4.3.3 Rear Impact Countermeasures

Automatic emergency braking is not appropriate to avoid all accidents or to mitigate in all traffic situations. Consider the following scenario: A forward collision with a lightweight traffic participant could be avoided by an automatic emergency braking, but a much heavier vehicle coming from behind and unable to decelerate fast enough crashes into the rear of the assisted vehicle. This is why FVCX-functionalities are often accomplished by systems protecting from the following traffic.

- **Hazard warning lights**: It is almost standard that the hazard warning lights are activated if the vehicle brakes with full capacity (automatically).
- **Rearward-looking sensors**: More sophisticated FVCX-systems exploit additional rearward-looking sensors to analyze whether the damage of an automatic emergency braking might exceed the benefits.

4.3.4 FVCM-Systems

FVCM-systems use actuators in the vehicles to reduce the severity of an impending collision.

- **Automatic partial braking**: Partial braking is applied automatically to reduce the relative velocity and to warn the driver more drastically. They are applied in different strengths when the accident is still avoidable, e.g., Audi A8: level $1:3\frac{m}{s^2}$ level $2:5\frac{m}{s^2}$ maximum allowed value according to ISO 22839 standard, $6\frac{m}{s^2}$ (ISO 22839 2011).
- **CMS-braking**: CMS (collision mitigation systems) were introduced in Japan starting from 2003. After a mandatory warning, the CMS-braking is activated if the driver cannot avoid an accident. The CMS must decelerate at least $5\frac{m}{s^2}$ (ISO 22839 2011), e.g., Honda CMBS (Bishop 2005; Sugimoto and Sauer 2005).
- **Automatic emergency braking (CM)**: An automatic emergency braking is triggered if an accident is unavoidable. Depending on the current friction coefficient, automatic decelerations can reach up to $6\frac{m}{s^2}$.
- **Reversible seat belt pretensioner**: Shortly before an unavoidable crash, the seat belt is tensioned by the reversible seat belt pretensioner in order to bring the driver and the passenger into an upright position and to avoid "submarining" (Mäkinen et al. 2007).
- **Closing windows and sunroof**: Once a dangerous situation is detected, the windows and the sunroof are closed automatically. This functionality was introduced first by Daimler as part of the pre-safe system (pre-safe: first version in 2002, Schmid et al. 2005).
- **Seat adjustments**: As part of the pre-safe system, the position of the passenger is influenced by adjusting the seats as well (Schmid et al. 2005).

4.3.5 FVCA-Systems

FVCA-systems try to avoid an accident.

- **Target braking**: A target braking is an extension of the hydraulic brake assist. In dangerous situations, the braking driver is assisted by additional brake pressure supplied automatically in order to avoid the accident.

- **Automatic emergency braking** (**CA**): An automatic emergency braking is triggered right in time to avoid an accident. Depending on the current friction coefficient, automatic decelerations can reach up to 1,0 g.
- **Other systems**: ▶ Chaps. 46, “Fundamentals of Collision Protection Systems,” and ▶ 50, “Intersection Assistance”.

5 Levels of Assistance in an Actual Realization

Currently available FVCX-systems are exemplary sketched for An Audi A8 with “Braking Guard” and “PreSense” in this section. These systems reflect the current state of the art. They are adapted to other vehicles within the VW group as well (Bentley, Audi A6, Audi A7, and VW Touareg). With friendly permission of the Audi AG, further background information can be supplied.

The vehicle uses data from two forward-looking RADAR sensors (Bosch, ACC3, 77 GHz) (see ▶ Chap. 17, “Automotive RADAR”), a monocular video camera (Bosch, 2nd generation), and ultrasonic sensors (see ▶ Chap. 16, “Ultrasonic Sensors for a K44DAS”). Rearward-looking RADAR sensors are used to monitor the following traffic supplying mainly data to the lane change assist system (Hella, 24 GHz, 2nd generation; see Sect. Hella 24 GHz Mid Range RADAR) (see Fig. 5).

FVCX-systems are activated in several steps. In the first phase, the brakes and the dampers are preconditioned (Prefill, hydraulic brake assist, adaptive dampers). At the escalation level, the driver is warned – first acoustically and optically, then by a warning jerk. In parallel, the reversible seatbelt pretensioner reduces the slack of the belts of the driver and the codriver. If the driver still does not react

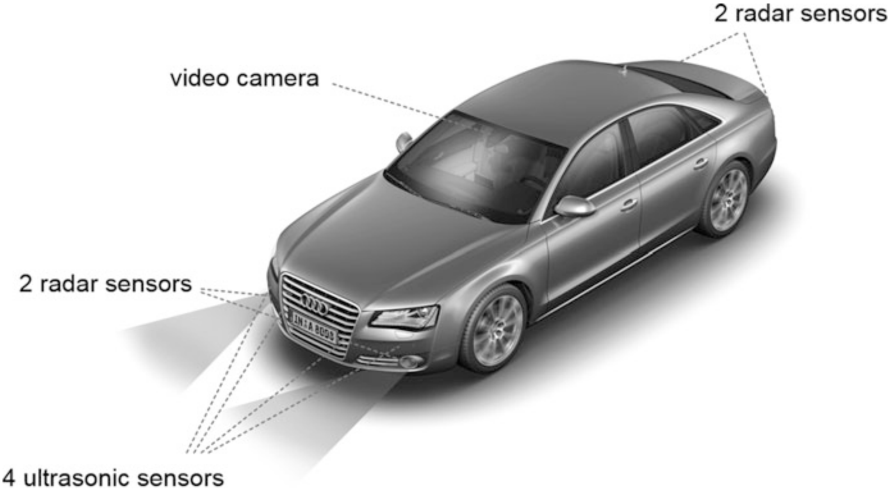


Fig. 5 Sensors for environmental machine perception in the Audi A8 (Duba GP (2010) personal communication)

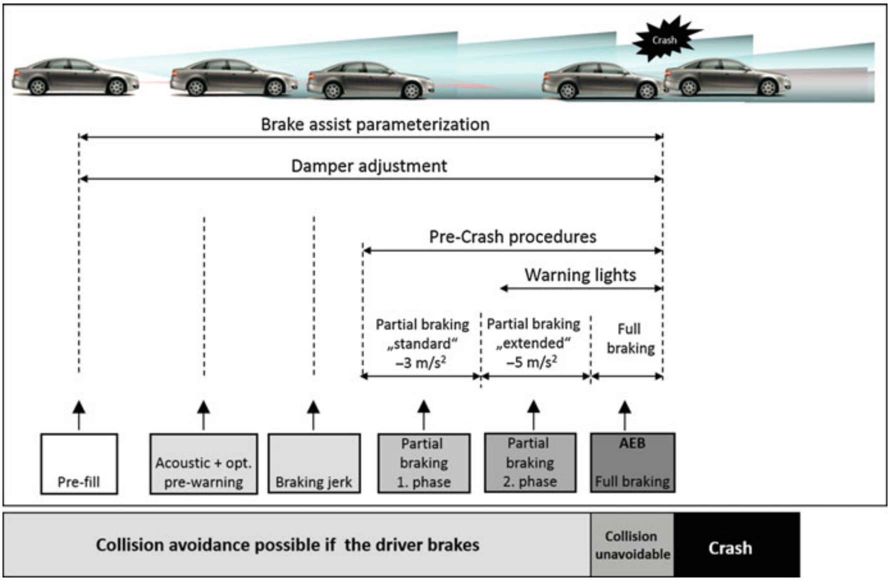


Fig. 6 Levels of escalation in the Audi A8; *AEB* automatic emergency braking (Duba GP (2010) personal communication)

appropriately, the automatic partial braking (first, $-3\frac{m}{s^2}$; second, $-5\frac{m}{s^2}$) and, after reaching the CU-criterion, the automatic emergency braking are activated in rapid succession. In addition, the sunroof and the windows are closed automatically; the reversible belt pretensioner increases the tightening force again. The automatic braking is signaled to the following traffic by hazard warning lights activated automatically as soon as the stronger partial braking has been triggered (see Fig. 6).

6 System Architecture

6.1 Overview

The demand on redundant multimodal environmental sensors leads to big data streams in the communication systems of modern vehicles. Availability, reliability, and the overall system safety require adequate communication technology. Time-triggered transmission and architectures in electronic control units (ECUs) would be helpful when fusing sensor data, but eventually also for the precise control of innovative actuator systems (e.g., smart airbags).

Therefore, system architecture and its thorough planning are key factors for mastering the complexity of connected safety systems. Advisably, machine perception is taken into account already in the planning phase of the topology of the vehicle’s communication systems. Data streams occurring at the fusion of sensor

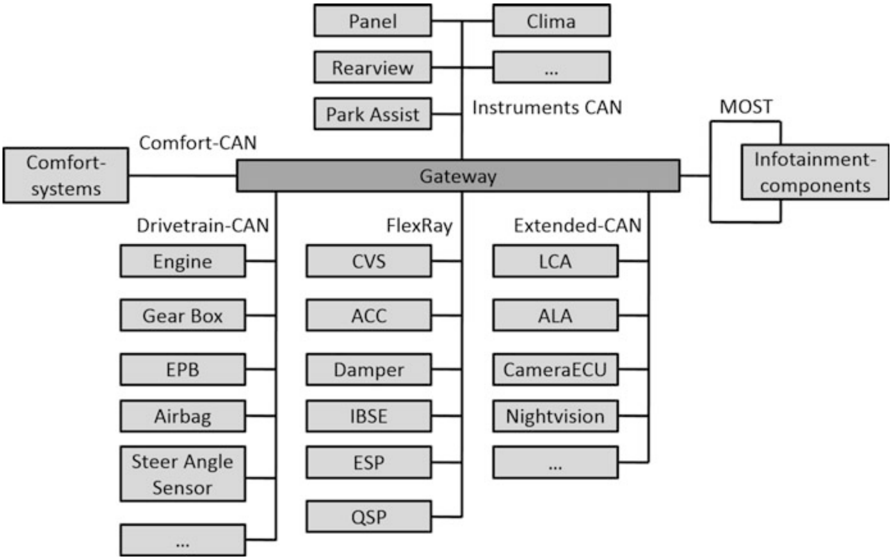


Fig. 7 Electronic hardware architecture of an Audi A8i (Kötz J (2000) personal communication)

data derived from environmental perception can determine the topology of in-vehicle communication systems.

As an example, Fig. 7 shows the electronic hardware architecture of the current Audi A8 (Kötz J (2000) personal communication). A central gateway connects several CAN networks, a MOST bus for multimedia purposes, and a FlexRay cluster for driver assistance and FVCX-systems. The latter provides connection to ECUs for video-based environment perception, ACC and damper control, a specialized ECU for an inertial-based state estimation, the ESP, and Quattro Sport. All components have to collaborate precisely in order to provide an appropriate FVCX functionality.

Up to now, system architecture was discussed in a quite traditional way, as collection of ECUs, vehicle networks, and gateways. For future systems, it is necessary to include additional aspects to handle the more and more complex systems of modern vehicles.

The functional system architecture splits the overall system into their functional modules from the aspect of the overall functionality. It exploits ways of representation from the fields of system dynamics and control theory (Maurer 2000). Additionally, an explicit knowledge representation should be part of the system development to provide a central storage for the current state of the vehicle. This can be modeled by methods from object-oriented software development. In current systems, this knowledge is available but hidden, mainly represented in a central diagnosis structure. In regard of an increasing automation level of the vehicles, knowledge about the state of the vehicle is essential for providing suitable reactions (Maurer 2000).

The properties of the vehicles should be described from the perspective of the customers and independent of their technical realization to start with (Kohoutek et al. 2007).

These three aspects are ideally discussed independent of the hardware; they stay untouched when migrating to another hardware platform. The hardware itself and aspects of low-level programming belong to the hardware-dependent aspects of the system architecture.

In the framework of AUTOSAR (▶ Chap. 7, "AUTOSAR and Driver Assistance Systems"), also the low-level software is increasingly standardized, so these aspects can be discussed more independently from their actual technical implementation. Some car manufacturers may pay additional attention to these aspects of system design to allow for a proper handling of future systems' complexity.

6.2 Functional System Architecture

The functional system architecture discusses the structure of the system independently from the hardware. It has been observed that there is an ongoing change of the hardware architecture in the different phases of development (research phase, concept phase, predevelopment phase) driven by the fast progress of new technologies.

In comparison, functional system architectures will only be changed if functional extensions are necessary or completely redesigned if major paradigms are revolutionized. The functional system architecture enables hardware-independent interface analysis; it reveals how appropriate interfaces between two modules can be designed. The topology of the hardware should be derived from thorough analyses of the functional system architecture.

Figure 8 sketches a simple block diagram of the system driver–vehicle–environment–driver assistance system. In the system architecture, a parallel structure

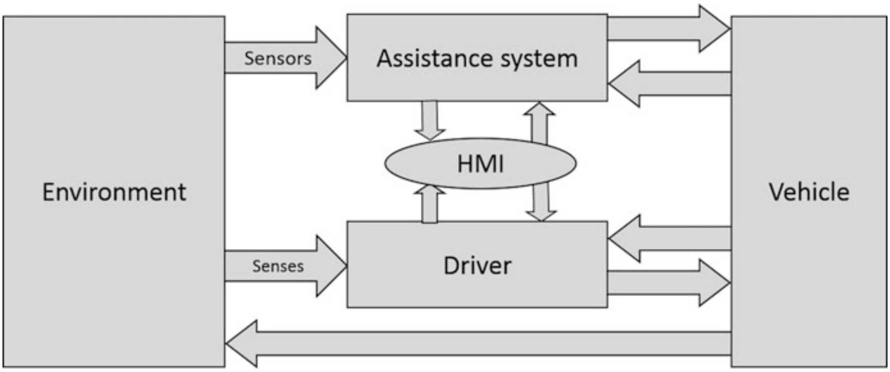


Fig. 8 Simplified block diagram of the system driver–vehicle–environment–driver assistance system (Kopf 2005)

results because the driver and the assistance systems run the same task in parallel by definition (see (Kraiss 1998)). Both the driver and the assistance system observe the environment and the ego-vehicle with their senses and the technical sensors, respectively. They influence the vehicle with appropriate actuators according to their goals. Driver and assistance system communicate via human–machine interface. Interactions between the vehicle and the environment are not displayed.

7 Design Process

7.1 Systematic Design of Driver Assistance Systems

Many developments and many design tools have been driven by military purposes. In the field of complex technical systems, the so-called V-model originally developed for defense systems has established a basic pattern for many other design schemes.

The V-model supports different fundamental design principles helping to structure complex systems. First of all, it supports the top-down design from overall requirements on the system level stepwise down to the detailed requirements on the component level. In the V-model, it is very important to specify appropriate test cases for each requirement. Corresponding to the top-down structure of the requirements, a bottom-up structure of the test cases occurs.

The introduction of the V-model as a paradigm in the design process of electronic vehicle systems leads to a significantly more structured way of development with car manufacturers and their system partners (e.g., Breu et al. 2007). The more the requirements are specified in detail, the more obvious it becomes that the test coverage of complex assistance systems is limited.

It is discussed critically in scientific publications that the V-model may not be appropriate if the information base is not yet complete at the beginning of the design process and therefore the system cannot be developed top-down (e.g., Reif 2006). In reality, the design proceeds incrementally and iteratively; many steps of the V-model or even the whole V-model are processed several times (Schäuffele and Zurawka 2006).

This demand for iterative design loops was regarded by a simple design model, which was developed during the research project “Automatic Emergency Braking” at Audi. The process was visualized in a simple diagram.

Figure 9 shows a full circle containing a complete iteration loop. An abbreviation path is defined after less than half of the circle leading back to the starting point of the development process. A more technical form of the notation was presented in 2006, but not continued during the last years (Glaser 2006).

As a result, two iterative loops emerge from the structure described above: The first loop is much shorter and saves resources; it requires expert knowledge from different departments. The tasks are performed either theoretically or in more advanced labs supported by a chain of concatenated X-in-the-loop tools (Bock 2009); no prototypes are built up during the inner iterative loop. The approach is

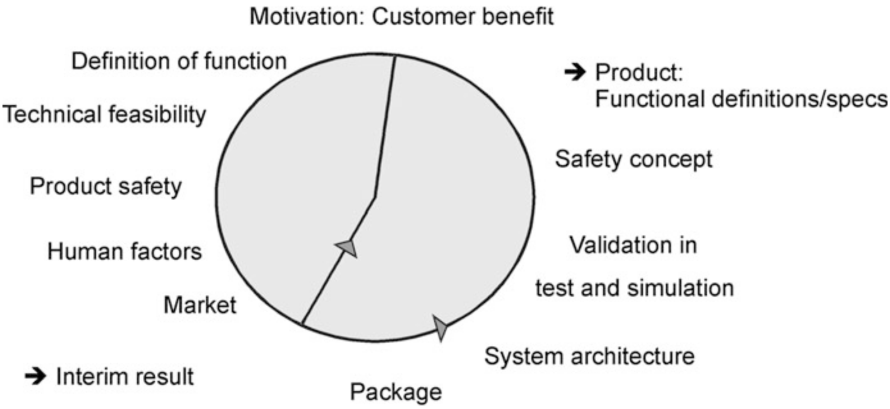


Fig. 9 Systematic design of driver assistance systems (Maurer and Wörsdörfer 2002)

extremely powerful if the experts available within the car manufacturer – supported by external experts if necessary – identify the basic design conflicts within the inner iterative loop; in addition, they profoundly distinguish between realizable and desirable, but non-realizable assistance functionalities.

Prototypic systems will only be built up if the experts agree to a function definition as a preliminary result solving all design conflicts revealed during the theoretical discussion. Sometimes experimental setups may be required even in the iterative loop to solve basic questions.

The needs of the driver and the assistance to him are always the starting point of the design process. This may sound trivial. However, the reader interested in design of automobiles will immediately recall many examples where the driver needs were not in the focus of the system design. Newspapers and journals are full of examples (e.g., Bloch 2007). Note that for the commercial success of the system, the subjective driver needs, not the objective impact of the system, can make the difference.

Based on these ideas, possible assistance functions are derived and tested by the expert whether they can be realized with state-of-the-art technology. Can the functional gaps and the systems failures be controlled by untrained users in each situation? Is a user-transparent design of the assistance function and its limitations possible? Are there any sensible human–machine interfaces? Is the system to be designed financially affordable for the customer? Does it fit to the branding of the car manufacturer? A more detailed discussion will include a practical example and a full iterative loop in the following section.

The approach described above supplements the class of design processes collected in the field of integrated product development (e.g., Ehrlenspiel 2003). This design scheme should be taken into account in any research and development phase of a system. User-centered and holistic design should be mandatory in academic research.

In the phase of industrial research and predevelopment, the design processes are important for the commercial success of the manufacturer. The fine adjustment is performed during the series development phase especially if innovative machine perception is involved; prototypes of the sensors only available shortly before market introduction reveal whether the specification gathered at the beginning of the project will be met by real production-type sensors. If they do not fulfill the specifications, it may become necessary to adjust the functionality shortly before start of production by adding yet another loop in the design scheme.

Of course, there should be open research and predevelopment projects, not directly addressed to particular user needs. But it is important that these projects are declared accordingly and do not suggest specific customer benefits.

7.2 Example: Systematic Design of an "Automatic Emergency Brake"

7.2.1 User-Oriented Definition of the Function

Analyses of accident research have revealed that many drivers do not exploit the full deceleration capacity of their vehicles. In Fig. 1, the statistical evaluation of a database for accident research was displayed. Remember that in the tests, a significant percentage of the driver either did not brake at all or applied a comfort braking although an emergency braking would have been the appropriate reaction to avoid the accident or at least to mitigate its impact (Zobel R (1999) personal communication; Kopischke S (2000) personal communication; cited from Maurer 2009).

Based upon these identified assistance needs, a first function is defined to start the conceptual development phase:

An automatic emergency braking, i.e., a braking intervention with maximum deceleration, is triggered if the accident cannot be avoided by any means due to the limits of handling. So the driver is given any freedom; he is only overridden by the automatic system if he cannot avoid the collision assuming perfect driving capabilities and often even perfect weather conditions (Kopischke 2000) (CU-criterion; see Sect. 4.1).

The definition of the function incorporates the knowledge available at the beginning of this conceptual development phase: From the beginning the system, functionality is limited to accident mitigation to avoid later liability claims of drivers and their relatives who could otherwise argue that the automatic emergency brake had been triggered too early and had even caused the accident.

During the first design loop, the experts report that RADAR, LIDAR, or video sensors are available to realize this function as long as the scenarios are easy to describe and the weather conditions are within a certain specification. As a constraint, it has to be analyzed during this first loop of iteration whether the function can be realized just by a RADAR sensor of a conventional ACC system. Latest during the first risk analysis, it becomes obvious that there are many possible traffic situations which exceed every single sensor principle. Missing triggers of the

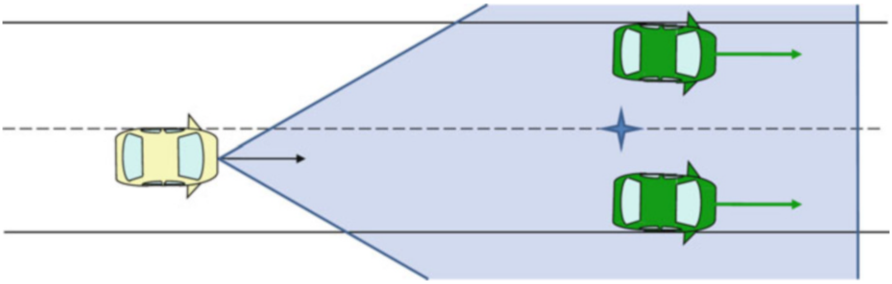


Fig. 10 “Ghost” objects perceived by a RADAR system

automatic emergency brake are regarded less critically; in this case, the assisted vehicle is not less safe than a conventional vehicle.

In contrast, an event is considered hazardous if an automatic emergency brake is activated although the CU-criterion is not fulfilled. As the fundamental principles of the single sensors are known, it is obvious to the experts that unintended activation of the automatic emergency brake may be rare, but not impossible due to the current state of the art.

RADAR experts are familiar with situations in which “ghost” objects occur. That means the sensor reports objects which do not even exist. For example, this can be the case if two vehicles move with a very similar velocity and are interpreted as a single “ghost” vehicle driving in between the two real vehicles (Fig. 10). It is easy to imagine how a “ghost” object could cause an erroneous reaction of the automatic emergency brake.

In the design discussions, it will be taken seriously when experts in product liability argue that in the event of a failure, courts will look for analogies. A likely analogy will be taken from machine perception where important state variables are measured or derived redundantly in a vehicle dynamics control system. Therefore, the automatic emergency brake has to be designed with redundant means for perceiving the decisive parameters (▶ Chap. 41, “Vehicle Dynamics Control with Braking and Steering Intervention”).

Even in this very early phase, experts underline that the expected limits of the function should be comprehensible for the driver and the other traffic participants. They emphasize that the manufacturer is responsible for the expectations of the customer. This expert knowledge is incorporated in variant design aids, thanks to the projects in the framework of the “Response” projects (e.g., Kopf 1999; Donner et al. 2007).

As an additional requirement, the system has to monitor itself, realize any substantial degradation, and warn the driver adequately. In order to prove that the system has worked without any malfunctions, data recorders are regarded as a sensible extension of advanced driver assistance systems.

In the discussion of possible hazards, it is crucial whether false triggers of an automatic emergency brake are controllable for the driver and the following traffic. The thorough analysis of this question requires building prototypes for the first time

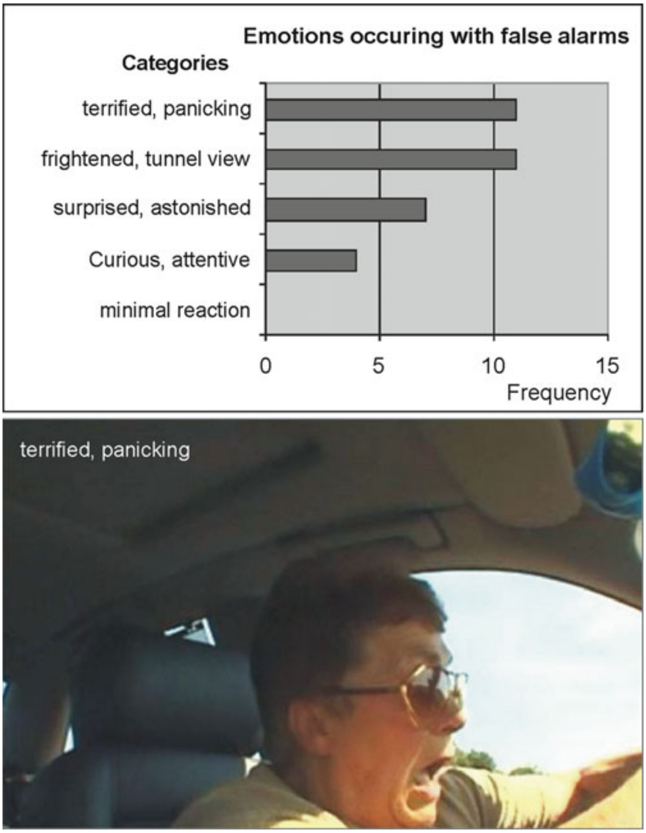


Fig. 11 Drivers' reaction to an unjustified trigger of an automatic emergency brake (Färber and Maurer 2005)

in the concept phase and therefore a first iteration of the outer design loop. The test results are unambiguous: More than a third of the drivers' reaction are categorized as "terrified, panicking." Another third of the drivers react "frightened, with tunnel view." It cannot be excluded that the comparably mild ("surprised" or "curious") reactions were evoked because the experiment was performed at a test track closed for public traffic (Fig. 11, Färber and Maurer 2005).

These analyses underline that false alarms of an automatic emergency brake can cause a significant risk for the driver, the following traffic, the car manufacturer, and the system partner. In addition to the technical, ergonomic, and legal aspects, product marketing should be addressed already during the concept phase. How do costly technical innovations contribute to the manufacturer if they do not fit into the image of the brand? They will not be promoted by the manufacturer and bought by the customers. If it comes to assistance functions, the situation is even more complicated: as a result of the expected functional limits, the products will not be promoted aggressively anyway. The manufacturer is responsible for the customers' expectations.

From the first iterative design loop, the following is learnt: A functional definition has been identified with a big potential impact to traffic safety. The sensors specified in the development task limit the benefits to longitudinal traffic. The realization with ACC sensors already launched in the market would be inexpensive. Comparisons with other safety systems reveal that a redundant perception of the most relevant state variables is strictly required. Experimental analyses which had to be performed during this early stage of development underlined that false triggers of the functional definition mentioned above are not acceptable.

As a consistent preliminary result was not found during the first loop, the further development has to be modified fundamentally. Within a longer-term perspective, the false alarm rate should be minimized by perception systems as complementary as possible. In the short-term, a consistent function should be reached by varying the functional definition (Sect. 7.1).

In the tests, the effect of the false emergency braking was impressive. Could not a weak braking jerk point the driver to a hazard ahead by just warning him with a haptic jerk? In case of a mistimed jerk, the following traffic would then not be endangered by a sudden unexpected deceleration of the velocity.

Experimental analyses confirm both expectations. The warning jerk is an efficient warning device; with an appropriate braking system, there is no noticeable deceleration. In the next iterative loop, a warning system is therefore matured for the market assisting the driver as described above. As the intervention is uncritical even if it is unjustified, a false alarm rate of 1 per 10.000 km seems to be acceptable.

This time the preliminary results look promising: The warning via the haptic channel is both direct and effective. Thus, a significant customer benefit is predicted. The impact again is limited to longitudinal traffic as the system design is based upon ACC sensors. The function can be realized without any additional sensor hardware. False alarms are both controllable and accepted by the drivers. This function can be offered to the market shortly after the concept phase (product name: Audi, "Audi Braking Guard"; VW, "Front Scan"; market introduction, 2006).

The further development of the original idea of an automatic emergency brake requires more sophisticated solutions. Quantitative prognoses can be given for the functional definition developed during the first design loop. The analysis which parameters influence the benefits most also has great importance. Figure 12 shows how the relative energy reduction and, therefore, the impact on accident mitigation depend on the system delay. This representation may even be helpful to communicate the benefit of a faster brake system within the car manufacturer or to select sensors which fulfill the requirements on the dynamics (Kopischke S (2000) personal communication) quoted by Maurer (2009).

During the first design loop, it had become apparent that crucial state variables are to be perceived redundantly. The selection of an appropriate sensor configuration is one of the most challenging design tasks when developing an innovative driver assistance system. In the general case, appropriate sensors which fulfill all

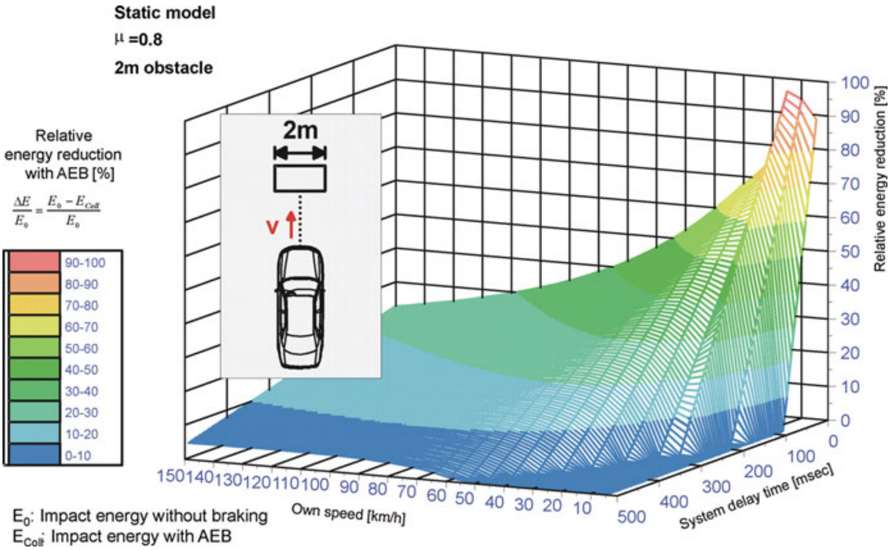


Fig. 12 Sensitivity of the relative energy reduction of an automatic emergency brake to the delay time (Kopischke S (2000) personal communication)

requirements derived from the functionality are not available on the market. Even measures to compare different sensor sets and perception algorithms are not yet established. So the selection of the sensor set will be based on the performance of the current prototypes and the performance predicted by the developers for their sensors in series production.

Apart from this uncertainty, the robustness of the machine perception can be increased by combining appropriate sensor principles. To establish both robustness of the perception system and formal redundancy, combinations of many different sensor principles are taken into account. A long-range RADAR sensor is the preferred sensor of the developers responsible for the design of adaptive cruise control (▶ Chap. 45, “Adaptive Cruise Control”) due to its performance in adverse weather conditions. A mono-video camera is becoming standard for lane departure warning and traffic sign recognition systems. At the point of decision, it was not sure whether additional redundancy would be needed. Therefore, stereo vision, laser sensors (▶ Chap. 18, “Automotive LIDAR”), and photonic mixing device (PMD) were analyzed as an additional forward-looking sensor. Finally, the sensor system consists of two RADAR sensors and a video camera (Duba GP (2010) personal communication; Lucas et al. 2008).

The reliability of this highly sophisticated sensor set is significantly higher compared to a single ACC RADAR sensor. In order to further reduce the likelihood of dangerous false interventions of an automatic emergency brake, the data of the rearward-looking RADAR system are exploited as well. An emergency braking is only triggered with full braking power if there is no vehicle following the ego-vehicle closely.

The benefit of the system is again limited to longitudinal traffic. The likelihood of false alarms is minimized. In this third iterative loop, a few aspects from the outer loop are focused on, too: packaging and functional tests. Aspects of the system architecture are treated in another section (Sect. 6).

Outside the car manufacturer, the integration of the sensors in the design concept is widely underestimated. Package spaces are threatened in any car especially if needed for other functions than the basic functions of the vehicle; this is even true if they have been reserved in the early concept phase. The integration of the ACC sensors in the Audi A8 was done quite elegantly. The fog lights were directly integrated into the head lights; the “old” package space for the fog lights was reused by the RADAR sensors.

7.2.2 Functional Tests of Driver Assistance Systems

Nowadays the usage of the term “testing” is not specific in the automobile development. It encompasses different test categories like functional tests, usability tests, testing user transparency, customer acceptance tests, testing electromagnetic interference, climate stress tests, acoustic tests, crash tests, electric and electronic tests including hardware-in-the-loop tests, and software integrity tests. The list can easily be extended. In the field of driver assistance, each topic itself is challenging, complex, and worth its own chapter in this book. Many groups in the technical development and in the quality assurance department are in charge of and contribute to the testing process.

This section is limited to the aspect of functional testing, again illustrated by the design example of an automatic emergency brake. Two failures are of outstanding importance because they decide on how the system is perceived by the driver and the public. It has already been reported that the drivers’ reactions to unjustified interventions were significant. So they have to be avoided by all means. The experimental setup for tests of justified interventions is even more challenging as they will always lead to a crash according to the functional definition.

The false alarm rate (activation of an emergency brake without regarded as necessary by human drivers) is a crucial factor for the acceptance of an automated emergency braking system with the given functional extend. It is still an open question which error rate would be accepted in society. The IEC standard 61508 specified acceptable error rates depending on the safety integrity level (SIL). This standard is both detailed and modified by the ISO standard 26262 especially for automotive applications, and fundamental design principles tend to replace strict error rates (see ▶ Chap. 6, “Functional Safety of Driver Assistance Systems and ISO 26262”).

7.2.3 Test Case “Justified Intervention”: Vehicle-in-the-Loop

A few requirements for testing the test case “justified intervention” are given here:

- An automatic emergency brake will be triggered.
- There will be a crash.

- The driver and the vehicle must not be exposed to danger.
- The situations should look realistically to the driver.
- The test should be as reproducible as possible.

Simple test setups or examinations in the driving simulator do not fulfill all criteria. In the driving simulator, the threat to the driver may not be as realistic as necessary; the dynamics of the vehicle are limited in the simulator. If the real vehicle crashes into foam cubes, jibs, and small mobile vehicles, again the driver will not panic at all. The most advanced setup is reported by Hurich et al. (2009) in which FVCX-systems are challenged by real automated cars. Of course, this very expensive setup does not include intentional crashes.

All the requirements mentioned above are met by a new testing method for driver assistance systems called Vehicle-in-the-Loop (VIL, Bock et al. (2007), ▶ Chap. 10, "Vehicle in the Loop"). The basic idea is to simulate only the other traffic participants, and real hardware will be used for all other elements. The (real) human driver drives a real vehicle at a real testing site. The real environment will be extended by see-through glasses, which simulate the other traffic participants for the driver. Experiments show that test drivers react in a realistic way, although they could distinguish real from simulated objects upon request.

7.2.4 Error Probability for "Unjustified Interventions": Trojan Horses

It is as challenging as the test of the justified intervention described above to ensure that the error probability rate is very little – as an example no more than 10^{-8} errors per hour (▶ Chap. 6, "Functional Safety of Driver Assistance Systems and ISO 26262"; ISO 26262 2011). Assuming that a vehicle only drives 30 km per hour in average, this means that 3 billion test kilometers are to be driven without any false alarm. This vast mileage cannot be driven in standard fleet tests due to financial reasons. Again, alternative testing methods are necessary.

In Winner (2002), the implementation of Trojan horses in vehicles of the customers is proposed. The idea is that the customers would purchase a comfort function like ACC realized with the same sensor configuration and the same perception software, for example, a functional extend of ACC Stop&Go. In addition, the software would contain all functions of an automatic emergency brake, but the FVCM-module would not have access to the braking system. Instead of intervening, the FVCM-system would write an entry in the permanent memory of the ECU. If entries are detected during a later service, they either result from an accident which should be known or they would have been caused by a false alarm. In principle, all information would be available to determine the wanted error probability rate. The authors are not aware of any active discussions among the car manufacturers whether the process is appropriate for future testing. However, it cannot be ruled out that car manufacturers or their system partners use this method already without communicating.

8 Conclusion

Front vehicle collisions are responsible for a major percentage of heavy traffic accidents. Thus, suitable warning and intervening systems increase the traffic safety. Different system extends were summarized by the term FVCX-systems. They can be distinguished by their influence on the driver–vehicle–environment system. Basically, they can be classified as conditioning, warning, mitigating, and avoiding systems.

The specification of already available systems in the market can only be understood when the capability of machine-based perception is taken into account. Progress in this topic enables for warning and avoidance systems. Current system extends have a limit perceptional capability compared to the situation-aware driver, which has to be regarded when designing FVCX-systems.

A systematic design approach is recommended for the development of FVCX-systems. The motivation for FVCX-systems always has to be derived from traffic accident research. Already in early conceptional phases, aspects of functional safety, legal aspects, system ergonomics, and marketing should be regarded. Only with a consistent function definition that further development is reasonable. A testing and evaluation concept should be part of these early stages as well.

References

Bainbridge L (1983) Ironies of automation. In: Automatica 19, issue 6

Benmimoun M, Pütz A, Aust M, Faber F, Sánchez D, Metz B, Saint Pierre G, Geißler T, Guidotti L, Malta L (2012) euroFOT SP6 D6.1 final evaluation results

Bishop R (2005) Intelligent vehicles technology and trends. Artech House, Norwood, MA

Bloch A (2007) Tech No. www.auto-motor-und-sport.de, 16 2007

Bock T (2009) Bewertung von Fahrerassistenzsystemen mittels der Vehicle in the Loop-Simulation (Evaluation of driver assistance systems via Vehicle in the Loop-Simulations). In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme, 1st edn. Vieweg und Teubner, Wiesbaden, p 76

Bock T, Maurer M, Färber B (2007) Vehicle in the Loop (VIL) – a new simulator set-up for testing advanced driving assistance systems. In: Driving simulation conference North America, Iowa

Breu A, Holzmann M, Maurer M, Hilgers A (2007) Prozess zur Komplexitätsbeherrschung bei der Entwicklung eines Stillstandsmanagements für ein hochvernetztes Fahrerassistenzsystem (Process of complexity management at the development of a vehicle standstill management in the context of a highly coupled driver assistance system). In: Tagung stillstandsmanagement, Haus der Technik, Essen

Buld S, Tietze H, Krüger HP (2005) Auswirkungen von Teilautomation auf das Fahren (Effects of partly-automated systems towards driving tasks). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung. Springer, Berlin Heidelberg

Donner E, Winkle T, Walz R, Schwarz J (2007) RESPONSE 3 – code of practice for the design and evaluation of ADAS. In: Technischer Kongress des VDA, Sindelfingen

Ehrlenspiel K (2003) Integrierte Produktentwicklung (Integrated product development). Hanser, München

eSafety (2010) eSafetySupport. www.esafetysupport.org

Eskandarian A (ed) (2012) Handbook of intelligent vehicles. Springer, London

Färber B (1986) Abstandswahrnehmung und Bremsverhalten von Kraftfahrern im fließenden Verkehr (Perception of distance and braking behavior of car drivers in moving traffic). Zeitschrift für Verkehrssicherheit (32):9–13

Färber B, Maurer M (2005) Nutzer- und Nutzenparameter von Collision Warning und Collision Mitigation Systemen (User- and usage-parameters of Collision Warning and Collision Mitigation systems). In: Maurer M, Stiller C (eds) Workshop Fahrerassistenzsysteme Walting

Gelau C, Gasser TM, Seeck A (2009) Fahrerassistenz und Verkehrssicherheit (Driver assistance and traffic safety). In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme, 1st edn. Vieweg and Teubner, Wiesbaden, p 26

Gibson JJ (1950) The perception of the visual world. Houghton Mifflin, Cambridge, MA

Glaser H (2006) Fahrwerk und Fahrerassistenz – eine ideale Kombination? (Chassis and driver assistance – a perfect combination?). In: 7. Symposium zum Thema Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel. AAET, Braunschweig

Hurich W, Luther J, Schöner HP (2009) Koordiniertes Automatisiertes Fahren zum Entwickeln, Prüfen und Absichern von Assistenzsystemen (Coordinated automated driving for development, test and validation of driver assistance systems). In: 10. Symposium zum Thema Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel. AAET, Braunschweig

ISO 15623 (2011) International Organization for Standardization (ISO): intelligent transport systems – forward vehicle collision warning systems – performance requirements and test procedures, Geneva

ISO 22839 (2011) International Organization for Standardization (ISO): ISO 22839:2013 intelligent transport systems – forward vehicle collision mitigation systems – operation, performance, and verification requirements, Geneva

ISO 26262 (2011) International Organization for Standardization (ISO): ISO 26262:2011 road vehicles – functional safety, Geneva

Kiesewetter W, Klinkner W, Reichelt W, Steiner M (1997) Der neue Brake Assist von Mercedes Benz – aktive Fahrerunterstützung in Notsituationen (The new Brake assist of Mercedes Benz – Active driver assistance in emergency situations). In: ATZ Automobiltechnische Zeitschrift 99, issue 6

Kohoutek P, Dietz J, Burggraf B (2007) Entwicklungsziele und Konzeptauslegung des neuen Audi A4 (Development objectives and concept design of the new Audi A4). In: ATZ/MTZ extra – Der neue Audi A4, Vieweg

Kompass K, Huber W (2009) Integrale Sicherheit – effektive Wertsteigerung in der Fahrzeugsicherheit (Integral safety – effective increase of value in vehicle safety). In: Fahrzeugsicherheit und Elektronik, Umwelt und Energie, 11. Technischer Kongress des VDA

Kopf M (1999) RESPONSE Checklist for theoretical assessment of advanced driver assistance systems: methods, results and assessment of applicability

Kopf M (2005) Was nützt es dem Fahrer, wenn Fahrerinformationssysteme und – assistenzsysteme etwas über ihn wissen (Evaluating the benefit for drivers from driver information and assistance systems knowing its driver). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung. Springer, Berlin Heidelberg

Kopischke S (2000) Entwicklung einer Notbremsfunktion mit Rapid Prototyping Methoden (Development of an emergency breaking function with rapid prototyping methods), PhD thesis, Technische Universität Braunschweig

Kötz J (2000): Personal communication

Kraiss KF (1998) Benutzergerechte Automatisierung – Grundlagen und Realisierungskonzepte (User-friendly automation – principles and development concepts). In: at – Automatisierungstechnik 46, issue 10, Oldenbourg, p 457–467

Lucas B, Held R, Duba GP, Maurer M, Klar M, Freundt D (2008) Frontsensorsystem mit Doppel Long Range Radar (Frontal sensor system with double Long Range Radar). In: Maurer M, Stiller C (eds) 5. Workshop Fahrerassistenzsysteme Walting

Mäkinen T, Irion J, Miglietta M, Tango F, Broggi A, Bertozzi M, Appenrodt N, Hackbarth T, Nilsson J, Sjogren A, Sohnke T, Kibbel J (2007) APALACI final report 50.10b

Maurer M (2000) Flexible Automatisierung von Straßenfahrzeugen mit Rechnersehen (Flexible automation of road vehicles using computer vision). Fortschritt-Berichte VDI, Reihe 12: Verkehrstechnik/Fahrzeugtechnik. Bd. 443

Maurer M (2009) Entwurf und Test von Fahrerassistenzsystemen (Development and Test of driver assistance systems). In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme, 1st edn. Vieweg und Teubner, Wiesbaden

Maurer M, Wörsdörfer KF (2002) Unfallschwereminderung durch Fahrerassistenzsysteme mit maschineller Wahrnehmung – Potentiale und Risiken (Collision mitigation by driver assistance systems with machine-based perception – Potentials and Risks), Unterlagen zum Seminar Fahrerassistenzsysteme und aktive Sicherheit

Naab K, Reichart G (1998) Grundlagen der Fahrerassistenz und Anforderungen aus Nutzersicht (Principles of driver assistance systems and requirements from users' perspective). Seminar Fahrerassistenzsysteme und aktive Sicherheit

Neukum A, Lübbeke T, Krüger HP, Mayser C, Steinle J (2008) ACC Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen (ACC Stop&Go: Driver behavior at functional system limits). In: Maurer M, Stiller C (eds) Workshop Fahrerassistenzsysteme Walting

Reif K (2006) Automobilelektronik – Eine Einführung für Ingenieure (Automotive electronics – an introduction for engineers). ATZ/MTZ-Fachbuch, Vieweg, Wiesbaden

Schäuffele J, Zurawka T (2006) Automotive software engineering 3rd edn. ATZ/MTZ-Fachbuch, Vieweg, Wiesbaden

Schmid V, Bernzen W, Schmitt J, Reutter D (2005) Eine neue Dimension der Aktiven und Passiven Sicherheit mit PRE-SAFE und Bremsassistent BAS PLUS in der neuen Mercedes-Benz S-Klasse (A new dimension of active and passive safety with PRE-SAFE and braking assist BAS PLUS in the new Mercedes-Benz S-Class). In: 12. Internationaler Kongress Elektronik im Kraftfahrzeug, VDI-Berichte Nr. 1907

Sugimoto Y, Sauer C (2005) Effectiveness estimation method for advanced driver assistance system and its application to collision mitigation brake system. In: Proceedings of the 19th International Technical Conference on the Enhanced Safety of Vehicles, pp 5–148

Weinberger M (2001) Der Einfluss von adaptive cruise control Systemen auf das Fahrverhalten (Influence of adaptive cruise control systems towards driving behavior). PhD thesis, Technische Universität München

Wiesbeck W (2006) Radar system engineering. Universität Karlsruhe, Lecture notes, 13rd edn. http://www2.ihe.uni-karlsruhe.de/lehre/grt/RSE_LectureScript_WS0607.pdf

Winner H (2002) Einrichtung zum Bereitstellen von Signalen in einem Kraftfahrzeug (System for the provision of signals in a vehicle). Patent DE 101 02 771 A1, Deutsches Patent- und Markenamt, date of filing 2001/23/01, date of publication: 25 Jul 2002

Yerkes R, Dodson J (1908) The relation of strength of stimulus to rapidity of habit-formation. J Comp Neurol Psychol 18(5):459–482

Zanten van A, Kost F (2009) Bremsenbasierte Assistenzfunktionen (Brake-based driver assistance functions). In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme, 1st edn. Vieweg und Teubner, Wiesbaden, p 392

Lateral Guidance Assistance

48

Arne Bartels, Michael Rohlfs, Sebastian Hamel, Falko Saust, and Lars Kristian Klauske

Contents

A. Bartels (✉)
Fahrerassistenz und integrierte Sicherheit, Volkswagen AG – Konzernforschung, Wolfsburg, Germany
e-mail: arne.bartels@volkswagen.de

M. Rohlfs • S. Hamel • F. Saust
Volkswagen AG, Wolfsburg, Germany
e-mail: michael.rohlfs@volkswagen.de; sebastian.hamel@volkswagen.de; falko.saust@volkswagen.de

L.K. Klauske
Carmeq GmbH, Berlin, Germany
e-mail: lars.kristian.klauske@carmeq.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_49

Abstract

Unintentional lane departures are the root cause for more than 1/3 of all accidents with severely injured occupants on German roads. Therefore, in recent years, Lane Departure Warning (LDW) and Lane Keeping Assistance (LKA) systems have been introduced on the market, which support the driver in lane keeping. LDW informs the driver by means of tactile, visual, and/or audible feedback if the vehicle is about to leave the lane unintentionally. LKA supports the driver by intervening in lateral vehicle control to help keep to the lane.

This chapter provides a clear classification of those lateral guidance assistance systems. General requirements are provided, referring to regulations and standards as ISO 17361, ISO DIS 11270, UN ECE R-79, and Euro NCAP. Key technologies and system components of current implementations are described, e.g., environment sensors, warning algorithms, HMI for driver information, and lateral controllers. Exemplary implementations of four OEMs are described in detail. A system evaluation from customer perspective is given referring to surveys published by ADAC and Auto-Bild. Finally, an outlook is given how to further improve the achieved performance for future systems.

1 Motivation

Steering, and thus keeping the vehicle in the current lane, is a primary driving task which the driver must continuously perform during the entire drive. Unfortunately, drivers do not always manage to perform this task without errors. This is evidenced by the accident statistics in Fig. 1. The figure shows the percentage of occupants sustaining severe injuries in car accidents (MAIS 2+) broken down by accident and road type. What becomes apparent is that in more than 1/3 of all accidents with severely injured occupants (37.9 %) on German roads, the cause was the vehicle unintentionally departing from the lane. The majority of these accidents occur in extraurban areas, e.g., on motorways, A roads, and rural roads (29.4 %).

The conclusion that can be drawn from these accident statistics is that the driver requires support in lateral control of the vehicle. A system which informs the driver before the vehicle unintentionally departs from the current lane or is trying to prevent this by an active intervention in lateral vehicle guidance can be expected to have a positive effect on accident statistics, in particular in extraurban areas, e.g., on motorways, A roads, and rural roads.

Additionally, the act of continuously stabilizing the vehicle in the center of the lane can be perceived as exhausting by the driver, especially when making long-distance travels. An assistance system which offloads a part of this lateral control task could relieve the driver and enhance traveling comfort.

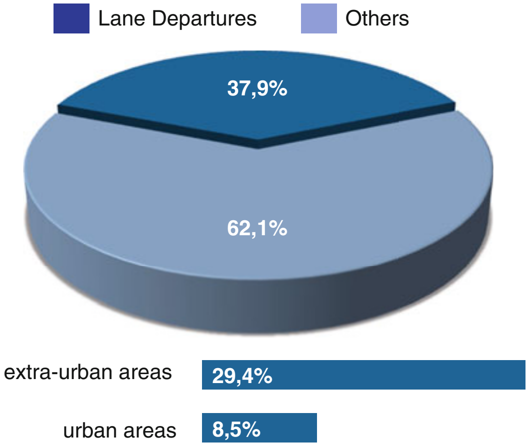


Fig. 1 Accidents with severely injured passenger car occupants (MAIS2+): percentage of all occupants by accident and road type (GIDAS 2012)

2 Requirements

Lateral guidance assistance systems are designed to help prevent vehicles leaving a lane unintentionally by

- Informing the driver in a timely manner
- Steering the departing vehicle back into the lane if possible
- Actively supporting the driver in keeping the vehicle in the center of the lane

To inform the driver, the system must be capable of determining the position of the vehicle relative to the lane boundary, which is typically a line. Driver information systems must at least be capable of detecting this line (Single Line Detection).

Systems that steer the vehicle back into the lane, or keep the vehicle in the center of the lane, must be capable of detecting the vehicle's position relative to the center of the lane, its future direction of motion, and the lane's progression ahead of the vehicle. In addition to appropriate vehicle sensors, this requires environment sensors with a high degree of forward vision and accuracy. If lane detection is accomplished by reference to the left and right lane markers, consequently the system must be capable of detecting both lines (Dual Line Detection). Ideally, line detection should work on almost all roads in all countries, even under adverse environmental conditions.

On the other hand, unnecessary driver information or interventions should be avoided with respect to lateral position and orientation of the vehicle. In addition, information/intervention should be avoided for intentional lane changes, i.e., during overtaking or if the driver intentionally "cuts corners."

Driver information on unintentional lane departures should be clearly perceptible but not “annoying.” For this purpose, a careful balance of visual, audible, and tactile driver information must be sought.

Active system intervention in lateral control is required to return a vehicle to a lane or to keep a vehicle within a lane. This intervention must be designed so that the driver can always override the system.

The lateral guidance assistance system should emulate natural steering characteristics, i.e., it should avoid continuous, high-frequency steering movements when keeping the vehicle within a lane.

The driver must be informed in a transparent and clear way as to whether the system is switched on and active. In addition, it must be easy for the driver to switch the system on and off.

The system must be capable of assisting the driver in keeping to the center of a lane without completely offloading this task. The responsibility for lateral vehicle control remains with the driver, and he or she must not ignore this task. This is essential especially for systems with active intervention in lateral vehicle control. The driver must thus always be involved in physically steering the vehicle.

3 Classification

From an engineering point of view, lateral guidance assistance systems can be classified into two groups: (a) Lane Departure Warning (LDW) systems which provide driver information and (b) Lane Keeping Assistance (LKA) systems which intervene with lateral vehicle control (Table 1). LDW systems inform the driver by means of tactile, visual, and/or audible feedback in imminent cases of the vehicle unintentionally leaving the lane. LKA systems support the driver by actively intervening in lateral vehicle control in order to keep to a lane. This can occur in one of two ways: Type I initially attempts to prevent the vehicle from leaving the lane in the best possible way by actively intervening with lateral vehicle control. If crossing of lane markers is not prevented (despite this intervention), the driver is informed as per an LDW. Type II supports the driver by keeping the vehicle at the

Table 1 Classification of lateral guidance assistance systems

LDW	Lane Departure Warning	
	Informs the driver by means of tactile, visual and/or audible feedback if the vehicle is about to unintentionally leave the lane.	
LKA	Lane Keeping Assistance	
	Supports the driver by intervening in lateral vehicle control to keep to the lane.	
	Type I	Initially makes the best possible attempt to prevent the vehicle leaving the lane through correctional intervention in the vehicle’s lateral control; also informs the driver if necessary (see LDW), safety-targeted feature.
	Type II	Supports the driver by keeping the vehicle at the center of the lane by actively intervening in the vehicle’s lateral control; also informs the driver if necessary (see LDW), convenience-targeted feature.

center of the lane by actively intervening with lateral vehicle control; it also warns the driver if necessary (see LDW). This differentiation allows the different types to be classified by functional characteristics. While type I contributes to vehicle safety, type II additionally takes convenience aspects into account.

A lane-centered lateral guidance system without driver involvement, e.g., as required for automated driving functions, can be designed from a purely technical point of view around the LKA system components described below. However, this is not covered by this chapter on lateral guidance assistance; nor are production systems with combined longitudinal and lateral control (for this, see ▶ Chap. 51, “Traffic Jam Assistance and Automation”).

4 Regulations, Standards, and Tests

The following section lists some examples of important regulations and standards that relate directly to lateral guidance assistance systems. The listed documents define requirements that must be taken into account for the design of production systems. In addition to threshold values, also requirements on the system’s override capability and HMI (Human-Machine Interface) concepts are specified.

ISO standard 17361 “Intelligent transport systems – Lane departure warning systems Performance requirements and test procedures” (ISO 2007) and ISO DIS 11270 “Intelligent transport systems – Lane keeping assistance systems (LKAS) – Performance requirements and test procedures” (ISO 2013) specify minimum functional requirements, basic HMI elements, and methods of testing for LDW and LKA systems for passenger cars, commercial vehicles, and buses on motorways and similar roads. However, ISO DIS 11270 does not differentiate between type I and type II systems. Changes to the requirements are still possible as the standard was still in the draft phase when this document was authored.

The regulation UN ECE R-79 describes “Uniform provisions concerning the approval of vehicles with regard to steering equipment” and groups the lane keeping assistance systems into corrective steering functions and automatically commanded steering functions (ECE 2005). For corrective steering functions, it demands that the driver shall be able to override the function as well as a limited duration of the intervention. Automatically commanded steering functions with continuous control action are currently permitted only up to a maximum speed of 12 km/h in the current version of the UN ECE R-79. In addition, advanced driver assistance steering systems shall be designed such that the driver can, at all times, choose to override the assistance function by deliberate action. Furthermore, according to UN ECE-R79, the driver remains in primary control of the vehicle.

The Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) in Japan has created framework requirements for LKA systems in a “Technical Guideline” which defines threshold values for the lateral acceleration which may not be exceeded during a steering intervention as well as requirements for the HMI. Here, they differentiate between cornering (max. 2 m/s^2) and driving on straight routes (max. 0.5 m/s^2) (MLIT 2013).

Comparable requirements for lateral acceleration limits are also specified in ISO DIS 11270. The maximum lateral acceleration and jerk which is allowed to be induced by a lane keeping action is limited to 3 m/s^2 and 5 m/s^3, respectively. In addition, the driver shall be provided with means to override the lane keeping action at any time. These means of suppression are specified driver activities such as operating the turn signal or steering intervention by the driver.

EU regulation 661/2009 requires the installation of LDW systems for commercial vehicles of classes M2, M3, N2, and N3, as of 01/11/2013 for all new vehicle types and as of 01/11/2015 for all new vehicles. In a separate regulation 351/2012, the EU states type approval regulations for LDW systems. For additional information regarding lateral guidance assistance systems for commercial vehicles, see ► Chap. 52, "Road Assistance for Commercial Vehicles".

In the USA, all new vehicles, including their LDW systems, are evaluated by the New Car Assessment Program (NCAP). All relevant test procedures are fully described in the document "Lane Departure Warning System confirmation test and Lane Keeping Support performance documentation" from the National Highway Traffic Safety Administration (NHTSA). LDW systems for the European market are evaluated as per Euro NCAP as of 2014. The NHTSA test procedure provides the basis for this; however, it was revised for the European market.

Figure 2 shows the NHTSA test procedure for LDW systems as an example, which establishes requirements for production systems. At a defined test speed of 72 km/h, and limited lateral speed ($0.1 \text{ m/s} \le v_{\text{latr}} \le 0.6 \text{ m/s}$) and yaw rate ($\dot{\psi} \le 1°/\text{s}$), the key requirement for the LDW system is that driver information for crossing a lane marker is output in a predefined zone of 0.75 m in front of the lane marker and 0.3 m behind the lane marker. Reproducibility of the test in terms of the defined maximum lateral speeds and the yaw rates of the vehicle is achieved by driving the vehicle on a defined track delimited by start and end traffic cones. The driver information can be provided via audible, tactile, or visual feedback.

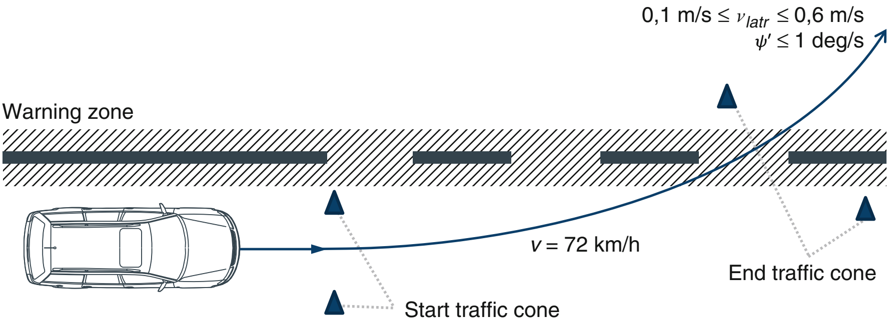


Fig. 2 Test requirements based on NHTSA

5 System Components

Figure 3 shows the block diagram of a lane departure warning (LDW) and a lane keeping assistant (LKA) with the respective system components. The white blocks are required by both the LDW and the LKA system. The grey blocks show the additional components required by an LKA system in comparison with an LDW system.

Environment sensors (e.g., cameras) generate measurement data (e.g., images) from which specific environmental data (e.g., position of lane markings) are extracted by signal processing. It is possible to include other environmental characteristics in signal processing by the use of sensor data fusion. The warning algorithm determines the driver information requirement within the function module (Fig. 4). Based on the vehicle and system status, a state machine determines whether the driver needs to be informed. The HMI includes the driver information

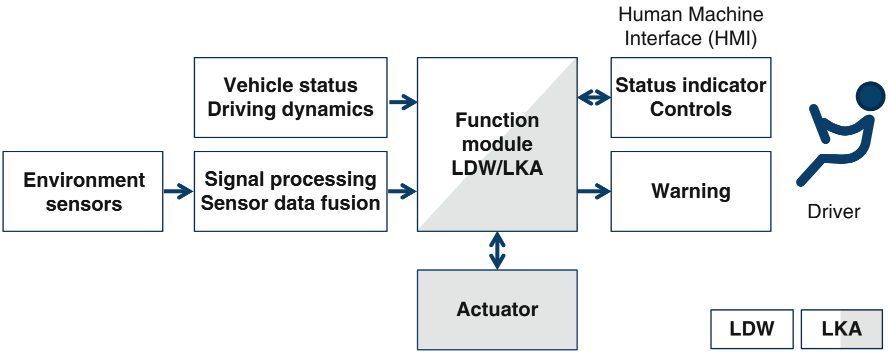


Fig. 3 System components for a lateral guidance assistance system

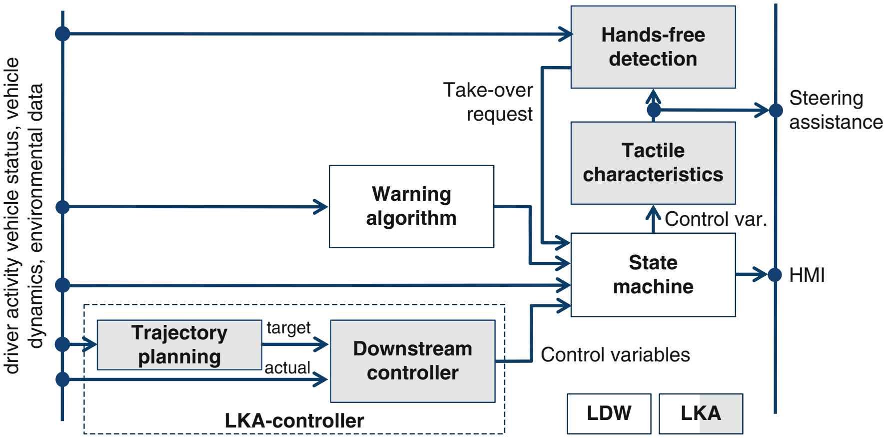


Fig. 4 LDW/LKA function module

along with the system status output. Control elements allow the driver to switch the system on and off and also to configure the system (e.g., adjusting the time for driver information).

LKA systems additionally require a lateral controller. The lateral controller computes the control values (e.g., steering torque) and sends them to an actuator (e.g., steering) for suitable lateral control intervention. Ideally, the driver will be able to perceive this as tactile feedback. The hands-free detection shall prevent the driver from driving without having his hands on the steering wheel.

The following describes the individual components required for a lateral guidance assistance system in detail.

5.1 Environment Sensors

Not only cameras but also infrared diodes and laser scanners are, in principle, suitable for detecting lane markers. However, monocular cameras are more typically used. They are located behind the windscreen at the same height as the interior mirror, where they are invisible to the driver (Fig. 7c) with a forward-looking direction of view.

Camera-based systems stand out due to a large field of view and high resolution. For example, given a ±21° field of view and ranges of up to 80 m, angular accuracy of ±2° can be achieved (TRW 2013). Lane markers can be detected well ahead with a comparably high degree of accuracy, which is beneficial for active lateral guidance in lane keeping systems. Furthermore, camera-based systems offer the potential for multipurpose use of the sensor, e.g., for road sign recognition or high-beam assistant systems.

If color cameras are used, the system remains operational in construction zones marked by yellow and white lane markers since the ambiguity of these lines can be assessed. Where cameras with a high dynamic range are used, the system has proven to be robust against extreme changes in lighting conditions, e.g., at tunnel entrances and exits or when driving along tree-lined roads in summer.

The first vehicle manufacturers have started using 3-D technologies, enabling spatial perception, which not only support more precise detection of objects and pedestrians but also facilitate the classification of raised structures, such as guardrails and curbs (Hegemann et al. 2013). Besides stereo vision (see ▶ Chaps. 20, “Fundamentals of Machine Vision,” and ▶ 21, “Stereovision for ADAS”) three-dimensional object detection can also be achieved with a monocamera solution using “Structure from Motion” (Derendarz et al. 2010).

Infrared diode-based lane marker detection (Fig. 8a) was unable to establish itself on the market, presumably because of its lack of adequate sensor range, multipurpose use, and accuracy. Poor sensor range, in particular, makes infrared-based diodes unsuitable for LKA systems, as firstly lane markers are only detected shortly before they are crossed and secondly ambiguities in construction zones are resolved either very late or not at all. Both cause delays in driver information output. Furthermore, detection of “Botts’ dots,” which are mainly found in the USA, cannot be guaranteed. Compared to cameras, infrared diodes – which are installed near the ground – are exposed to heavy soiling. However, due to their vertical direction of view, they are not sensitive to oncoming lights or rain, as opposed to cameras.

Laser scanners for detecting lane markers are currently only used in research projects (Montemerlo et al. 2008; Homm et al. 2011). Localization-based approaches that rely on highly accurate digital maps are not currently available in production, nor are infrastructure-based systems that rely on magnetic nails or guide cables.

For detailed information on environment sensors, please refer to Part IV of this handbook. “Sensors for DAS.”

5.2 Signal Processing

The image processing algorithms for detecting the lane markers are authoritative for the quality of camera-based LDW and LKA systems. The predominant task is that of detecting lane markers. An example of an algorithm for this is presented in ▶ Chap. 20, “Fundamentals of Machine Vision”. The general requirements for detecting lane markers are

- Available for the widest possible range of infrastructural conditions
- Robust when faced with adverse environmental conditions

Road infrastructure diversity is a challenge. White (Europe) and yellow (USA, Canada) lane markers need to be detected on dark asphalt, or bright concrete, as well as marker nails in construction zones or “Botts’ dots” in the USA. The line lengths, spacing, and widths vary greatly throughout the world (see ISO 17361 Annex A). Well-maintained marker lines on motorways should be detected as well as worn or weather-beaten lines on minor roads. Line structures formed by bitumen joints, tarmac seams, or guard rails should be identified as irrelevant, as should skid marks or tracks in snow.

Adverse environmental conditions that can impair the visibility of marker lines include line coverage through soiling, foliage or snow, or overgrowth in the form of grass or shrubs. Lane markers on wet roads in the dark with oncoming headlights, or during daylight with the sun low on the horizon, are difficult to detect even for the driver. The same is true in case of heavy rain, spray, or fog.

Ideally, the vehicle’s lane is detected even if the marker lines are briefly or permanently invisible due to such adverse environmental conditions. Short drop-outs in detecting the line markers can be bridged using suitable algorithms such as Kalman filters or particle filters. If marker lines are not detected over a longer period, then other environmental characteristics can be used for lane detection (see Sect. 9).

5.3 LDW/LKA Function Module

An LDW system with driver information can already be implemented using this environmental data in combination with the vehicle status, vehicle dynamics, and driver activity information already available in today's vehicles. On top of this, a type I or type II LKA system can be implemented if suitable actuators are additionally available in the vehicle such as an electromechanical steering system (electric power-assisted steering (EPS)). The key element in this is the function module.

Figure 4 shows a structure example for this type of software component. It features a central state machine and warning algorithm, along with the components necessary for an LKA system as hands-free detection, tactile characteristics computations, trajectory planning, and controller for lateral vehicle control. These components are described in the following. In current vehicle architectures, the hardware component responsible for the warning algorithm and LKA lateral vehicle control is typically the environment sensor control unit (e.g., the evaluation module in the camera).

5.3.1 State Machine

The state machine is a major component of LDW and LKA systems (Fig. 4). It checks whether all the boundary conditions for driver information and/or intervention in lateral vehicle control are met. The respective system must be switched on (On/Off Button) and operational (Self-Diagnostics); the sensors must not be inoperable or soiled; the turn signal must not be set, and the vehicle speed must lie within the activation limits (Vehicle Status). To help prevent continuous activation/deactivation, a hysteresis may be applied to the speed threshold. For LKA systems, it is additionally checked if the driver is actively steering, and/or the driver's hands are on the steering wheel, or is letting the system steer the vehicle (Hands-Free Detection). A link between the LDW/LKA systems and other driver assistance systems can be implemented in the state machine. In combination with lane change assistance systems (see ► Chap. 49, "Lane Change Assistance"), the LDW or LKA system can output driver information, for example, if a neighboring lane is occupied and the driver has indicated the intent to change lanes by activating the turn signal. LDW and LKA systems can thus help to prevent accidents with vehicles in the neighboring lane during lane change maneuvers.

5.3.2 Warning Algorithm

The "Distance to Line Crossing" (DLC) d_{LC} is the most basic criterion for driver information in case of imminent lane departure. It denotes the lateral distance between a certain part of the vehicle and the lane marker. A warning zone is set up by defining a minimum and maximum DLC. It starts just before the lane marker and ends just after it (Fig. 5, left). The driver is informed if the vehicle enters this warning zone. The driver information stops when the vehicle leaves the warning zone. The DLC can use simple sensors with small sensor range, such as infrared diodes. The use of DLC to identify critical situations may have adverse effects:

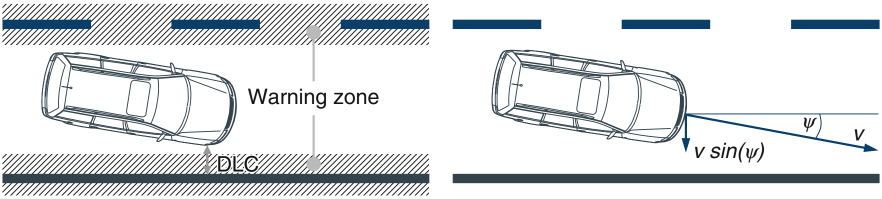


Fig. 5 *Left*: DLC and warning zone, *right*: lateral vehicle speed for determining the TLC

For example, if a vehicle is driving very close and parallel to a lane marker, driver information will still be output even though the vehicle is not about to leave the lane.

"Time to Line Crossing"(TLC) t_{LC} is better suited as a criterion for lane departure warning since it is predictive. By this means, it helps to prevent unnecessary driver information, as described in the DLC section above. The TLC denotes the time span after which a vehicle will probably cross a lane marker based on the vehicle position and motion. It is calculated as $(t_{LC} = d_{LC}/v \cdot \sin(\Psi))$, where $v \cdot \sin(\Psi)$ is the approach speed toward the lane boundary, along with the vehicle's longitudinal speed v and the vehicle's orientation relative to the lane marker Ψ (Fig. 5 right). The curvature of the vehicle trajectory and the road should also be considered for a generically applicable approach to computing the TLC. Computations for this are found in (Winsum et al. 2000; Mammar et al. 2006). In the simplest case, the driver is informed as soon as the TLC falls below a predefined threshold. Highly accurate sensors with a large sensor range, such as cameras, are particularly well suited to determining the TLC.

It is desirable for the driver to be able to adjust this threshold; depending on the style of driving and the route, it may be useful to output driver information just before, while, or even after crossing the lane marker.

If the driver intentionally "cuts corners" on a winding road, or is driving along a narrow road close to the lane boundary, or starting an overtaking maneuver, but without using the turn signal, then driver information can be considered unnecessary or even irritating. Undesirable driver information can be prevented with the aid of driver intention recognition (see ▶ Chap. 37, "Driver Condition Detection"). Additionally, evaluating the environmental and context information, such as the vehicle acceleration, accelerator pedal position, steering wheel angle, yaw angle, lane curvature, and types of marker lines on the left and right can help to detect intentional "corner cutting" or overtaking in many cases, and therefore avoid outputting unnecessary driver information. Additionally, the time for outputting driver information can be delayed for driving on narrow roads.

If driver status information is available as per ▶ Chap. 37, "Driver Condition Detection", it seems reasonable to configure the time at which the driver is informed based on the driver's activity: for example, to inform earlier if the driver is distracted, tired, or driving inattentively, or to delay providing driver information when the driver is active and thus enhancing system acceptance while active driving.

5.3.3 Lateral Control

Various approaches are used for lateral control in type I and type II LKA systems (for example, see Gayko 2012; Koelbl 2011; Mann 2008). In the case of the exemplary lateral controller described in the following section, the target and actual behavior of the vehicle in lateral direction are expressed as accelerations. This approach already accounts for the vehicle speed in planning the trajectory, thus facilitating speed-independent configuration of the actual controller. In addition, lateral acceleration is key to the vehicle response perceived by the driver.

The lateral controller in our example is broken down into two modules: a preliminary trajectory planning module computes the desirable behavior of the vehicle as a lateral acceleration on the basis of the environmental data. Then, a downstream controller calculates the required actuator control variables by reference to the vehicle data (actual lateral acceleration) and a vehicle behavior model.

Trajectory planning occurs primarily on the basis of the lane marker data, which can be provided as approximate clothoids by the lane marker detection system. A point $y(x)$ on a marker line at a distance of $x = v \cdot \tau$ (vehicle coordinates) is given by

$$y(x) = y_0 + x \cdot \sin \Psi + \tfrac{1}{2} \cdot x^2 \cdot \kappa + \tfrac{1}{6} \cdot x^3 \cdot M$$

with a lateral deviation of y_0, yaw angle of Ψ, actual curvature of κ, and change in curvature of M; the yaw angle portion can be approximated by $\sin \Psi = \Psi$ having regard to the measuring uncertainty of yaw angles pertinent to LKA systems.

Representing the lane markers as approximate clothoids internally is an obvious choice, as the course of roadways in many countries is actually approximated by the use of clothoids.

Furthermore, objects and roadside structures identified by the environment sensors can be taken into consideration – as a replacement for missing lane markers, for example, or to limit the drivable space. Just like lane markers, the drivable space is delimited by clothoids on the left and right.

To implement an LKA system, the trajectory planning module computes the target lateral acceleration ahead of the vehicle in a defined period of time τ taking the width of the vehicle into account. If the driver may configure when driver information is output, then the target value can be optionally configured by the driver by reference to a safety distance.

5.3.4 Haptic Characteristics

Although the actuator is responsible for actually implementing the actuating value computed by the LKA lateral controller, the haptic characteristics and controllability of an LKA system that intervenes with a vehicle's lateral movement are also essential components of this system. They are thus considered to be components of the LKA function module in the example described here, although they can be implemented in other components within the architecture of current LKA systems (e.g., as a module in the EPS or ESC control unit for LKA systems with course-correcting brake intervention).

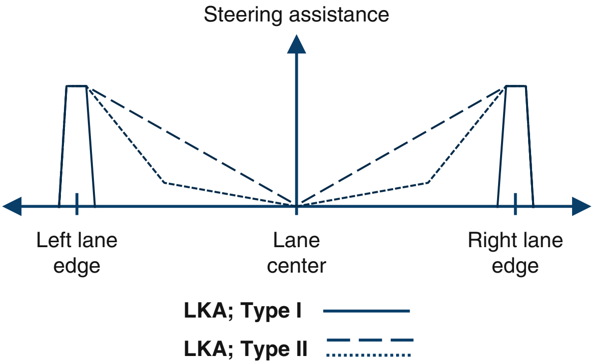
Due to the immediate and regular (type I LKA) or even permanent (type II LKA) interaction with the driver through the lateral control actuators, the haptic characteristics have an imminent impact on the driver's experience with the system. Because driver requirements and LKA lateral control, in particular for LKA type II systems, frequently compete (e.g., the control actively keeping to the center of the lane, while the driver wants to follow the edge of a lane), the haptic characteristics module determines when, and with what level of perceptibility, the actuators should implement the controller's requests, based on details of the vehicle state, driver activity, and environmental data.

Two authoritative scenarios exist for haptic characteristics in terms of immediate driver activity: for countersteering, the driver activity is in a direction opposite to the control request; the driver perceives the control request as counterforce acting on the steering wheel. The haptic characteristics module proportionally reduces the request to the actuator, according to the desired degree of steering assistance by the LKA system. Although this impacts the warning character and the control performance, the system appears to be more convenient and less irritating. In "combined steering," the driver action and the control request are aligned. The driver perceives the control request as an unusually intense vehicle reaction. Many drivers evaluate this behavior as unpleasant.

As the LKA's warning character and control performance impact an active driver's user experience during countersteering in particular, the haptic characteristics module shown as an example here additionally takes into account the vehicle's position in the lane. If the vehicle is at the edge of a lane, the haptic characteristics module induces stronger steering assistance when countersteering. If the driver continues to drive in the center of a lane, the module reduces the steering assistance. By this means, the control request is reduced to a greater extent for an active driver.

Figure 6 illustrates this relationship by means of two possible characteristic curves: for type I LKA systems, the focus is on the warning character at the edge of the lane; steering assistance thus remains strong here – the type I LKA system does not provide steering assistance at the center of the lane. Type II LKA systems also need to implement an unambiguous warning characteristic at the edge of the lane;

Fig. 6 Assisted steering depending on the lateral position for type I and type II LKA systems



steering assistance is at a high level here. When driving in the center of the lane, areas with reduced steering assistance can be set by system configurations, or the driver, to improve the user experience.

The controllability of assistance systems that actively intervene with lateral vehicle control can be considered from two points of view. On the one hand, functional safety as per ISO 26262 requires a risk assessment, based on a controllability evaluation in various driving situations, in order to determine the risk, taking exposure and severity into account; see ▶ Chap. 6, "Functional Safety of Driver Assistance Systems and ISO 26262". An evaluation of the controllability of the steering torque applied by the system is given in (Schmidt 2009). This reference particularly draws attention to the fact that beside the maximum amplitude of the steering torque, especially the steering torque gradient is relevant to driver controllability.

Requirements are also imposed by directives and regulations (see Sect. 4). LKA systems in particular are subject to regulations and standards that affect the implementation of controllability and specify permissible limits. As a result, the restrictions that apply to driver controllability of LKA systems must be implemented in the LDW/LKA function module, namely, in the interaction of the haptic characteristics module and the state machine.

5.3.5 Hands-Free Detection

To fulfill legal regulations and standards (see Sect. 4), the driver must continue to perform system monitoring tasks; because of this, today's LKA systems are not intended to support automated driving, whereby a hands-free detection is necessary.

The electromechanical power steering systems installed as standard equipment in many of today's vehicles (see ▶ Chap. 31, "Steering Actuator Systems") already provide the data required to evaluate driver activity via the integrated sensors of the steering system. Analyzing the steering activity helps to distinguish between the driver's steering wheel movements and exterior influences, caused by, for example, poor road surfaces. The different frequencies of the various excitation types can be leveraged to help do this. Because it becomes increasingly difficult to make a distinction as the steering activity decreases, for example, in the case of long straight runs, false positives can occur in hands-free detection. Beyond this, other methods of detecting driver activity exist, for example, using driver monitoring cameras, or capacitive or pressure sensors in the steering wheel.

If the system fails to detect sufficient steering activity, this can indicate that the driver no longer has his hands on the steering wheel. In this case, the driver needs to be prompted in a suitable way (audible, tactile, visual warning) to resume steering. If the driver does not follow this prompt, the system is switched off after an appropriate delay.

5.4 Driver Information

According to ISO standard 17361, "(. . .) an easily perceivable haptic and/or audible warning shall be provided. (. . .) If the haptic and/or audible warnings are not designed to indicate the direction, a visual cue may be used to supplement the warning." (ISO 2007)

The general requirements for driver information for LDW and LKA systems include

- *Explicit*, so that driver information is easily perceivable even by an inattentive driver
- *Intuitive*, so that that the type of driver information prompts the intended driver reaction
- *Exclusive*, so that the driver can react quickly without needing to think
- *Side selective*, so that the driver can immediately conclude the direction in which to steer
- *Only perceptible by the driver*, so that other occupants do not notice the driver information
- *Cost effective*, in that no additional components are required, if possible

Tactile driver information can be provided, for example, by a vibrating steering wheel, a vibrating seat, or a safety belt pretensioner. Steering wheel vibration can be designed into the steering wheel via vibration motors, or alternatively generated as a function of the electric power-assisted steering (EPS). Interventions in lateral control can also be used as tactile driver information, for example, by relying on the EPS steering actuator to produce steering torque characteristics similar to Fig. 6 or, alternatively, by using noticeable course-correcting brake intervention via the ESC system.

Audible driver information can be provided by "auditory icons", for example, special information sounds, or the sound of a rumble strip. These can be output, for example, via the radio/navigation system's stereo loudspeakers. Music and voice output should be muted for driver information. As an alternative, the buzzer or gong in the instrument cluster can be used. It is evident that audible driver information does not fulfill the requirement stated previously that only the driver should perceive the information.

Visual driver information should lie within the driver's primary field of view, for example, as an image or symbol in the instrument cluster or head-up display (HUD) (see Figs. 9b and 10b).

An evaluation of the different driver information options for LDW and LKA systems in line with the abovementioned criteria is given in Table 2.

Clearly, unequivocal perception of the driver information is possible with almost all versions. In some cases, seat vibration may not be perceivable due to thick clothing in winter. For visual-only driver information in principle, a distracted driver missing the symbol in the instrument cluster or head-up display cannot be excluded. For this reason, as per ISO 17361, visual driver information can complement, but never replace, tactile or audible driver information.

Intuitive driver information that prompts the intended driver reaction can be provided through lateral control intervention and steering wheel vibration, or also by using the sound of a rumble strip. A symbol or image in the HUD or the

Table 2 Evaluation of driver information for LDW and LKA systems

Driver information			Explicit	Intuitive	Exclusive	Side selective	Driver only
Type	Medium	Actuator					
Tactile	Lateral control intervention	EPS steering	+	+	+	+	0
		ESC	+	+	+	+	–
	Steering wheel vibration	EPS steering	+	+	+	–	+
		Vibrator	+	+	+	–	+
	Safety belt tug	Seatbelt tensioner	+	0	+[a] –[b]	–	+
	Seat vibration	Vibrator	0	0	+	+	+
Audible	"Rumble strip"	Loudspeakers	+	+	+	+[c] –[d]	–
	Special info sound		+	0	+	+[c] –[d]	–
	Gong, buzzer	Instrument cluster	+	0	–	–	–
Visual	Image, symbol	Instrument cluster	–	+	+	+	+
		Head-up display	–	+	+	+	+

Belt tensioner exclusively for LDW:
[a] Yes
[b] No
Stereo loudspeakers available:
[c] Yes
[d] No

instrument cluster is also suitable, in principle. An appropriate reaction is not necessarily guaranteed in case of information through the belt pretensioner, vibration information via the seat, audible information, and a buzzer or gong in the instrument cluster. These types of driver information do not necessarily indicate the need for steering intervention.

When providing tactile driver information, the sensory channel used here is exclusively available. The driver can easily associate the information with the LDW and/or LKA systems without much hesitation. This prompts a fast driver reaction. In the case of the seat belt pretensioner, this is only true if it is not used for other applications (then " + "). Special notification sounds, and the sound of a rumble strip, allow an unambiguous assignment, as do images and symbols in the instrument cluster and HUD, the prerequisite being that they are clearly perceptible, and easily understandable. By contrast, an unambiguous assignment of gongs or buzzers in the instrument cluster is not possible as they are used by many other applications.

Side selective driver information can only be provided by intervening with lateral control, seat vibration, and via a suitable symbol or image in the instrument cluster or HUD. Stereo loudspeakers are also suitable (+), but mono loudspeakers are not (−).

Safety belt tugs, steering wheel and seat vibration, and visual driver information are only perceptible to the driver. Lateral control intervention via EPS can be very moderate so that other occupants hardly notice it. After all, the recommended action for the driver is conveyed by the assisting steering torque and not the vehicle movement. In contrast, lateral control intervention via ESC should be intense to allow the driver to clearly perceive the vehicle movement and thus deduce a suitable recommended action. ESC intervention is thus distinctly perceived by all occupants, as is the respective audible driver information; this can affect acceptance of the system. Frequent false positives are especially irritating for this type of driver information and also impact acceptance.

In general, driver information is cost effective if the required components are already installed in the vehicle as standard equipment.

Table 2 could be used as a decision-making aid in choosing driver information for a lateral guidance assistance system. However, the weighting of the factors depends to a great extent on type of vehicle, vehicle equipment level, and vehicle manufacturer. In a vehicle with standard EPS, driver information via steering intervention seems to be effective. If EPS is not installed, then driver information via vibration in the steering wheel or seat is an interesting solution. Alternatively, ESC intervention is sensible if returning the vehicle to a lane should be particularly noticeable. On the other hand, for commercial vehicles or transporters (typically without a front passenger) without a suitable steering actuator, the sound of a "rumble strip" or specific information can be prioritized. Taking cost factors into account, this can also be an attractive solution for passenger vehicles. Apart from the driver's subjective evaluation criteria, the acceptance of driver information from LKA systems also depends on influences in culture and society. Where Asian vehicle manufacturers often tend to use audible driver information, European manufacturers tend to use tactile and visual information channels.

5.5 Actuators

Electric power-assisted steering (EPS) systems are typically used in passenger vehicles as lateral control actuators for LKA systems, as described in detail in ▶ Chap. 31, "Steering Actuator Systems". Their influence on the steering torque can be immediately perceived by the driver as tactile feedback. By this means, informative steering wheel vibrations as well as additional steering torque can be implemented indicating recommended actions to the driver. This can be implemented neither for power-assisted hydraulic steering systems nor for superimposed steering systems without using additional actuators.

Targeted braking of individual wheels can also influence the vehicle's lateral control. This effect is used by the ESC system to stabilize the vehicle while driving at the dynamic limits (see ▶ Chap. 39, "Brake-Based Assistance Functions"). With a view to fuel consumption and brake wear, course correcting braking intervention should only be used temporarily and not continually. This makes ESC interventions more suitable for a type I LKA system with a return to lane function than for a type II system with lane-centered control function.

5.6 Status Display and Controls

The system status display must provide the driver with information on the LDW or LKA system's current status in an easily perceptible, but unobtrusive and easily understandable, way. This information is normally given visually. In the simplest case, the availability of the system is indicated to the driver by means of an illuminated LED within an On/Off button for the system (Fig. 7a). Image

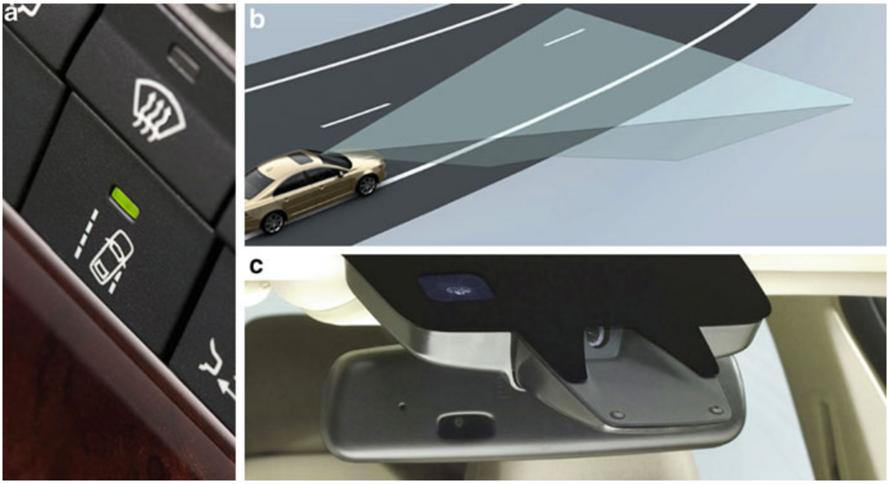


Fig. 7 Lane Departure Warning by Volvo (**a**) On/Off button with status LED (**b**) Camera field of view (**c**) Integration of mono-camera (Source: Volvo)

Fig. 10b shows a more complex solution: an image in the instrument cluster display shows the detected lines, the vehicle, and the vehicle position relative to these lines. Other variants are also used in production systems, resulting from combinations of the two solutions mentioned above.

The transition from "ready for operation" to "not ready for operation" is indicated to the driver by, for example, the LED in the On/Off button being unlit or the lines in the instrument cluster display image changing color. Audible information for the status change is not normally given.

Controls for switching the LDW or LKA system on and off are widely used (pushbutton). Optionally, the driver may be offered the possibility of configuring the system, that is, adjusting the driver information limits, switching specific driver information on and off, and switching an LKA system between type I and type II (in the instrument cluster menus).

6 Exemplary Implementations

LDW systems were introduced in commercial vehicles in 2000 in Europe and shortly after that in the USA. For passenger vehicles, the systems were introduced as of 2001 in Japan, as of 2004 in North America, and as of 2005 in Europe. LKA systems were first introduced in Japan in 2002 and in 2006 in Europe. Today, lateral guidance assistance systems are offered by almost every manufacturer of passenger vehicles ranging from the luxury class to the compact class. Further democratization of these technologies into the small and micro car segments is expected.

The individual vehicle manufacturers' systems can be classified according to the following differentiation characteristics:

(a) LDW (driver information) or LKA (driver information and lateral control intervention)
(b) Type I (return to lane) or Types I & II (return to lane and lane-centered control)
(c) Primary driver information: audible or audible and tactile or tactile

Table 3 shows an overview of lateral guidance assistance systems available in Europe from the various vehicle manufacturers grouped according to these three criteria. OEMs offering both LDW and LKA systems are listed twice. Unless otherwise noted, a mono camera is used as the environment sensor, tactile driver information is provided via steering wheel vibration, and an EPS steering system is used as the lateral control actuator.

It becomes obvious that most manufacturers offer their systems either as LDW or as packages of type I and type II LKA. Type II LKA-only systems (lane-centered control without return to lane) are not available. The primary driver information is either audible or tactile. Audible driver information is preferred by Asian vehicle manufacturers (Daihatsu, Mazda, Honda, Hyundai, Lexus, Toyota) while European OEMs mostly use tactile driver information (Audi, BMW, Citroën, Ford, Peugeot,

Table 3 Overview of lateral guidance assistance systems by various vehicle manufacturers

Primary driver information	LDW	LKA	
		Type I	Type I and Type II
Audible	Daihatsu, Mazda, Opel, Renault, Volvo		Honda, Hyundai, Lexus, Toyota
Audible & tactile	Hyundai[a]	Infiniti[e]	
Tactile	Audi, BMW, Citroën[b,c], Ford, Mercedes-Benz, Peugeot[b,c], VW	Audi, Ford, Lancia, Mercedes-Benz[d,e], Seat, VW	Audi, Ford, Seat, Škoda, Volvo, VW

[a] Seatbelt tensioner
[b] Seat vibration
[c] IR diodes
[d] Stereo camera
[e] ESC intervention

Mercedes-Benz, Škoda, Seat, Volvo, VW). Systems with combined audible and tactile information are an exception (Hyundai, Infiniti).

Apart from Citroën and Renault (infrared diodes) and Mercedes-Benz (stereo camera), all other vehicle manufacturers use a mono camera as the environment sensor. Apart from Hyundai (safety belt pretensioner) and Citroën (seat vibration), all vehicle manufacturers use steering wheel vibration or steering intervention as tactile driver information. Only Infiniti and Mercedes-Benz use course correcting brake intervention to return to a lane via ESC; all other OEMs use EPS. Manufacturers that offer both LDW and LKA systems distinguish between these functional differences via the system names. Ford differentiates between the “Lane Keeping Alert” and “Lane Keeping System.” Mercedes-Benz differentiates between “Lane Keeping Assist” and “Active Lane Keeping Assist.” Volvo uses the product names “Lane Departure Warning” and “Lane Keeping Aid”.

In the following sections, two LDW and two LKA systems by different vehicle manufacturers are shown as examples (see underscore in Table 3). The sections show the differences in terms of feature set, sensors, and driver information in detail. For lateral guidance assistance in commercial vehicles, see ▶ Chap. 52, “Road Assistance for Commercial Vehicles”. All details relate to the information available at the time of authoring this document.

6.1 “Lane Departure Warning” by Volvo

Volvo’s “Lane Departure Warning” informs the driver of an unintentional impending crossing of the lane markers. The system is activated automatically when starting the vehicle and is available as of a speed of approx. 65 km/h. It is automatically disabled below 60 km/h, or by the driver pressing the On/Off button. The system’s operational readiness is indicated via an LED in the On/Off button (Fig. 7a). A camera is used for continuous detection of the lane markers (Fig. 7b).

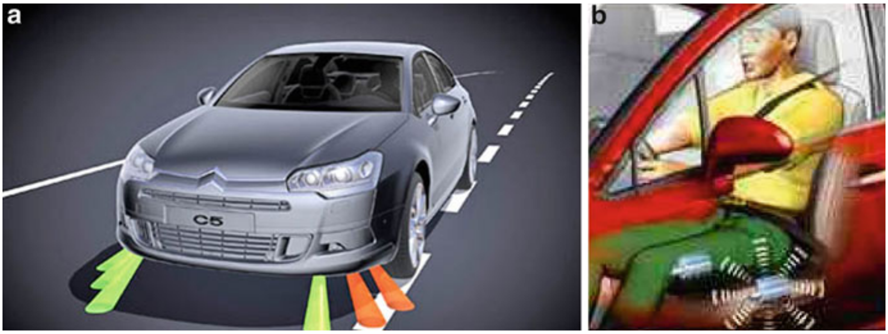


Fig. 8 AFIL from Citroën (**a**) IR diodes' field of view (**b**) Vibration alarm in seat (Source: Citroën)]

It is located behind the windscreen in the base of the interior rear view mirror (Fig. 7c). When the vehicle is about to cross a marker line without an active driver intervention being detected (e.g., turn signal not actuated), the driver's attention is drawn to the situation by means of an audible signal. The sensitivity of the system can be changed from "normal" to "increased" by the driver; in this case, information on unintentional leaving of a lane is provided earlier.

Volvo offers its Lane Departure Warning system in an equipment package named Driver Alert for the S60, S80, V60, V70, XC60, and XC70.

6.2 AFIL from Citroën

The AFIL system by Citroën (Alerte de Franchissement Involontaire de Ligne; which translated means alarm on involuntary line crossing) detects whether a lane marker line is crossed without previously operating the turn signal at speeds above 80 km/h. Three infrared sensors behind the front trim on each side register crossing of the lane markers (Fig. 8a). The driver is informed of the vehicle leaving the lane by a vibration in the seat on the side on which the lane was crossed (Fig. 8b). The driver can press a button to deactivate the function, for example, when traveling on motorways (Renault 2008).

Citroën offers the AFIL system in the DS5, C4, C4 Grand Picasso, C5, and C6.

6.3 Active Lane Keeping Assist by Mercedes-Benz

The Active Lane Keeping Assist system by Mercedes-Benz (Mercedes 2013) monitors the area ahead of the vehicle using a camera system mounted at the top of the windscreen. In addition, RADAR sensors monitor various areas in front of, behind, and at the sides of the vehicle. When a front wheel moves onto a detected

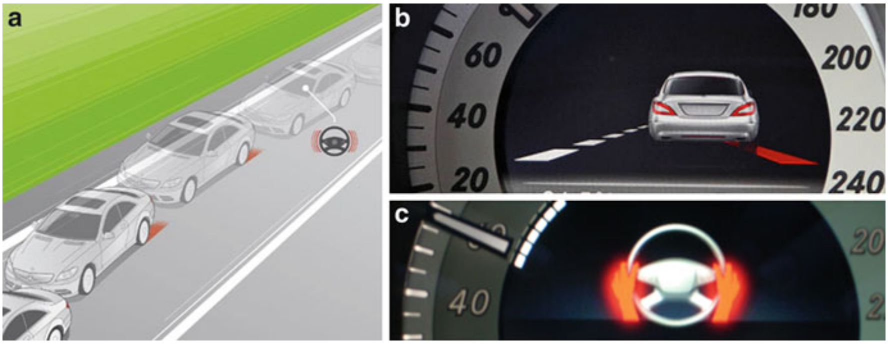


Fig. 9 Active Lane Keeping Assist system by Mercedes-Benz (**a**) Functional principle (**b**) Lane departure warning (**c**) Take-over prompt (Source: Mercedes-Benz)

continuous or dashed lane marker line, driver information is provided as steering wheel interval vibration with duration of up to 1.5 s. If the driver ignores this information, lane correcting brake intervention can return the vehicle to the original lane. The multifunction display then shows a message as depicted in Fig. 9b. If the turn signal is operated beforehand, or a distinct steering wheel movement is performed, the system assumes an intentional lane change, and the driver information is suppressed. The driver can switch the system on or off, or configure the system, using the multifunction display menu. The function is available within a speed range of 60–200 km/h.

Course correcting brake intervention only occurs if the RADAR sensor system is functional. In addition, a lane with marker lines on both sides must have been detected. In the case of dashed lane markers, lane correcting brake intervention will only occur if a vehicle is simultaneously detected in the neighboring lane. Vehicles that may influence brake intervention include oncoming vehicles, overtaking vehicles, and vehicles driving in parallel.

The driver can cancel inappropriate intervention at any time by lightly countersteering, actuating the turn signal, and braking or accelerating noticeably. Lane correcting brake intervention is automatically cancelled as soon as a vehicle safety system intervenes (e.g., vehicle dynamics control by ESC) or if lane markers are no longer detected.

If the driver takes their hands off the steering wheel for a longer period, they are prompted to put their hands back on the steering wheel by a message in the multifunction display (Fig. 9c) in combination with an informative sound. If the driver fails to take over the steering wheel, the Active Lane Keeping Assist system switches off after approx. 5 s. The driver is informed of this via the system status display.

Mercedes-Benz offers its Active Lane Keeping Assist system for E, GLK, SL, and S class vehicles as part of an equipment package named Driver Assistance Package Plus.

6.4 Lane Assist from VW

Lane Assist by Volkswagen (Gies and Brendes 2012) uses a camera to detect lane markers – continuous and dashed lines are detected – and in combination with the vehicle dynamics data, it computes the risk of leaving a lane. If this risk is acute, Lane Assist warns the driver via steering wheel vibration. Depending on the vehicle, the system also gently countersteers to keep the vehicle inside the lane markers within the specified system limits (Fig. 10a). Lane Assist is designed for motorways as well as A roads and rural roads with well-developed infrastructure.

A further development of Lane Assist, which was introduced with the Golf 7 generation, allows the system to be configured. When Adaptive Lateral Guidance is active, Lane Assist does not wait until the risk of the vehicle leaving the lane is imminent before providing assistance. If the lane is marked by marker lines on the left and right, the system continuously attempts to support the driver through gentle steering intervention, thus keeping the vehicle at the center of the lane. The system adapts to the position preferred by the driver within the vehicle's lane. If the driver prefers to drive with a slight offset to the center of the lane, the system learns the new position within a few seconds; this leads to an offset of the ordinate axis in Fig. 6.

Lane Assist can be activated at speeds above 65 km/h; the system switches off at speeds below 60 km/h. The system also works in darkness and poor weather. The driver can always override Lane Assist with very little force and is not relieved of their responsibility for driving the vehicle. To monitor this requirement, the system continuously analyzes the steering activity to discover whether the driver is steering or is letting the system steer the vehicle. In the latter case, the driver is prompted to take control by an audible and visual warning. If the driver ignores this prompt, the system switches off. A change to passive state occurs when the turn signal is actuated, the driver brakes intensively, no markers are detected, or ESC is deactivated.

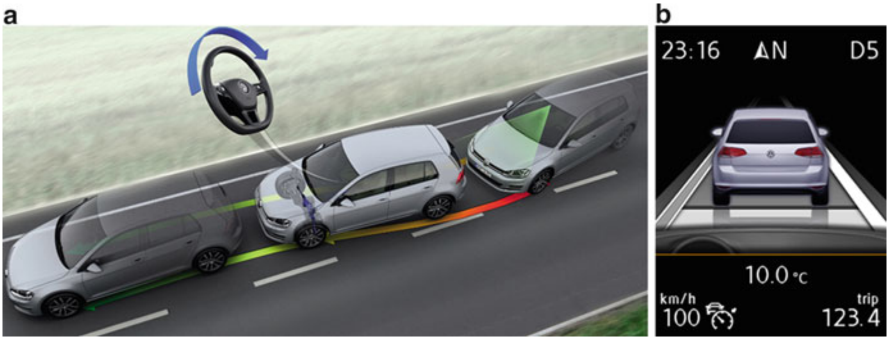


Fig. 10 Lane assist by Volkswagen (**a**) Steering intervention (**b**) Multi-function display (Source: Volkswagen)

In rarely occurring situations, the driver will be prompted to take over steering by steering wheel vibration. Vibration occurs if correcting steering intervention is insufficient to keep the vehicle within a lane, or the system fails to detect lane markers during intensive steering intervention.

Lane Assist is available for almost all models by Volkswagen.

7 System Evaluation

All lateral guidance assistance systems support the driver in keeping to a lane, prevent unintentional leaving of a lane in many cases, and thus make a contribution toward positively influencing the accident situation. However, on taking an in-depth look at the specific system characteristics, differences are evident that affect the accident avoidance potential as well as customer acceptance.

The accident avoidance potential and the driver support are possibly higher with type I LKA systems than with LDW systems. This is suggested by analysis of accident events and research with driving simulators (Daschner and Gwehenberger 2005; Navarro et al. 2007). Customer benefits from LKA systems could also be enhanced once again by type II systems. Type II systems continuously support the driver in keeping the vehicle at the center of a lane, thus allowing for a relaxed and comfortable driving experience. In contrast to that, the driver will experience type I systems only if a potentially critical lane departure occurs, a situation that is rare (Freyer et al. 2010).

Customer acceptance of the various characteristics of LDW and LKA systems was presented by the German automobile club ADAC in 2012 in a survey (ADAC 2012). The requirements for an ideal system were derived from the user responses.

According to the ADAC, drivers with a lane departure warning (LDW) system prefer a warning by means of steering wheel vibration when crossing a line, as this warning is exclusively perceived by the driver. Ideally, lane marker detection should be visualized in a display, or head-up display if available. The ability to adjust the warning points (proximity to lane markers) and the vibration intensity was much appreciated. The respondents evaluated adaptive settings as good; this helped to avoid confronting an active driver with too many unnecessary warnings when driving on winding roads, which drivers could perceive as irritating and lead to them switching the system off.

According to the ADAC, lane keeping assistant (LKA) steering corrections were well accepted. Respondents wanted steering intervention to be adjustable so that preferred settings could be selected, either for keeping the vehicle to the center of a lane (type II) or for pushing the vehicle away from the edge of the lane (type I). Intervention in the form of ESC braking of the individual wheels was also appreciated by many users. Ideally, respondents wanted the identified lane markers to be shown in the instrument cluster display, or head-up display if available. Steering intervention requires assured lane marker detection. If this were not possible due to poor markings, then at least a lane departure warning should be issued.

Audible driver information with system-specific sounds that allowed the sides to be distinguished was praised; however, informative sounds were generally perceived to be less than intuitive, and sometimes irritating. The ability to differentiate between sides was positively accepted for vibrating seats. However, the respondents were divided on the assessment of vibrating seats in general; some found them useful as they allowed for unambiguous allocation of driver information. Others found vibrations in the seat annoying and irritating, even after a long drive. In the case of an ESC intervention, the respondents were a bit unhappy with the associated speed reduction. Lane detection on county roads provoked complaints from almost all respondents.

The crucial factors for customer acceptance are driver information design as well as system availability. This is shown in a comparison test by the German car magazine Auto Bild (Auto Bild 2011). Given similar levels of availability, LKA systems asserted themselves against LDW systems. However, LDW systems with good availability were on a par with, or even ahead of, LKA systems with poor availability.

8 Achieved Performance

The performance of lateral guidance assistance systems has been continuously improved in the last few years. This was achieved by the use of high-resolution and highly dynamic color cameras, adaptive warning algorithms with driver intention recognition, and more robust image processing algorithms.

Nevertheless, these systems still have many limitations. This becomes evident on reading the user manuals by the vehicle manufacturers. These manuals point out that lateral guidance assistance systems are merely an aid, and that responsibility for lateral control of a vehicle remains with the driver. Depending on the manufacturer, the function is only available above speeds of between 60 and 70 km/h. This rules out complex lane marking scenarios in cities to a large extent, and lateral guidance assistance is not available.

Lane markers cannot always be unambiguously detected, for example, in cases of poor visibility, dazzling, covered or worn lines, soiled or sticker-covered windscreens, in areas with construction zones, winding roads, tree-lined roads, and entrances to and exits from tunnels. These cases can impair (missing or unnecessary driver information) or disable a system. In the case of LKA systems, steering intervention may not always be sufficient to return the vehicle to the lane. Road and weather conditions are not taken into account in steering intervention. Thus, from a customer's point of view, the availability and robustness of lateral guidance assistance systems could be improved.

9 Outlook

The availability and robustness of lateral guidance assistance systems could further be improved by considering in the vehicle's lane detection and lateral control not only marker lines but also other environmental features such as other vehicles, kerbs, edging strips, or guard rails. If only a few features are used, then a simple rule-based approach might be appropriate. For a large number of features, a model-based approach appears meaningful. This approach involves supporting or refuting a lane hypothesis based on environmental features. In doing so, lane markers can be given higher priority than other features. Already today, many of these features are detected by existing environment sensors. Consequently, the challenge lies in the subsequent data processing and/or road modeling.

Mercedes-Benz's Steering Assist shows an initial approach to road modeling. At speeds above approx. 60 km/h, it uses available lane markers for orientation. In a speed range of 0–60 km/h, it orientates its behavior on vehicles moving ahead – detected by RADAR sensors – while also taking lane markers into consideration, for example, for stop-and-go driving in traffic jam situations (Mercedes 2013). The system is not intended to notify the driver if they unintentionally leave a lane but to enhance the driving experience in stop-and-go driving situations. In stop-and-go driving, if the leading vehicle were to leave the lane at a motorway exit, the Steering Assist system would follow – that is, it would also change lanes – without informing the driver. One thing is clear, however: vehicle lateral control using the positional data of the vehicle driving ahead is possible, although there are some restrictions. This demonstrates the potential of road models for lateral control that use the largest possible set of environmental features in addition to lane markings.

If the robustness and availability of lateral guidance assistance systems can be further improved, there is potential for many other systems with combined lateral and longitudinal vehicle control ranging from Traffic Jam Assist (▶ Chap. 51, "Traffic Jam Assistance and Automation") to fully automated driving (▶ Chap. 62, "Autonomous Driving").

References

ADAC survey (2012) Warnsignale von Assistenzsystemen (Warning signals from driver assistance systems) http://www.adac.de/infotestrat/tests/assistenzsysteme/assistenzsysteme_2012/. Accessed 27 Oct 2014

Auto Bild comparison test (2011) On-board-computersysteme im Vergleich: Wer spurt am besten? (On board computer systems compared: who has the best lane keeping performance?). Auto Bild 32:36–46

Daschner D, Gwehenberger J (2005) Wirkungspotenzial von Adaptive Cruise Control und Lane Guard System bei schweren Nutzfahrzeugen (Potential effectiveness of adaptive cruise control and lane guard system for heavy commercial vehicles). Allianz Zentrum für Technik GmbH, Report no. F05-912, commissioned by MAN for the Federal Ministry of Education and Research (BMBF) project Safe Truck, Munich

Derendarz W, Graf T, Wahl FM (2010) Monokamerabasierte Umfelderkennung für komplexe Umgebungen (Mono camera-based environment detection in complex environments). In:

4. VDI-Tagung Optische Technologien in der Fahrzeugtechnik (VDI congress optical technologies in vehicle technology), VDI Reports 2090, pp 141–156

ECE-R 79 (2005) Uniform provisions concerning the approval of vehicles with regard to steering equipment

Freyer J, Winkler L, Waarnecke M, Duba GP (2010) Eine Spur aufmerksamer – Der Audi Active Lane Assist. (A tad more attentive - Audi Active Lane Assist). ATZ 112(12):926–930

Gayko J (2012) Lane Keeping Support. In: Winner H (ed) Handbuch Fahrerassistenzsysteme (Handbook of driver assistance systems), 2nd edn. Springer, Wiesbaden, pp 554–561

GIDAS (German In-Depth Accident Study) database excerpt (2012) Passenger car occupants with severity of injury MAIS2+ in initial collisions

Gies S, Brendes C (2012) Der neue Golf – Fahrwerk – Modularität als Prinzip (The new Golf – Running Gear – Modularity as a Principle). ATZ extra 11(2013):52–63

Hegemann S, Lüke S, Nilles C (2013) Randsteinerkennung als Teil der Urbanen Fahrerassistenz (Kerb detection as a component in urban driver assistance). ATZ 115(11):894–899

Homm F, Kaempchen N, Burschka D (2011) Fusion of Laserscannner and video based lanemarking detection for robust lateral vehicle control and lane change maneuvers. IEEE Intell Vehicle Symp IV:969–974

ISO standard 17361 (2007) Intelligent transportation systems – lane departure warning systems – performance requirements and test procedures. British Standards Institution

ISO DIS standard 11270 (2013) Intelligent transport systems – lane keeping assistance systems (LKAS) – performance requirements and test procedures. International Organization for Standardization

Koelbl C (2011) Darstellung und Potentialuntersuchung eines integrierten Quer- und Längsreglers zur Fahrzeugführung (Implementation and potential analysis of an integrated lateral and longitudinal controller for vehicle guidance). Dissertation, Cuvillier, Göttingen

Mammar S, Glaser S, Netto M (2006) Time to line crossing for lane departure avoidance: a theoretical study and an experimental setting. IEEE Trans Intell Transp Syst 7(2):226–241

Mann M (2008) Benutzerorientierte Entwicklung und fahrergerechte Auslegung eines Querführungsassistenten (User-oriented development and driver-appropriate design of lateral guidance assistance systems). Dissertation, Cuvillier, Göttingen

Mercedes-Benz E Class Operating instructions (2013)

MLIT – Ministry of Land, Infrastructure Transport and Tourism (2013) Technical guidance to lane-keeping assist devices of motor vehicles. In: Technical guidelines in Blue Book 11-6-1-3

Montemerlo M, Becker J, Bhat S, Dahlkamp H, Dolgov D, Ettinger S, Haehnel D, Hilden T, Hoffmann G, Huhnke B, Johnston D, Klumpp D, Langer D, Levandowski A, Levinson J, Marcil J, Orenstein D, Paefgen J, Penny I, Petrovskaya A, Pflueger M, Stanek G, Stavens D, Vogt A, Thrun S (2008) Junior: the Stanford entry in the urban challenge. J Field Robotics 25(9):569–597

Navarro J, Mars F, Hoc JM (2007) Lateral control assistance for car drivers: a comparison of motor priming and warning systems. Human Factors 49(5):950–960

Renault press release (2008) Peugeot 308 SW – Spurhalteassistent AFIL erhöht aktive Fahrsicherheit (Lane keeping assist AFIL improves active driving safety) www.peugeot-presse.de/download/rtf/PM_308SW_200807.rtf. Accessed 27 Oct 2014

Schmidt G (2009) Haptische Signale in der Lenkung: Controllability zusätzlicher Lenkmomente (Tactile signals in steering: controllability of additional steering torques) In: Berichte aus dem DLR-Institut für Verkehrssystemtechnik (Reports by the DLR Institute of Traffic System Technology), vol 7, Dissertation

TRW product information (2013) Skalierbare Kamera (Scalable camera). http://www.trw.de/technology_information/electronics/driver_assist_system_electronics. Accessed 27 Oct 2014

van Winsum W, Brookhuis KA, de Waard D (2000) A comparison of different ways to approximate time-to-line crossing (TLC) during car driving. Accid Anal Prev 32:47–56

Lane Change Assistance

49

Arne Bartels, Marc-Michael Meinecke, and Simon Steinmeyer

Contents

Abstract

More than 5 % of all accidents involving injury to people take place as the result of a lane change. Therefore, it is sensible to provide the driver with a lane change assistant in order to provide support in this driving maneuver.

A. Bartels (✉) • M.-M. Meinecke
Fahrerassistenz und integrierte Sicherheit, Volkswagen AG – Konzernforschung, Wolfsburg, Germany
e-mail: arne.bartels@volkswagen.de; marc-michael.meinecke@volkswagen.de

S. Steinmeyer
Volkswagen AG, Wolfsburg, Germany
e-mail: simon.steinmeyer@volkswagen.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_50

ISO standard 17387 "Lane Change Decision Aid System" differentiates between three different types of system: the "blind spot warning" which monitors the blind spot on the left and right adjacent to the driver's own vehicle, the "closing vehicle warning" which monitors the adjacent lanes to the left and right behind the driver's own vehicle in order to detect vehicles approaching from behind, and the "lane change warning" which combines the functions of "blind spot warning" and "closing vehicle warning."

Almost all major vehicle manufacturers are now offering systems that assist the driver to change lanes. Systems with "blind spot warning" are available from Citroën, Ford, GM, Jaguar, Jeep, Land Rover, Lexus, Mercedes, Nissan/Infiniti, Opel, and Volvo. Systems with "lane change warning" are available from Audi, BMW, Mazda, Porsche, Volvo, and VW. All vehicle manufacturers use an optical indicator in or near to the exterior mirrors in order to show information for the driver. The majority of vehicle manufacturers use RADAR sensors that are installed in the rear of the vehicle. Two-level, escalating driver information is only offered in some of the systems. The type of escalation (optical, acoustic, tactile, lateral guidance intervention) usually differs from one vehicle manufacturer to another.

The performance capability of the lane change assistants described above is already quite considerable. However, all of these systems have their limits, and the vehicle manufacturers need to inform drivers of these in the owner's manual, for example.

1 Motivation

Driver assistance systems support the driver in his driving task. The customer benefit expected from a driver assistance system is particularly high if the driving task in which the driver is to be assisted is one which harbors a high potential for error. The lane change is one of these driving tasks with a high error potential.

This is indicated by a statistical analysis of accidents involving injuries to people, which have been collected in a database held at Volkswagen Accident Research and GIDAS (German In-Depth Accident Study). Figure 1 shows the accidents for the years 1985–1999 where changing lanes with a car was the main cause of the accident for urban, rural, and motorway traffic, as a percentage. It becomes clear that on average, more than 5 % of all accidents are the result of changing lanes. It is also clear that the majority of these accidents occur on rural roads or motorways.

These considerations indicate the importance of providing drivers with a system that can support them when changing lanes. Initially, this support is to be configured for rural road and motorway scenarios.

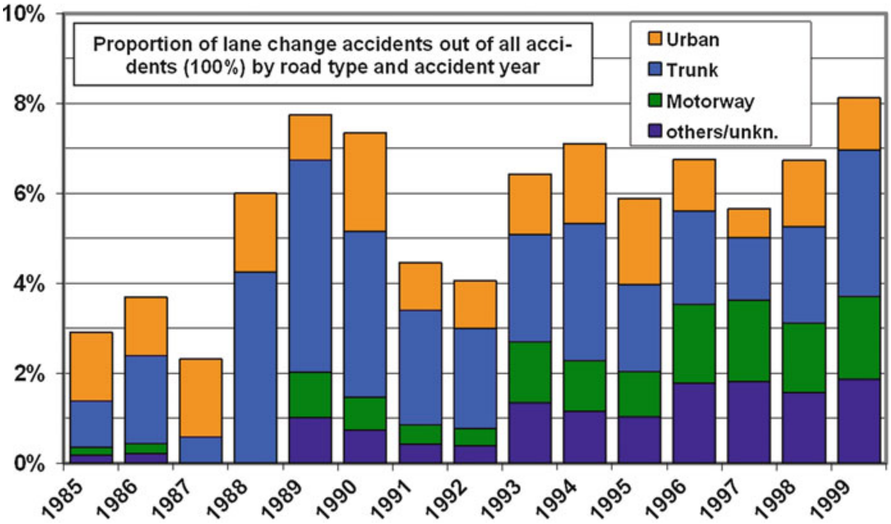


Fig. 1 Lane change accidents with a passenger car as the main causal factor as a proportion of all accidents by road type and accident year (GIDAS 2008)

2 Requirements

The driver needs to be able to preclude any risk to other road users when making a lane change. According to current regulations, it is the driver's responsibility to check the area at the rear and to the side of the vehicle before changing lanes. It is mandatory both to look in the exterior and rearview mirrors and to glance over the shoulder, as part of this procedure. Omitting the glance over the shoulder, or if the exterior mirrors are not set correctly, or if the driver is simply inattentive, may lead to the risk of failing to notice other road users in the blind spot. If a lane change maneuver is initiated under such circumstances, then this can result in a collision with the vehicle in the adjacent lane.

Another frequent cause of accidents when changing lanes is the failure to estimate the speed of overtaking vehicles correctly. On motorways and fast roads in particular, drivers frequently underestimate the approach speed of vehicles in the far distance and that of vehicles which are coming up quickly from behind. In this situation, a lane change can lead not only to a collision with the overtaking vehicle if it is unable to brake sufficiently but also to rear-end collisions with other road users who are unable to respond in good time to the rapid deceleration of the overtaking vehicle.

The driver also needs support when changing lanes toward the passenger side. This maneuver is mandatory in Germany, due to the obligation to keep to the right wherever possible. Following an overtaking maneuver, the driver is obliged to return to the right-hand lane as soon as the traffic situation permits. In contrast to this, overtaking on the passenger side is also practiced in many other European countries. Furthermore, in the USA, it is an everyday reality that other road users will be driving in the blind spot of the subject vehicle at almost the same speed in both adjacent lanes.

The above analysis leads to the following functional requirements of a lane change assistant:

- The lane change assistant should inform the driver about hazardous situations that result due to the driver failing to monitor the area around the vehicle adequately.
- For this purpose, the assistance function should be capable of detecting road users approaching rapidly from behind as well as detecting other road users in the blind spot of the subject vehicle.
- The assistance function should operate for both adjacent lanes equally, on the driver and the passenger sides.
- Ideally, the assistance function will be available for all road, weather, and traffic conditions with approximately the same level of quality.

The human-machine interface (HMI) between the driver and lane change assistant is highly important. If the system's monitoring of the area around the vehicle indicates that the lane change is potentially problematic, then the driver is informed about this in a suitable and timely manner. The information can in principle be delivered via the visual, acoustic or haptic sensory channels. In configuring the HMI, however, attention should be paid to encouraging the driver to look in the mirror, as this remains an obligation on the driver even when the lane change assistant is activated. Positioning optical indicators in or near to the exterior mirrors represents a solution to this requirement. The spatial proximity between the exterior mirror and indicator lamps ensures that the driver can see the visual information from the assistance system when looking in the mirror. The brightness of these indicator lamps should be configured so that they can be easily seen by the driver under all ambient conditions that arise. On the other hand, neither the driver nor drivers of other vehicles should be distracted or dazzled by the indicator lamps, particularly at night.

When configuring the HMI, it is also necessary to decide whether the driver information should be provided on a single- or two-level basis. In a two-level driver information system, escalation from information level 1 to information level 2 takes place as soon as the driver's intention to change lanes is detected. This escalation does not take place in a single-level driver information system.

In information level 1, the driver is shown each vehicle that is a potential hazard when changing lanes. This happens even if the driver is not intending to change lanes. The display of level 1 information should be noticeable to the driver, but it should cause neither interference nor distraction, even when activated frequently. If indicator lamps are positioned in or close to the exterior mirrors, then this can be achieved, for example, by appropriate regulation of the light brightness depending on the ambient light level.

In information level 2, the driver's intention to change lanes is additionally detected, e.g., by actuation of the turn signal lever. If the driver intends to make a lane change and this lane change is evaluated as potentially dangerous based on the system's detection of the surroundings, then more intensive information should be provided to the driver. In terms of providing information for the driver via optical indicators in or close to the exterior mirrors, this can be by means of a very bright, brief flash of the indicator lamps, for example. Haptic or acoustic information can also be employed.

Another important factor for the lane change assistant is an intelligent information strategy. In order to ensure sufficient customer acceptance, the lane change assistant must be capable of reliably displaying all traffic situations that are detected as being potentially dangerous. On the other hand, the driver must not be given unnecessary information. Unnecessary in this context includes information about a vehicle in the adjacent lane that, although detected by the environment sensors, is moving sufficiently slowly and is still far enough away to allow a lane change to be performed without risk. Another unnecessary item of information would be about a vehicle driving straight ahead in the next lane but one. The information strategy therefore needs to evaluate the measurement data from the environment sensors and, based on this, decide very carefully whether the driver should be informed or not.

3 Classification of System Functions

The current ISO standard 17387 "Lane Change Decision Aid System" specifies various configurations of the lane change assistant and classifies them into various subtypes. Furthermore, a system status diagram is defined with system statuses and transition conditions. These are presented below.

3.1 Classification According to Performance of Environment Perception

Three different system types are defined in ISO standard 17387. These differ in terms of the zones monitored by the environment sensors. Table 1 shows an overview.

The specified system types have the following functions:

- Type I systems provide information about vehicles in the blind spot on the left and right sides. They do not provide information about vehicles that are approaching from the rear on the left or right sides.

Table 1 Classification by zone coverage (ISO 2008)

Type	Left adjacent zone coverage	Right adjacent zone coverage	Left rear zone coverage	Right rear zone coverage	Function
I	X	X			Blind spot warning
II			X	X	Closing vehicle warning
III	X	X	X	X	Lane change warning

Table 2 Classification by target vehicle closing speed and roadway radius of curvature (ISO 2008)

Type	Maximum target vehicle closing speed	Minimum roadway radius of curvature
A	10 m/s	125 m
B	20 m/s	250 m
C	30 m/s	500 m

- Type II systems provide information about vehicles that are approaching from the rear on the left and right sides. They do not provide information about vehicles in the blind spot on the left or right sides.
- Type III systems provide information both about vehicles in the blind spot and vehicles approaching from the rear, in both cases on the left and right sides.

Type II and type III systems are themselves subdivided into three subcategories. In the aforementioned standard, these are distinguished according to the maximum permitted relative speed of the target vehicle approaching from the rear v_{max}, as well as the minimum permitted roadway radius of curvature R_{min}. Table 2 shows an overview.

The maximum relative speed between the subject vehicle and the vehicle approaching from behind has a direct influence on the required sensor range, given the computation time of the system and a specified minimum response time of the driver. At $v_{max} = 20$ m/s, a computation time of the system of 300 ms and a required minimum response time of 1.2 s, the minimum sensor range is $20\,\text{m/s} \times (1.2 + 0.3\,\text{s}) = 30\,\text{m}$. The sensor range must be increased if information is to be provided in good time at even faster approach speeds. A speed of $v_{max} = 30$ m/s means the minimum sensor range needs to be as much as 45 m.

There are two reasons for classification with regard to the minimum roadway radius of curvature. Firstly, early detection of the target vehicle can be made more difficult by the restricted detection range of the environment sensors used. For example, given a cone-shaped detection range, the aperture angle of the sensor is a significant factor in achieving good coverage of the relevant lane when driving through a bend. On the other hand, for a given curve radius and a typical subject vehicle speed, the maximum relative speed of the target vehicle approaching from behind is limited by the dynamic driving properties of the target vehicle.

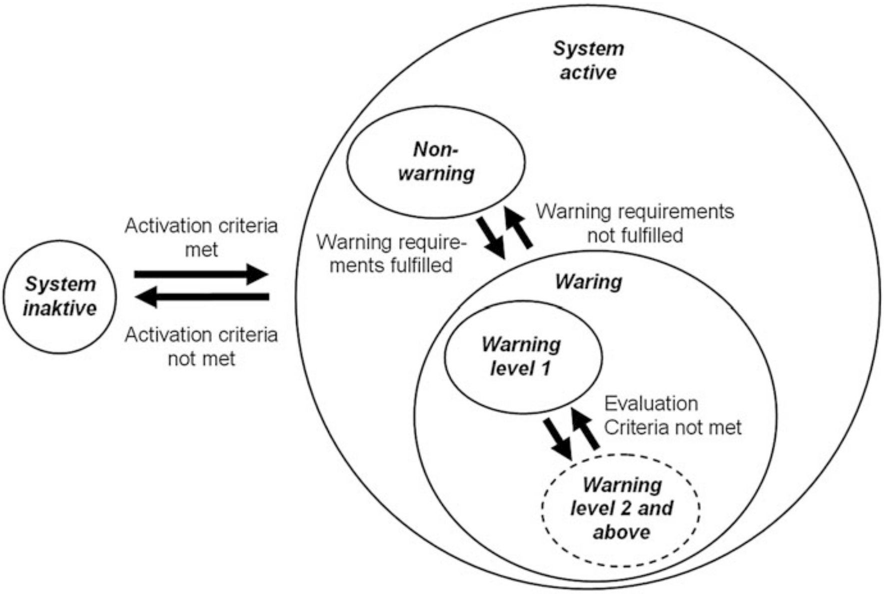


Fig. 2 System state diagram for a lane change assistant according to ISO 17387 (ISO 2008)

3.2 System State Diagram

ISO standard 17387 specifies a system state diagram with system statuses and transition conditions for the lane change assistant. This is shown in Fig. 2.

No information is given to the driver if the system is inactive. Certain criteria have to be met in order for the system to be activated. For example, the system can be activated by pressing a button if the subject vehicle is driving faster than a specified minimum activation speed. The system is deactivated, for example, if the driver presses the off button or is driving below the minimum activation speed.

If the system is active, then the information is only given to the driver providing certain conditions are met, for example, a vehicle is detected in the blind spot or else a vehicle is approaching from the rear at high speed. No information is given to the driver if these conditions are not met.

The information can be given to the driver using several distinct levels. Information level 1 involves “discreet” information to the driver at a lower level of urgency than level 2 driver information. It is informative in character. The driver is given level 2 information if certain evaluation criteria are met, indicating that it is the driver’s intention to make a lane change. These selection criteria can include:

(a) Operation of the turn signal lever
(b) An analysis of the steering angle or steering torque
(c) The position of the subject vehicle within the lane
(d) The lateral distance from a vehicle in the adjacent lane

In case (c), it is possible to make use of synergies with a system that is possibly already present for detecting the lane markings, for example. The driver information of level 2 can occur in principle staggeredly and/or in multiple levels. If the driver has been clearly informed when, for example, the turn signal lever is operated while a vehicle is in the blind spot and in spite of the information the driver moves the vehicle into the adjacent lane, then the intensity of driver information can be increased again or even an intervention in the lateral vehicle guidance can take place.

4 Examples of Implementation

Driver assistance systems which support the driver when making a lane change have been available from several vehicle manufacturers for several years. Initially, these were used in upper category vehicles: at Audi in the A8 and the Q7, at VW in the Phaeton and the Touareg, at BMW in the seven series, and at Mercedes in the S-class. However, a democratization of this particular driver assistance system can now be observed. Many vehicles in the mid-range class and lower mid-range class are now fitted with lane change assistance systems, for example, the Audi A4 and the A3, the BMW 3 series, the Ford Focus, the Mercedes A and B class, the Mazda 3, the Volvo V40, and the VW Passat.

The characteristics of the systems from individual vehicle manufacturers can differ markedly from one another, although the majority of them can be assigned to one of the categories of ISO 17387, as was described in Sect. 3. The primary differences are:

(a) The various sensors used for environment perception
(b) The number of information levels
(c) The classification into type I and type III systems

Table 3 shows an overview of the lane change assistant systems currently available on the market, listed according to these three criteria. A third information level was added in order to simplify a system comparison. Vehicle manufacturers that offer their systems with the option of a two-level or three-level driver information system are listed twice.

It is apparent from the table that many manufacturers offer type I or type III systems for their models. Type II systems, which do not have blind spot information and which only provide information about vehicles approaching from the rear, are not currently offered on the market. Ultrasonic, camera, and RADAR sensors with sensor ranges of 3–5 m are used for systems of type I which exclusively monitor the blind spot. RADAR sensors with a range between 70 and 100 m (class B) are used only for type III systems, which provide information about vehicles in the blind spot and about vehicles approaching from the rear. Due to their restricted sensor range, ultrasonic and camera sensors are not used for type III systems, and the same also applies to RADAR sensors with a limited range (class A).

Table 3 Overview of lane change assistance systems from various manufacturers listed according to type, warning levels, and environment sensors

Warning	Type I		Type III	
1 level	Ultrasonics	Citroën, Opel		
	Camera	Volvo		
	RADAR (A)	Ford, Jaguar, Jeep, Land Rover, Lexus		
2 levels	RADAR (A)	GM, Mercedes	RADAR (B)	Audi, BMW, Mazda, Porsche, Volvo, VW
3 levels	RADAR (A)	Mercedes, Nissan/Infiniti		VW

It is also apparent that multilevel driver information is currently only available with RADAR-based systems.

The different vehicle manufacturers have each chosen different product names in order to differentiate themselves from their competitors and almost certainly to emphasize their proprietary system functions. The lane change assistant system from Audi is known as "Audi Side Assist." The almost identical system from VW is known as "Side Assist." At BMW, the system is known as "Lane Change Warning." The system from Citroën is called the "Blind Spot Monitoring System." Ford uses the name "Blind Spot Information System." GM uses the name "Side Blind Zone Alert." Mercedes-Benz calls its system "Blind Spot Assist." Mazda, for its part, uses the name "Rear Vehicle Monitoring System." Nissan/Infiniti uses the name "Side Collision Prevention." Volvo has called its system the "Blind Spot Information System."

In the following, manufacturers' series systems are provided as examples for each category listed in Table 3 (underscored). The focus here is on the differences between functions, sensors, and driver information.

4.1 "Blind Spot Monitoring System" from Citroën

The "Blind Spot Monitoring System" from Citroën informs the driver as soon as a motor vehicle or motorcycle enters the blind spot of the vehicle. In this case, a warning LED lights up in one of the exterior mirrors (Fig. 3a). A second warning level that is activated by the turn signal lever, for example, is not available. The system can be activated by a button and is active between speeds of 10 and 140 km/h. No warning is given to the driver outside these limits. According to manufacturer's details, the system supports the driver in complex traffic situations at low relative speeds therefore mainly in urban areas in city traffic and on urban motorways as well as on multilane rural roads.

The system monitors the area approx. 5 m behind the rear bumper and up to 3.5 m to the side of the vehicle. Four ultrasonic sensors are used, and these are fitted into the sides of the front and rear bumpers. Both rear sensors are used to monitor the blind spots (Fig. 3b). Both of the front sensors are used only as a plausibility check.

Citroën first introduced the "Blind Spot Monitoring System" in the C4 in 2010 and currently offers it as part of a package with the Tire Pressure Monitor for €290.

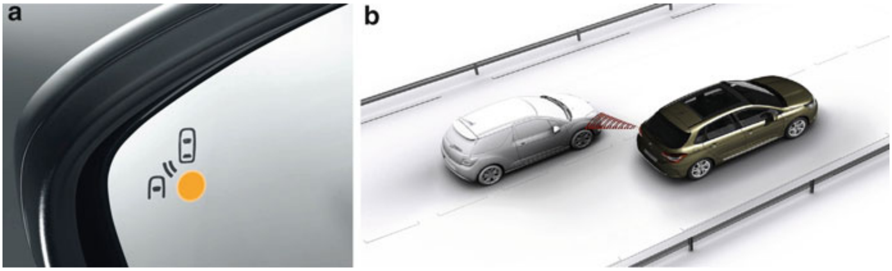


Fig. 3 "Blind Spot Monitoring System" from Citroën (Heise 2010; Citroën 2013). (**a**) *Yellow* light behind the mirror glass of the exterior mirror. (**b**) Schematic diagram of function

4.2 "Blind Spot Information System" (BLIS) from Volvo

The BLIS from Volvo informs the driver of vehicles that are in the driver's blind spot. In particular in dense traffic, this is intended to avoid traffic accidents due to lane changes. The system is based on two digital cameras integrated into the exterior mirrors. These rearward-pointing cameras monitor the traffic in both adjacent lanes to the right and left of the subject vehicle. If a vehicle enters the blind spot, a light in the right or left A-pillar lights up discreetly to inform the driver of this (see Fig. 4a). There is no escalation of the driver's information, e.g., when the turn signal lever is operated.

The monitoring range of the cameras is restricted to a corridor about 3 m in width and 9.5 m in length on the left and right of the subject vehicle (see Fig. 4b). BLIS detects all objects that are moving up to 70 km/h faster or 20 km/h slower than the subject vehicle.

The camera-based version of BLIS was first introduced by Volvo in model year 2005 and gradually introduced into almost all of the Volvo car range. In 2012, Volvo introduced the RADAR-based Enhanced BLIS (type III system) (Volvo 2012). It is currently available for the S60, V40, V60, and XC60 models. The camera-based BLIS is currently available for the S80, V70, XC70, and XC90 models. The system is currently available in the V40 for €540. In all other models, the additional charge is €620.

Currently Volvo is the only manufacturer to offer a camera-based lane change assistance system.

4.3 "Blind Spot Information System" from Ford

The "Blind Spot Information System" from Ford automatically monitors the blind spot adjacent to the vehicle while driving, with the help of short-range RADAR sensors. A yellow warning light in the exterior mirror on the side in question provides an active indication to the driver as soon as another vehicle (truck, car, motorcycle, bicycle, etc.) is located in the vehicle's blind spot (see Fig. 5a).

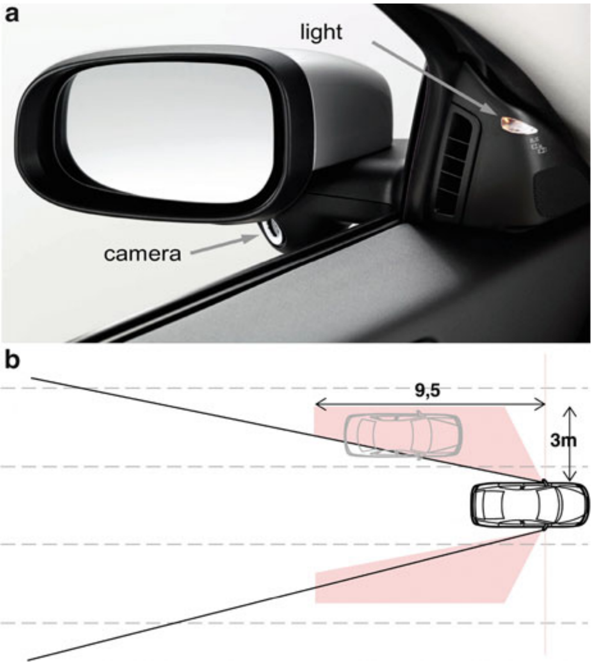


Fig. 4 Blind Spot Information System (BLIS) from Volvo (Gizmag 2007). (**a**) Camera integrated into the exterior mirror and *yellow* light in the A-pillar. (**b**) Monitoring range of the sensors

According to Ford, the “Blind Spot Information System” provides enhanced visibility and therefore reduces stress on the driver, as well as increasing road traffic safety. The system’s activation speed is 10 km/h.

The RADAR sensors are positioned at the sides of the rear bumper. They operate in the frequency range around 24 GHz. The side area is monitored using several pronounced antenna lobes, allowing the bearing of the objects to be assigned (Fig. 5b). The area monitored on the driver and passenger sides is approx. 3 m wide and extends from the exterior mirror backwards to approx. 3 m behind the rear of the vehicle.

Currently Ford offers their customers in Germany the “Blind Spot Information System” for the Focus, Mondeo, C-Max, S-Max, Kuga, and Galaxy as an option for an additional charge between €390 and €675, depending on the model and the equipment.

Systems with similar RADAR sensors and driver information are currently offered by Jaguar, Jeep, Land Rover, and Lexus.

4.4 “Active Blind Spot Assist” from Mercedes-Benz

The “Active Blind Spot Assist” from Mercedes-Benz monitors the blind spot of the subject vehicle on the driver and passenger sides using short-range RADAR sensors.

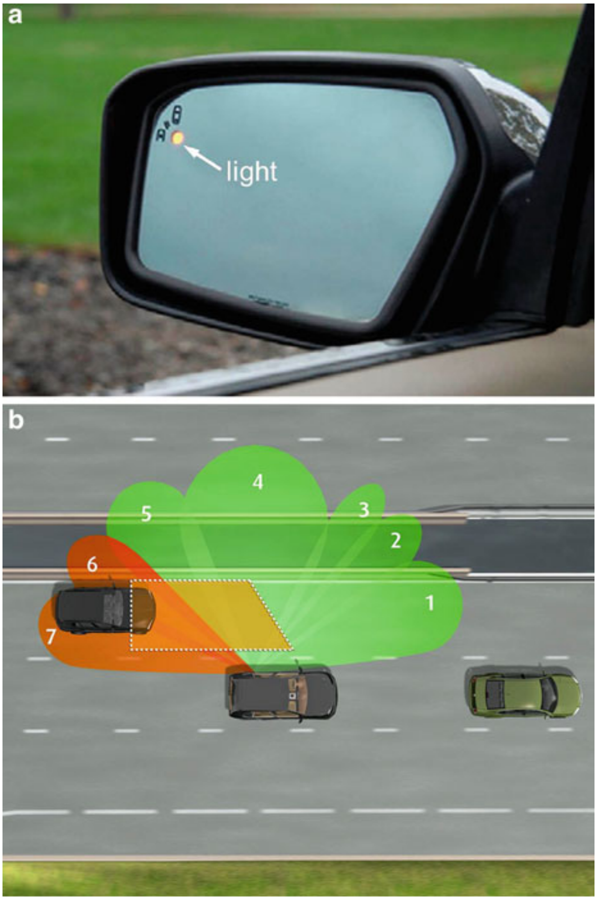


Fig. 5 “Blind Spot Information System” from Ford. (**a**) *Yellow* light behind the mirror glass of the exterior mirror (IndianCarsBikes 2010). (**b**) Detection area of the RADAR sensors (Prova 2007)

These employ wide-band transmission with a mean frequency of 24 GHz and are integrated into the front and rear bumpers of the vehicle where they cannot be seen from the outside.

As soon as a vehicle is detected in the monitored area, the driver is informed by means of a red light illuminated continuously in the virtually invisible indicator lamps integrated into the exterior mirrors (Fig. 6a). An indication about the threat of a collision is given if a vehicle has been detected in the blind spot monitoring area and the driver has activated the turn signal. There is a one-off double beep and the red light flashes. If the turn signal continues to be activated, detected vehicles are permanently indicated by the red light flashing continuously. There is no repetition of the acoustic driver alert.

If the driver ignores this indication and initiates a lane change to the adjacent lane by turning the steering wheel, there is the danger of an immediate collision. This is detected by the Active Blind Spot Assist with the aid of a camera, which detects the lane markings, as well as the front and rear RADAR sensors which

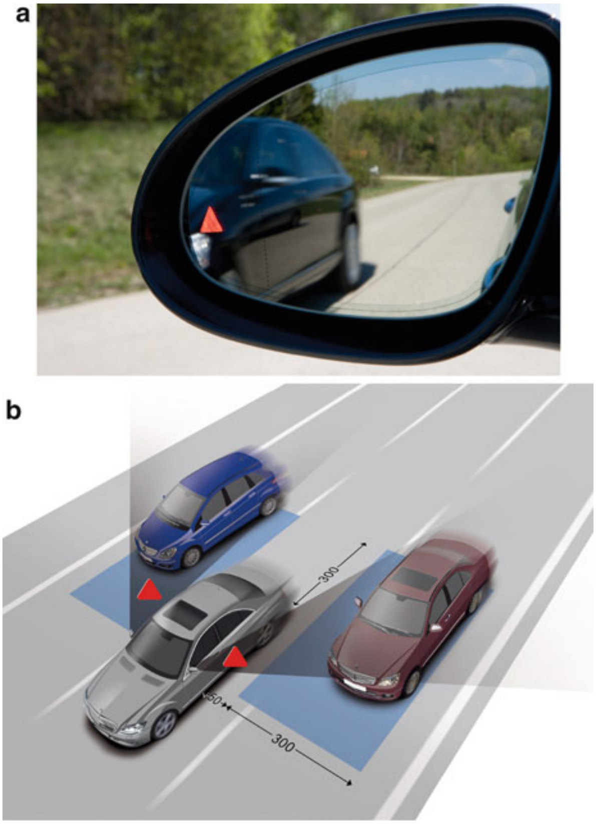


Fig. 6 "Blind Spot Assist" from Mercedes-Benz (Heise 2007). (**a**) Integration of lamp in exterior mirror. (**b**) Detection range

detect other road users. The system then initiates a corrective braking intervention, which can always be overridden by the driver. In addition, the driver is warned by a double beep, continuous flashing of the red light, as well as an indicator lamp in the multifunction display. The corrective braking intervention by the Active Blind Spot Assist is available at speeds of between 30 and 200 m km/h. The Blind Spot Assist is inactive at speeds below 30 km/h, and the indicator lamps in the left and right exterior mirrors light up yellow.

Figure 6b shows the detection range of the sensors. The monitored area is about 3 m wide, starting at a distance of about 50 cm from the side of the vehicle. The monitored area starts with the driver's shoulder and extends to about 3 m beyond the rear of the car.

The RADAR sensors of the "Blind Spot Assist" must be switched off in certain countries and in the vicinity of radio astronomy facilities. This is due to the restricted radio certification of 24 GHz radars with wide-band transmission used for automotive applications.

Based on the "Blind Spot Assist" introduced in 2007 in the S-class , the "Active Blind Spot Assist" with corrective braking intervention was first presented by

Mercedes-Benz in 2010 (Mercedes 2010). Currently, Mercedes offers this system in a package along with other driver assistance systems in the C, CL, GL, GLK, M, S, and SL classes for an additional charge of €2,677.50. As a separate system, the "Blind Spot Assist" is available as an option without corrective braking intervention for €535.50 in the A, B, C, CLA, CLS, E, and GLK classes, for €1,082.90 in the G class, and for €650.00 in the SLS AMG.

4.5 Audi Side Assist

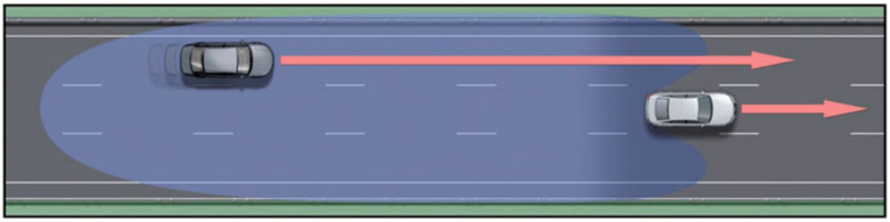
The "Audi Side Assist" informs the driver both about vehicles in the blind spot and vehicles which are approaching quickly from behind. This information is provided both on the driver and passenger sides.

The driver information is provided by lights integrated into the housing of the left and right exterior mirror. The light illuminates on the relevant side if the lane change appears potentially critical based on the situation detected by the system (Fig. 7a, b). This level 1 information is subtle, i.e., the driver only notices it by looking directly into the mirror. This means the driver is able to be continually aware of the system's features, even in situations that are not dangerous and without the lights causing interference or distraction for the driver. Level 2 information is activated if the driver operates the turn signal. In this case, the driver is informed of the hazard involved in changing a lane by the light flashing brightly several times (Fig. 7c, d). If the turn signal is activated continuously, then detected vehicles are indicated by the light remaining continuously lit, and there is no continuous flashing. Additional documents concerning the HMI of "Audi Side Assist" can be found under (Vukotich et al. 2008).

The "Audi Side Assist" is based on two 24 GHz RADAR sensors with narrow-band transmission that are integrated behind the left and right corners of the rear bumper, making them invisible from the outside. The rear-facing RADAR sensors of the latest generation have a range of 70–100 m. This also allows the driver to be informed in advance of vehicles rapidly approaching from behind. The side areas to the left and right of the vehicle are scanned by a distinctive and specially developed side lobe of the RADAR sensors. This means information can also be provided about vehicles in the blind spot. The current "Audi Side Assist" systems can be used to their full extent at speeds above 30 km/h. Further details on the "Audi Side Assist" can be found under (Popken 2006).

The RADAR sensors of "Audi Side Assist" are referred to as narrow-band systems within the specifications of the ISM band between 24.000 and 24.250 GHz. The output of maximum 20 dBm EIRP conforms to the European standard EN 300 440. A special revision of the radio certification regulations is not required for these RADAR sensors. They are not subject to the restrictions imposed on 24 GHz radars with wide-band transmission and do not have to be switched off in the vicinity of radio astronomy facilities.

The "Audi Side Assist" was offered for the first time in 2005 in the Audi Q7. Today, the system is available in almost all Audi models. The additional charge for a

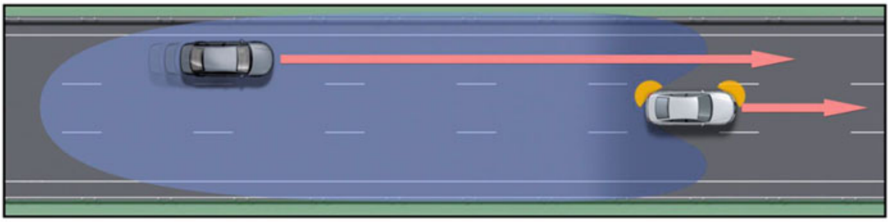


Approaching vehicle would be critical in case of a lane change

b

Continuously lit yellow
indicator light

c



Turn signal activated and critically approaching vehicle

d

Bright yellow indicator
light flashing briefly

Fig. 7 “Audi Side Assist” (Audi 2008) (**a**) and (**c**) detection range, (**b**) and (**d**) integration of the indicator light into the exterior mirror housing, (**a**) and (**b**) continuous activation of the *yellow* indicator light if the lane change is critical, (**c**) and (**d**) brief, bright flashing of the *yellow* indicator light if the turn signal is activated and the lane change is critical this system as an optional extra is currently €500 in the Audi A3, A6, A7, and Q3; €550 in the Audi A4, RS4, A5, and Q5; €600 in the Audi Q7; and €800 in the Audi A8 in the package with the “Audi pre-sense rear.”

Similar systems to the “Audi System Assist” are used by Porsche and VW. Similar sensors to those used by Audi are also used by BMW, Mazda, and Volvo.

4.6 “Side Assist Plus” from VW

The “Side Assist Plus” enhances the functions of the “Audi Side Assist” with the aid of a camera-based driver assistance system for lateral vehicle guidance (lane assist). The third warning level is the principal innovation of the “Side Assist Plus”: If the system detects, for example, the intention of the driver to change a lane in spite of traffic approaching from the rear, then a steering intervention is initiated, supplemented by a slight vibration of the steering wheel and flashing lights in the exterior mirrors. Here, the driver is always able to override the steering intervention.

The “Side Assist Plus” was first introduced in the Passat in 2011. In a package with the “Lane Assist,” the additional cost for the optional equipment in a CC or Passat is €1,100. The “Side Assist” without steering intervention on these vehicles costs €550. In a Touareg or Phaeton, the price for the “Side Assist” is €605 or €610.

This combination of a type III system with a three-level driver formation is so far unique and is not offered currently by any other vehicle manufacturer.

4.7 Commercial Vehicles

Lane change assistance systems have already begun to appear in the light commercial vehicle segment. VW Commercial Vehicles offer the “Side Assist” as part of a package for the models Caravelle, Multivan, California, and Transporter. The prices for these are between €952 and €1,154. Mercedes offers the “Blind Spot Assistant” in an equipment package for the Sprinter for €1,178.

Lane change assistance systems are currently rare in the segment of heavy-duty vehicles, even though there are considerably larger areas which are hardly visible to the driver in comparison with cars. The reason for this is probably the special requirement for heavy-duty vehicles that the sensors should only be installed on the tractor unit so that the system will still function when the trailer is changed. Consequently, the sensors should be installed on the tractor unit where possible. They must be installed in a position with a direct view of the side area to be monitored. Furthermore, there is currently no standard for these types of systems for the heavy-duty vehicle segment.

Nevertheless, first systems that cover a part of the side area are now available. Volvo currently offers a system that monitors the blind spot on the passenger side under the name “Lane Changing Support” (Fig. 8b). Volvo uses a 24 GHz RADAR

Fig. 8 "Lane Changing Support" from Volvo Trucks (Volvo 2011). (**a**) LED light on the A-pillar near the grabpole. (**b**) Monitored area on passenger side

sensor with a large beam width which is mounted above the wheel housing and beneath the storage compartment on the passenger side of the vehicle. The system informs the driver at speeds above 35 km/h via a light in the A-pillar on the passenger side if the monitored area is occupied, and the turn signal lever is operated (Fig. 8a). An acoustic warning can also be selected by the driver.

Currently, there are no systems in mass production for heavy-duty vehicles that completely monitor the left and right sides of the tractor unit. However, an example of how to implement such a system can be seen in a Scania research project (Degerman et al. 2012; Meinecke et al. 2013). Both side areas are completely monitored by mounting RADAR sensors in the left and right sides of the cab and an additional RADAR sensor in each exterior mirror of the truck (Fig. 9a). The rear-facing sensors in the exterior mirrors have a small beam width to achieve a long range at the same power output, whereas the laterally oriented RADAR sensors have a very large beam width (Fig. 9 middle).

Fig. 9 Prototype from Scania: (**a**) Mirror with additional RADAR sensor. (**b**) Zone-based warning system with three LED lights (Degerman et al. 2012)

A zone-based warning concept supports the driver to assess the position of vehicles at the side areas of the truck which are hardly visible to the driver even with the help of exterior mirrors. A total of three zones are defined, each represented with an LED light in the mirror or A-pillar (Fig. 9 right). The driver therefore always has information about in which of the three zones other road users are located.

A three-level information system was implemented in the prototype. Level 1 information is provided subtly with the aid of LED lights. Operating the turn signal lever despite the fact that the side area is occupied activates level 2 information. If the driver nevertheless initiates a lane change maneuver, the intensity of the LED lights is increased, and a corrective intervention to the lateral vehicle guidance takes place via a moderate steering torque which returns the vehicle to the original lane. However, the steering intervention can always be overridden by the driver.

5 System Evaluation

The performance of lane change assistance systems is significantly influenced by the choice of environment sensors. Class B RADAR sensors which monitor the adjacent lane beside the subject vehicle as well as up to 100 m behind the subject vehicle allow the detection of other vehicles in the blind spot, as well as fast approaching vehicles from the rear (type III system). Class A RADAR sensors as well as camera and ultrasonic-based systems can only detect vehicles in the blind spot of the subject vehicle (type I system). Consequently, the customer benefit with type III systems is higher than with type I systems. A benchmark test by Auto Bild came to the same conclusion (Auto Bild 2013). All the top places were occupied by type III systems with assessment scores of between 14.0 and 17.5 out of a possible 20 points. The scores for type I systems were between 8.0 and 10.5 points.

It can be assumed that RADAR-based systems offer greater availability and robustness than camera- or ultrasonic-based systems. Other road users can be reliably detected in poor weather conditions, such as rain, spray, or fog.

The activation speed of camera-, RADAR- (type A), and ultrasonic-based systems of type I is usually 10 km/h. RADAR-based (class B) type III systems mostly have an activation speed of 30 km/h. This is an advantage for type I systems as they can also be used at very low speeds, e.g., in stop and go traffic in urban areas or on motorways. This did not lead to a higher ranking of type I systems in comparison with type III systems in the Auto Bild benchmark test. Also, from the accident statistics in Fig. 1, one must assume that a greater sensor range is preferable to a lower activation speed.

The customer benefit of low-cost ultrasonic-based systems is also limited by the fact that they are mostly only available up to a vehicle speed of approx. 130 km/h.

Furthermore, the customer benefit is largely decided by the way and manner the driver information is provided. It should be provided in a subtle and unobtrusive manner as long as no lane change into an occupied adjacent lane is intended or indicated. If this occurs, the driver information should be intensified. This criterion can, in principle, only be fulfilled by systems using two- or three-level driver information. However, current systems with a single-level driver information system are limited to subtle visual information. It is probably for this reason that the Auto Bild benchmark test criticized that the level 1 driver information in all type I systems was not easily noticeable.

Lamps in or near the exterior mirror are now the established warning for level 1 driver information systems. The location of the lamps varies between the various manufacturers: A-pillar (Volvo), triangular mirror cover (Infiniti, Mazda), mirror housing of exterior mirror (Audi, BMW, Porsche, VW), and outer edge of the mirror glass (Citroën, Ford, GM, Jaguar, Jeep, Land Rover, Mercedes, Opel). Also, the detailed design of the indicator lamps varies between the various vehicle manufacturers between red and yellow indicator lamps that are configured as illuminated dots, pictograms, or surfaces. Through the position, size, intensity, and color of the lamps, the vehicle manufacturers are ultimately trying to provide suitable level 1 driver information for the respective system as they see fit.

Illuminated dots and pictograms located behind the mirror glass of the exterior mirror were classed as “too weak,” “hard to identify,” or “moderately identifiable” in the Auto Bild benchmark test. Furthermore, it was feared that they would not be visible with a bright background or due to the headlights of other vehicles. However, large lamps where the brightness can be adjusted by the driver were praised.

Lamps in or near the exterior mirror are also the established warning for level 2 driver information systems. Almost all vehicle manufacturers that offer two-level systems inform the driver through bright multiflashing sequences of these lamps. Some vehicle manufacturers also feature an acoustic warning (Mazda, Mercedes) or a vibration in the steering wheel (BMW). Whether it makes sense to also provide an acoustic or haptic warning alongside the visual warning in the second level should be left for the experts to decide. However, the increased benefit to the customer of multilevel driver information is obvious.

The essential difference with level 3 driver information is an intervention into the lateral vehicle guidance as soon as the vehicle begins to move into an occupied adjacent lane. This can be realized through a corrective braking intervention, where individual wheels are braked, resulting in a momentum that returns the vehicle to the original lane (Infiniti, Mercedes) or through a steering intervention by the electromechanical steering system (VW). Advantageous is the avoidance of an immediate critical situation. Disadvantageous is the requirement for an additional sensor to monitor lane markings, which is required for the technical realization of such a system and which in general at least doubles the cost of the system. This was also a criticism in the Auto Bild benchmark test.

Lane departure warning systems are hardly used in the heavy-duty vehicle segment. The specific requirements of the heavy-duty vehicle segment must be taken into account when designing such a system. However, the basic criterion for the design of the sensors and driver information is likely to be similar to those of cars.

6 Achieved Performance

The performance capability of the aforementioned lane change assistance has already reached significant levels. However, all of these systems have their limits, and the vehicle manufacturers need to inform drivers of these limits, for example, in the owner’s manual.

Almost all vehicle manufacturers are united in pointing out that their system is only an assistant, possibly does not detect all vehicles, and cannot replace attentiveness by the driver. Furthermore, all vehicle manufacturers point out that vehicles may not be detected adequately or, under certain circumstances, may not be detected at all if the sensors are soiled or during adverse weather conditions, such as rain, snow, or heavy spray.

Vehicles with 24 GHz radars installed in their rear bumper cannot use the systems if the sensors' field of view is blocked by objects such as bicycle carriers, trailers, or stickers.

In the case of "Audi Side Assist" and VW "Side Assist," it has been pointed out that the driver cannot be informed in sufficient time about vehicles approaching at very high speed. No information is provided to the driver on tight bends with a curve radius less than 200 m.

Furthermore, there may be errors when assigning lanes to vehicles, because the widths of adjacent lanes cannot be measured but are only estimated. Therefore, it has been pointed out that very wide lanes combined with a driving style which places the vehicles at the outer edge of their respective lanes may result in information about vehicles in adjacent lanes failing to be provided. When driving in narrow lanes combined with a driving style that places the vehicles on the inside edge of their respective lane, there may well be unnecessary instances of information being provided to the driver about vehicles in the next lane but one.

These errors with lane assignment can to some extent be prevented with addition of a system for lane detection which is mostly camera-based. Currently, these systems have many limitations, e.g., the need for lane markings and the need for them to be visible. The ability of the system to detect lane markings can be impaired, for example, by road soiling or soiled sensors and also by fog or heavy rain.

Therefore, it is apparent that the aforementioned lane change assistance systems could be improved further from a user's perspective.

7 Further Developments

The system functions of lane change assistants can be improved further by increasing the performance of environment sensors, for example, by increasing sensor ranges and the speed range within which the sensors can be operated reliably.

As explained in Sect. 6, errors when assigning lanes to other vehicles can lead either to driver information being provided superfluously or the information not being provided when it ought to be. This can be prevented when the system can detect the position of the adjacent lane as well as the position of the target vehicle. Currently, camera-based systems are mostly used for the detection of lane markings. These systems have various restrictions which effect their availability and robustness.

It is therefore desirable to create a lane model that is based on the data received from the camera as well as the data from other environment sensors. Therefore, in addition to the lane markings, other features can be used to support the modelling of the lane, such as the position of guardrails, the transition between the road and the median strip/grass verges, the driveway of other vehicles, as well as map data with information such as bends and the number and width of lanes.

At present, lane change assistants only help the driver to decide whether a lane change is possible or not. The lane change itself must be performed independently by the driver. However, this maneuver could also be assisted with the aid of environment sensors that detect the entire adjacent lanes before, next to, and behind the subject vehicle and the previously described model of the lane, including improved odometry estimates in combination with an electronically controllable steering actuator system. Lateral vehicle guidance when changing lanes could be supported (Habenicht 2012) or even automated with the aid of suitable steering torque.

References

Audi Q7 and VW Touareg owner's manuals (2008)

Auto Bild No 11 (2013) Was taugen Totwinkelwarner? (How useful are blind spot alerter?) see also http://www.autobild.de/artikel/assistenzsysteme-test-3920250.html. Accessed 20 Oct 2014

Citroën homepage (2013) Toter Winkel Assistent (Blind spot assist). http://www.citroen.de/technologie/fahrassistenzsysteme/toter-winkel-assistent/#/technologie/fahrassistenzsysteme/toter-winkel-assistent/. Accessed 20 Oct 2014

Degerman P, Ah-King J, Nyström T, Meinecke M-M, Steinmeyer S (2012) Targeting lane-change accidents for heavy vehicles. VDI/VW conference on driver assistance systems and integrated safety, Wolfsburg, VDI Verlag GmbH, pp 167–175

GIDAS (German In-Depth Accident Study) and accident database of Volkswagen Accident Research (2008)

Gizmag Internet magazine (2007) Volvo launches Blind Spot Information System (BLIS). http://www.gizmag.com/go/2937/. Accessed 20 Oct 2014

Habenicht S (2012) Entwicklung und Evaluation eines manöverbasierten Fahrstreifenwechselassistenten (development and evaluation of a maneuver based lane change assistance). Dissertation, Technical University of Darmstadt

Heise Online (2007) Mercedes: Neuer Totwinkel-Assistent (Mercedes: new blind spot assist). http://www.heise.de/autos/artikel/Mercedes-Neuer-Totwinkel-Assistent-421304.html. Accessed 20 Oct 2014

Heise Online (2010) Der neue Citroën DS4: Bilder und Details (The new Citroën DS4: pictures and details). http://www.heise.de/autos/artikel/Der-neue-Citroen-DS4-Bilder-und-Details-1069445.html. Accessed 20 Oct 2014

IndianCarsBikes (2010) Blind spots? Call Ford for help. http://www.indiancarsbikes.in/automotive-technology/ford-blind-spot-information-system-2010-uk-cars-s-max-galaxy-7751/. Accessed 20 Oct 2014

ISO standard 17387 (2008) Lane change decision aid system, ISO, Switzerland

Meinecke M-M, Steinmeyer S, Ah-King J, Degerman P, Nyström T, Deeg C, Mende R (2013) Experiences with a radar-based side assist for heavy vehicles. International Radar Symposium, Dresden, IEEE, pp 720–725

Mercedes-Benz press release (2010) Neue Fahrer-Assistenzsysteme – Premiere: Aktiver Totwinkel-Assistent und Aktiver Spurhalte-Assistent mit Bremseingriff (New driver assistance systems – premiere: active blind spot assist and active lane keeping assist with braking intervention). http://media.daimler.com/dcmedia/0-921-658892-49-1298797-1-0-0-0-0-1-11702-1549054-0-1-0-0-0-0-0.html. Accessed 20 Oct 2014

Popken M (2006) Audi side assist. In: Hanser automotive electronics + systems, Carl Hanser Verlag, Munich, 7–8, pp 54–56

Prova Internet magazine (2007) Jaguar Toter-Winkel-Überwachung: Das Auto blickt zurück (Jaguar blind spot monitoring: the vehicle looks back). http://www.prova.de/archiv/2007/00-artikel/0141-jaguar-toter-winkel/index.shtml. Accessed 20 Oct 2014

Volvo Cars Press Release (2012) The all-new Volvo V40 – safety and support: the most IntelliSafe Volvo model ever. https://www.media.volvocars.com/global/en-gb/media/pressreleases/42212. Accessed 20 Oct 2014

Volvo Truck Product Information (2011) Spurwechselunterstützung (Lane change support). http://productinfo.vtc.volvo.se/files/pdf/hi/LCS_Ger_02_880763.pdf. Accessed 20 Oct 2014

Vukotich A, Popken M, Rosenow A, Lübcke M (2008) Fahrerassistenzsysteme (Driver assistance systems) ATZ and MTZ 2 (special edition), pp 170–173

Intersection Assistance

50

Mark Mages, Felix Klanner, and Alexander Stoff

Contents

This publication is based on work done as a research assistant at the Institute of Automotive Engineering, Technische Universität Darmstadt, Germany.

M. Mages (✉)
Chassis & Safety, Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: mark.mages@continental-corporation.com

F. Klanner
BMW Forschung und Technik, München, Germany
e-mail: felix.klanner@bmw.de

A. Stoff
Fachgebiet Fahrzeugtechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: alexander.stoff@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_51

Abstract

Junctions and intersections are relatively complex areas in today's vehicle traffic. Due to this complexity, many traffic accidents occur between vehicles interacting in these areas. Each driver needs to assess parameters such as intersection geometry, right-of-way regulation, and the behavior of third-party vehicles, and a driver fault in any of these tasks can consequently lead to a collision. Thus, an intersection assistance system needs to consider different types of potential driver mistakes in these complex scenarios. This makes the avoidance of intersection collisions particularly challenging.

This chapter summarizes major challenges for assistance systems that address the most relevant accident types in intersections, and it introduces different approaches to solve the resulting issues. These approaches vary from intersection-based driver information to cooperative active safety systems using communication technology and/or data from onboard sensors such as RADAR, camera, or LIDAR. Furthermore, the importance of the capabilities of different sensors as well as communication technology for onboard intersection assistance systems is discussed.

The early recognition of driver mistakes and the design of a suitable intervention strategy require detailed analyses of common driver behavior in intersections. The results of respective driver behavior analysis are introduced. Different example prototype systems are described in detail, including criticality assessment and warning/intervention strategies. Validation tests with real test persons for some of the discussed systems prove that a positive driver acceptance for intersection assistance systems is feasible.

To maximize the overall benefit of intersection assistants during the period of marked introduction, when fleet penetration rates are low, the possibilities and limitations of assistance for collision avoidance in the prioritized vehicle are discussed.

1 Accidents at Intersections

One of the main causes of accidents in road traffic is the road user's misconduct in intersections and junctions. In 2012, for example, 42 % of all accidents with serious property damage, 37 % of all accidents with personnel injury, and even 19 % of all fatal accidents occurred on intersection-related accident types – turning into a road (accident type 2 according to Institut für Straßenverkehr (1998)) or crossing a road (accident type 3) (Statistisches Bundesamt 2010). That is the reason why intersections are in the focus of research from both a safety and engineering point of view. Based on a detailed accident data analysis out of the GIDAS database, the major causes for accidents in intersections could be identified (Hoppe et al. 2006; Intersafe 2005). Frequently occurring errors are:

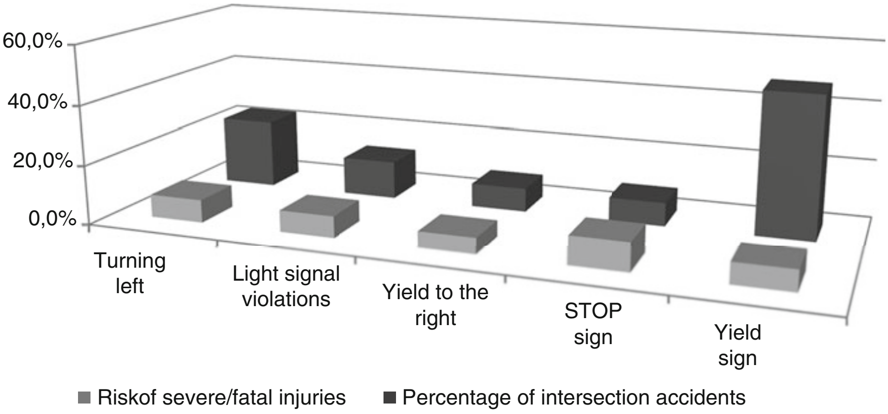


Fig. 1 Accidents at intersections (GIDAS 2010)

(a) Misinterpretation, i.e., the situation is perceived but misinterpreted by the driver. A typical example is misjudgment of the velocity of prioritized traffic or false interpretation of local traffic regulations.
(b) Inattention, i.e., distraction from the actual driving task that leads to prolonged reaction times.
(c) Lack of attention concerning possible obstructions on intersections. This may be cause by vehicle-related obstruction like a wide A-pillar which conceals bicycles/motorcycles or by external obstructions like parking vehicles or buildings. In addition, oncoming vehicles turning left may temporarily obscure the driver's view.

Number and severity of intersection accidents are also depending on traffic regulations in the intersection as represented in Fig. 1.

2 Intersection Assistance Systems

When approaching an intersection, there is one point when the driver needs to make a decision on whether to enter the intersection or to stop in front of it. This decision is based on multiple information including traffic rules and third-party vehicles. A critical situation with high risk of a collision results from a driver not being able to either receive or evaluate this information correctly. To avoid such situations, it is imperative to support the driver by predictive assistance systems with regard to situation interpretation and collision avoidance.

One major challenge for intersection assistance systems is the large number of potentially critical situations that might lead to an accident. Intersection accidents can be classified with respect to the type of accident and the current traffic regulations, shown in Fig. 1. The assistance systems introduced in this chapter address these categories accordingly. Different systems and suggestions for respective warning strategies are illustrated in Fig. 2:

- Top: “Stop” Sign Assistant (driver information and warning cascade)
- Center: Intersection assistant in give-way situations (driver information and warning cascade)
- Bottom left: Traffic Light Assistant with speed advice for the green light phase
- Bottom right: Left Turn Assistant

2.1 Stop Sign Assistant

A Stop Sign Assistant supports the driver while approaching an intersection. The objective is to avoid an inadvertently stop line crossing (stop sign, sign number 206 of German road traffic regulations – Stop! Yield!). If a driver fails to yield after stopping at a stop line or sight line, this is not addressed by a Stop Sign Assistant because sensor requirements to avoid a collision in such a situation are more challenging. The Cooperative Intersection Collision Avoidance System (Sect. 2.3) supports the driver for the purpose of collision avoidance in this situation. At this point it is important to indicate that a retrospective distinction between the two situation categories, for example, by the analysis of accident databases, is not always obvious. Therefore, additional analyses, e.g., accident reconstruction studies, may be necessary to classify the relevance of these two situations correctly.

In order to avoid inadvertently stop line crossing, it is not required to take into account other traffic participants as any driver must stop the vehicle at the stop line according to road traffic regulations. This regulation does not depend on periodic priority changes in contrast to traffic lights. In addition the vehicle’s stop position is clearly specified at the stop line as contrasted with, e.g., left turn situations. Hence, the Stop Sign Assistant addresses intersection scenarios with a relatively low complexity in comparison to situations including traffic light or a left turn accident.

A two-stage Stop Sign Assistance is suggested in Meitinger et al. (2004). The prototype system includes early information concerning the stop sign at the next intersection and combines this with a driver warning if necessary. This is a suitable compromise between expected effectiveness and required reliability. Furthermore, no unnecessary warnings are given to an attentive driver. The head-up display (HUD) is a convenient human-machine interface (HMI) for such a system because the driver is able to perceive both the environment and the information display without avert gaze.

One challenge of every warning system is to handle the conflict of objectives: warning in due time versus minimum number of false positives. This conflict is known as the warning dilemma. Consequently, in order to provide suitable Stop Sign Assistance with a warning that is both timely and acceptable, knowledge about traffic regulations and a detection of the driver intention (whether or not to stop at the stop line) is necessary.

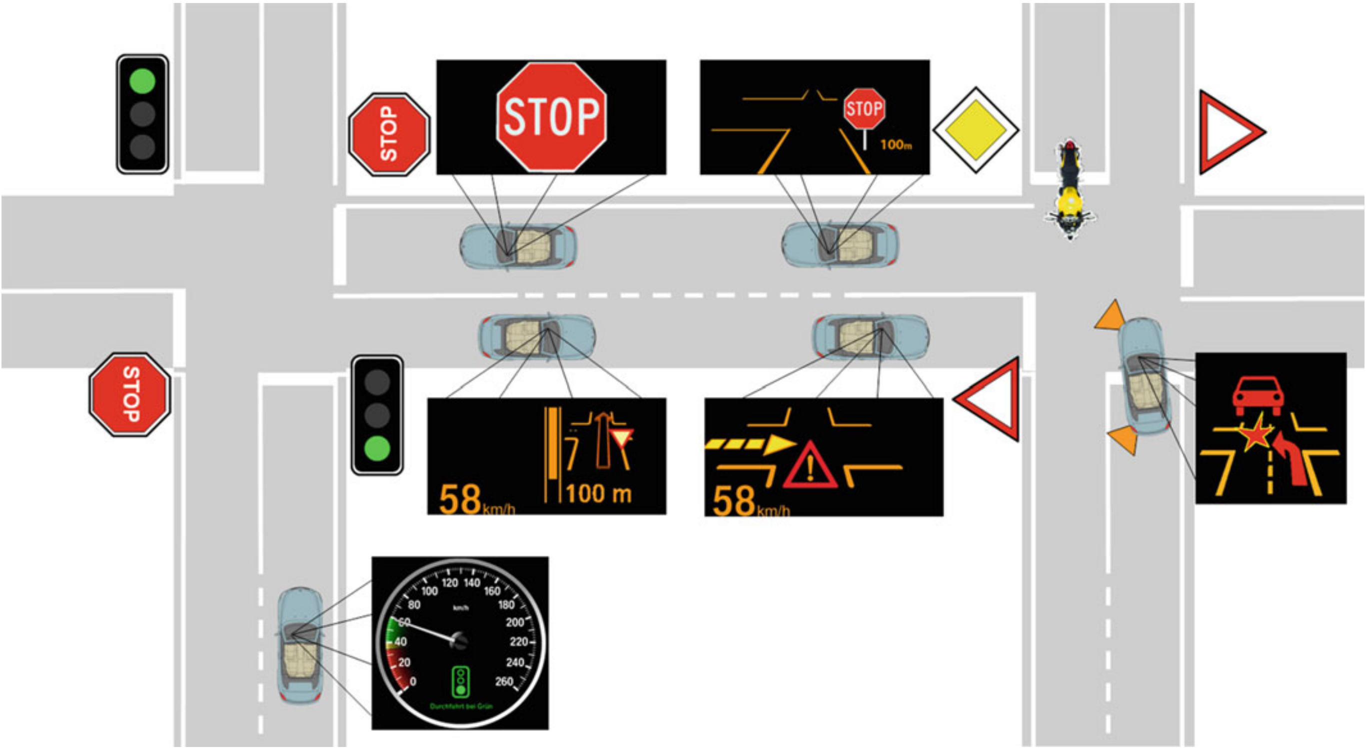


Fig. 2 Overview of different intersection assistance systems and intervention strategies

An additional challenge is the common driver behavior at intersections with stop signs. The road traffic regulations require a complete vehicle stop at the stop line. Nevertheless, a relevant number of drivers empirically do not come to a complete halt. If the traffic on the main road does not require a stop, in many cases only the vehicle speed is reduced. This circumstance would lead to a huge number of warnings that are in line with traffic laws but unnecessary with regard to safety aspects. To avoid annoyance of the driver in such situations, a Stop Sign Assistance system requires an additional distinction between inadvertently and intended stop line crossings.

2.2 Traffic Light Assistant

A Traffic Light Assistant supports the driver while approaching and stopping at an intersection with traffic light regulation. Accident database analyses show the following accident distribution at intersections with traffic light regulations: 2/5 of these accidents are collisions with a vehicle, “which turns into or crosses” (official type of accident number, 5) and 1/4 of these accidents are a collision with a vehicle, “which drives ahead or waits” (official type of accident number, 2) (Meitinger et al. 2004). The reason for the first type of accident (“cross-traffic collision”) is mainly a failed stop at a red traffic light. The reason for the second type of accident (“rear-end collision”) is mainly that different drivers assess the possibility differently to cross the stop line before the traffic light switches to red. It is a common objective to launch different measures to avoid accidents because of a failed stop at a red traffic light. It has to be considered though that systems that avoid a failed stop at red traffic lights – while reducing the number of cross-traffic collisions – may simultaneously have the opposing effect on the number of rear-end collisions (Garber et al. 2007).

In order to enhance traffic safety as well as traffic flow, traffic lights provide different useful measures:

- Specific phase change signals help reduce the criticality of surprising phase changes. Examples are counters next to the green and red light in Singapore and Taiwan. These counters signal the seconds until the next phase change. A different example is Austria, where the green light starts to flash seconds before a phase change. As a result, the driver gets more time to decide whether to cross or to stop at the stop line. The main advantage of the additional phase change signal is that it helps to avoid rear-end collisions. The disadvantage however is that the driver has to process additional data (in case of the counter also an additional source of information) which may lead to a reduced vehicle throughput at the intersection.
- Color-coded sections with a defined distance to the stop line that shall help the driver to decide whether to stop or to cross. This approach is tested in the United States (Yan et al. 2005). If the vehicle has passed the color-coded section with design speed before the traffic light switches to yellow, it will cross the stop line

in time before the red phase will start (assuming the vehicle speed remains constant), otherwise it will not. This static method requires less investment in infrastructure than the specific phase change signal. The downside is that the reliability of this information decreases if there is a significant deviation between the vehicle's velocity and the design speed. Furthermore, the yellow phase length needs to be fixed.

- Assistance systems which help to avoid a failed stop at red traffic lights and also support the driver while approaching traffic lights. Such systems require additional information concerning the traffic light. Still, no information about third-party traffic participants is necessary (similar to the Stop Sign Assistant). This type of Traffic Light Assistant uses static parameters like stop line position, the traffic light's working condition, and the current phase as well as information about the next phase change. All these information can be provided via wireless infrastructure to vehicle communication (Hopstock et al. 2007; Kosch and Ehmanns 2006). Systems that rely on onboard vehicle sensors (i.e., without communication technology) can at best guess the next phase change and may also have difficulties in observing the current traffic light status if the vehicle is standing close to the traffic light due to limited vertical field of view of current camera sensors.

If this information is available, the method to prevent a vehicle from running a red traffic light is comparable to the Stop Sign Assistance system already discussed. In addition, the vehicle shall be prevented from entering the intersection for the full duration of the red light phase of the traffic light. The system therefore includes a warning and/or braking intervention to prevent stop line crossing during a red phase after stopping. Situations such as this occur especially frequent when lanes of the same driving direction have different light signal phases.

Besides the safety enhancement aspects, Traffic Light Assistance may also provide comfort and economic benefits while approaching the intersection. For example, the Traffic Light Assistant can in advance estimate the expected signal phase for the moment the vehicle reaches the stop line as well as the next phase change. Based on this information the assistance system may then calculate a speed recommendation within the speed limit, allowing the driver to optimize the approach concerning safety, comfort, and efficiency.

It is known (Intersafe 2005) that when approaching a green traffic light without vehicles ahead, respectively, 1/3 of all drivers:

- Increase vehicle speed, in order to enhance the chance to reach the stop line during this green phase
- Decrease vehicle speed, in order to gain time for a decision in case the color switches
- Does not change the vehicle's speed

An increased vehicle speed is considered critical regarding traffic safety aspects and also unfavorable concerning fuel consumption. Decreased vehicle speed is unfavorable as well due to the limited capacity of urban roads, especially in areas of high population density. Thus, supporting the driver in optimizing the vehicle's speed while approaching a traffic light is beneficial in many regards.

The mentioned approaches only give an overview of different functions that Traffic Light Assistant systems can provide. It is important to highlight that assistance functions which base on infrastructure address safety and comfort aspects and also provide measures to solve challenges concerning traffic flow, energy consumption, and emission reduction.

2.3 Cooperative Intersection Collision Avoidance System

Cooperative Intersection Collision Avoidance Systems support drivers when their intention is to turn into or cross major roads. The objective is to avoid accidents at intersections with yield signs (yield sign, sign number 205 of German road traffic regulations – Yield!) and with right of way from the right.

Turn-into and crossing accidents are subdivided into collisions with and without stop at the stop line. As stop lines are not always painted, this also refers to a virtual stop line, i.e., the sight line of the intersection. A fundamental difference between Cooperative Intersection Collision Avoidance on the one hand and Stop Sign and Traffic Light Assistance on the other is the relevance of third-party vehicles: The driver only has to yield and stop the vehicle at the stop line if there are vehicles on the main road which make such an action necessary. Otherwise, the intersection can be crossed without stopping, in some cases without even slowing down. Hence it is required to provide the position and driving dynamic data of the vehicle on the side road as well as information about traffic participants on the main road as the general basis of such an assistance system.

Possibilities to realize Cooperative Intersection Collision Avoidance are:

1. Assistance with intelligence in infrastructure:
   In order to support the waiting drivers on the side road to turn into or cross the main road safely, the Rural Intersection Decision Support can be used (Donath et al. 2007). RADAR sensor systems capture the vehicles' position and speed on the main road and provide the data to the assistance system. Time gaps between vehicles on the main road and the vehicles' arrival times at the intersection are estimated. Waiting vehicles at the stop line are captured and classified by a video camera system. Based on this information it is decided whether it is safe to turn into or cross the intersection. If this is not the case a corresponding warning is issued to the driver. As a possible human-machine interface (HMI) concept, a common stop sign is extended by an additional risk warning.
2. Assistance with intelligence in vehicles:
   To support approaching as well as waiting drivers on the side road to turn into or cross the main road safely, several approaches have been published (Hopstock 2007; Klanner 2008; Mages 2009). A main difference between these approaches is the technology used to collect the required data. Some systems depend only on

vehicle-based environment perception sensors. Examples for such prototype systems work based on laser scanners, video cameras, RADAR, and a high-precision digital map (Intersafe 2005). Other systems gain data from relevant third-party vehicles via wireless vehicle-to-vehicle communication (Klanner 2008) which in return requires a suitable communication solution as well as an estimation of current position and driving dynamic data of all relevant vehicles. The positioning reference system is usually a Global Navigation Satellite System (GNSS) and the Global Positioning System's (GPS) time stamp. In addition, any active safety system for a crossing and turn-into assistant requires an early identification of potential collisions and a criticality estimation.

The decision-making process concerning assistance measures can be divided into two subtasks (Mages 2009):

(a) Estimation of the criticality of an imminent collision
(b) Calculation of the latest possible assistance point in time to avoid an entrance of the host vehicle into the collision zone

One approach for situation interpretation approaches is explained in more detail in Sect. 4. If collision-avoiding measures are necessary to support drivers to avoid critical traffic situations, different information and warning levels as well as automatic emergency maneuvers are conceivable. First, a driver information regarding the traffic regulation at the next intersection is issued. If vehicle speeds are highly critical, the criticality of traffic situations can be identified in an especially early stage while approaching the intersection. Nevertheless, it is difficult to decide whether assistance is necessary because the driver intention has to be taken into account as well. Hence, it is recommended to wait until the latest possible warning point. Suitable assistance measures are a combination of a warning symbol (e.g., in the head-up display), an audible warning signal, and an active deceleration/brake jerk during the expected reaction time of the driver (Hopstock 2007). The active deceleration of a brake jerk may reduce the time required by the driver for a suitable reaction due to haptic feedback and will also increase the available time for the driver to react (Mages 2009). Thus, using a brake jerk as part of the warning strategy is considered beneficial in multiple ways to stop the vehicle in front of the collision zone.

If the vehicle comes to a halt at the stop line and the driver subsequently intends to enter the intersection while there is crossing traffic on the major road (to which the driver should yield), a driver warning is not a sufficient countermeasure. The reason is that there is not enough time for the driver to react to a warning and stop the vehicle after setting off, since the vehicle is already very close to the potential collision zone. In this case, an active full brake intervention that prevents the driver from starting into the intersection in the first place seems to be the only possible measure to avoid a collision. Whether or not such a rigorous intervention is appropriate and accepted by the majority of drivers depends on the actual collision imminence and on the reliability and failure resistance of all data used to assess this collision risk. In order to keep the expected number of false-positive warnings as low as possible, it is necessary to adapt warning criteria to typical driver behavior (Sect. 4).

One additional challenge results from the necessity to distinguish between the driver's intention to turn into and to cross an intersection while the vehicle is still approaching the intersection. This is essential to estimate the future path of the vehicle as there is no collision thread if the path of the ego vehicle does not coincide with the path of another traffic participant on the main road. Also, the typical velocity profile while approaching an intersection is influenced by the maneuver, which in return needs to be considered in the system's assessment on whether or not the driver intends to stop prior to the intersection. Since drivers do not always make use of the turn signal, an assistance system should not rely on this indicator alone and include additional factors to estimate the driver's intention (Mages 2009).

3. Assistance with intelligence in vehicles and sensors in infrastructure:
   Combining parts of the introduced ideas leads to a different approach: A combination of infrastructure-based environment perception and vehicle-based situation interpretation as well as assistance. This can be beneficial if there are intersections with a considerably high number of accidents (Suzuki et al. 2007). At these intersections, sensor-based perception systems are used to perceive the intersection's transport scene and to provide this information to traffic participants via wireless communication. To receive this environment perception results in a vehicle, the vehicle has to be equipped with an appropriate communication module. The main benefit of the approach is that environment perception results comprise information concerning vehicles with and without communication modules (in contrast to intersection collision avoidance systems which only use vehicle-to-vehicle communication). As a result of this, equipping intersections that stand out as accident hotspots with such a system can increase traffic safety even in an early phase of vehicle-to-X communication, where the penetration rate of communication modules in vehicles is low. The downside is the necessary investment in both traffic infrastructure and vehicles.

2.4 Left Turn Assistant

The task of a Left Turn Assistant is to support the driver to avoid collisions with oncoming traffic while turning left (the wording in this chapter is assuming right-hand traffic). The official type of accident "turn collision" comprises a huge number of such situations, some of which include collisions with crossing traffic. In the focus of a Left Turn Assistant as described in this chapter are only collisions with oncoming traffic. Avoiding collisions with crossing traffic while turning left is – from a system's point of view – very similar to avoidance of collisions while turning right, as introduced in Sect. 2.3.

Different studies show (Hoppe et al. 2007; Pierowicz et al. 2000) that main reasons for collisions while turning are misjudgment of oncoming traffic's distance and speed, driver failure to perceive third-party vehicles (especially bicyclists and powered two-wheelers), and an obstructed view caused by oncoming turning traffic. Turning left is particularly challenging to the driver because of the maneuver's complexity. For a driver assistance system, it is even more difficult to assess the criticality of a potential collision in turning situations since no clearly defined turning point exists. As a consequence, a huge number of trajectories have to be considered. This is why the analysis of driver behavior and the prediction of the turning intention are very important for a Left Turn Assistant.

In addition a Left Turn Assistance also requires an assistance strategy which prevents the left turning vehicle from entering a potential collision zone, i.e., the path of an oncoming vehicle. It is obvious that the later the driver shows the left turn intention, the later it is possible to detect his intention reliably and thus the less time is available for system intervention. As a consequence, if a vehicle is very close to or directly at the (virtual) center line, the possibilities to support the driver are very limited (Meitinger et al. 2006).

In left turn situations, similar to cross-traffic scenarios, the necessity of a vehicle stop at the intersection depends on third-party traffic members, in this case oncoming vehicles. To classify the different collisions, the accidents at stop signs as well as left turn accidents distinguish between accidents with and without a vehicle stop at the intersection. Corresponding number of accidents' analyses show that both scenarios are relevant for the accident occurrence at intersections (Chovan et al. 1994; Meitinger et al. 2006).

The following situations are addressed by a Left Turn Assistant as described in this chapter:

- The driver decelerates the vehicle to a complete halt within the intersection and waits for a suitable gap in oncoming traffic.
  The upside of this situation (from an assistance system's point of view) is that a left turn intention can be estimated with a reliability of almost 100 % (Branz and Öchsle 2005; Meitinger et al. 2006). While waiting, the driver can be provided with visual information concerning potentially critical oncoming traffic to assist him in choosing a suitable gap. A convenient HMI concept is to display this information on the HUD (head-up display). This way the driver can perceive both external and internal information. The downside of this situation (again from the perspective of an assistance system) is that the vehicle is very close to the conflict zone, i.e., there is no driver reaction time available to be able to stop the vehicle outside the potential conflict zone. As a consequence, a driver warning will not suffice to avoid a collision if the driver accelerates the vehicle from standstill directly at the (virtual) center line. In this case a start inhibit brake intervention may be the only feasible option to avoid an accident.
- The driver approaches an intersection with the intention to turn left and while approaching identifies a gap in oncoming traffic that (according to his evaluation) allows him to turn left without stopping. For this scenario the major challenge from a systems' point of view is to estimate whether and when a drivers intends to turn left. This situation interpretation can be done based on vehicle signals including steering wheel angle and steering wheel velocity (Meitinger et al. 2006). Informing or warning the driver about oncoming traffic in this situation is not considered effective as the available time is too short for a

driver reaction. Thus, an active emergency deceleration has to be considered to avoid such collisions. However, the driver still has to be able to overrule this intervention. There is no doubt that the reliability of environment perception, driver intention detection, and situation interpretation has to be high to avoid false-positive brake interventions.

To evaluate oncoming traffic in both scenarios, it is necessary to take the driver's usual assessment of gaps in traffic into account. A comparison of different studies shows that the assessment of the gap in oncoming traffic is influenced by a variety of intersection-based factors (Institut für Straßenverkehr 1998; Meitinger et al. 2006). As a consequence, gaps accepted by the driver for turning left vary between 4 and 14 s. If the prediction of the left turn uses a driver model, as described in Meitinger et al. (2006), the influence of intersection geometry concerning the driver behavior and additional intersection-specific information requirements has to be taken into account.

2.5 Intersection Assistant for Prioritized Vehicle

The intersection assistant systems introduced so far have in common that they address potentially critical situations caused by a violation of right-of-way regulations by informing/warning the driver or by brake intervention in the vehicle obliged to give way. Assuming all vehicles were equipped with such systems, a significant amount of right-of-way violations could be excluded, and thus many intersection accidents could be avoided. However, such a scenario cannot be assumed within medium term. The average age of the German vehicle fleet in 2012 for instance was about 8.5 years (Kraftfahrt-Bundesamt 2012) which means that even if all new vehicle types of today would be equipped with such systems, it would take a considerable amount of time to achieve a penetration rate even close to 100 %. Because of this, the benefit of the intersection assistance systems described earlier is limited to scenarios, in which the vehicle equipped with such a system is also the vehicle which is about to violate the right-of-way regulation. If however the tables are turned and the prioritized vehicle is equipped with an intersection assistant while the waiting duty vehicle is not equipped, there will be no benefit.

As with all safety measures, the total benefit correlates with market penetration. But unlike road departure accidents for instance, intersection collisions usually involve multiple vehicles. If a collision can also be avoided by an intersection assistance system fitted to the prioritized vehicle, this would highly increase the overall benefit of intersection assistants during the years of introduction, when market penetration is fairly low. For this reason it is indicated to also consider active safety measures for the prioritized vehicle in case of an imminent right-of-way violation by a third-party vehicle.

When addressing the vehicle with right of way, user acceptance has to be considered even more than for intersection assistance systems addressing the waiting duty vehicle. For the latter it has been shown that with a suitable system design including information, warning, and emergency braking, drivers accept an intersection assistant system's interventions as adequate if in retrospect it is obvious to the driver that his own fault led to the respective situation. However, from the point of view of the prioritized driver, the same situation with similar system intervention measures presents itself in a different way. In this case there is no misbehavior of the driver – the situation is caused by a fault of a third-party traffic participant. Because of this, acceptance for any type of intervention may be lower. Each intervention considered unnecessary or annoying by the driver (even though it may well be objectively correct) will adversely affect the acceptance of such functionalities and thus needs to be avoided with highest priority. Otherwise, the driver may be tempted to manually deactivate the system.

The subjectively perceived necessity of an intervention can only be assessed by the driver and mainly depends on the behavior of the potential collision object. If in a critical situation the driver of the waiting duty vehicle reacts late but in time, he will come to a standstill before entering the potential collision zone. In this scenario any intervention in the prioritized vehicle is neither objectively nor subjectively correct. Thus, intervention measures in the prioritized vehicle may only be issued after the potential collision object is unable to avoid a collision. For acceptance reasons, this also should be transparent to the driver. Only then an emergency brake intervention is judged as necessary measure by the driver in retrospect.

Considering these facts, appropriate safety measures for the prioritized vehicle might not only include emergency braking but also emergency evasion steering. In a crossing scenario for instance, the choice of a suitable strategy depends on the question, which of the participating vehicles is able to avoid the collision by braking latest regardless of the behavior of the other collision partner (this means avoiding the collision by coming to a standstill before entering the potential collision zone). If the prioritized vehicle can still avoid a collision by braking, while the potential collision partner cannot, emergency braking at the latest possible moment is suggested as intervention strategy.

If the waiting duty vehicle is the last collision partner able to avoid the collision, the situation presents itself in a different way. Now the prioritized vehicle is unable to avoid the collision by stopping prior to the collision zone (referred to as avoiding geometrically). Still it might be able to avoid the collision by decelerating, thus gaining time and entering the collision zone after the potential collision partner has already left this area (referred to as avoiding temporally). Avoiding a collision temporally however is not independent of the behavior of the obstacle object – if for instance both vehicles decelerate, both vehicles may still collide in the intersection. An analysis of GIDAS database has shown that in almost 40 % of all considered cases, the driver of the obstacle objects has attempted to avoid a collision by braking within the last second prior to the collision (Stoff and Liers 2013). As a consequence, attempting to avoid a collision temporally in the prioritized vehicle will still lead to a collision in many cases, if both vehicles involved do not coordinate their behavior. It also cannot be guaranteed that such an intervention will not worsen the situation compared to an accident without any intervention even if speeds are lower (for instance, if one vehicle is struck at the passenger cabin, while without braking, the collision would have hit the vehicle front instead).

On the assumption that a worsening of the situation has to be avoided under all circumstances, temporal collision avoidance is not suggested as intervention strategy for the prioritized vehicle.

Therefore, if the waiting duty vehicle is the last collision partner able to avoid a collision and does not do so, collision avoidance by the prioritized vehicle independent of the behavior of the collision partner (if at all) is only possible by evasion. Since the time needed for such maneuvers is quite low, the potential benefit with regard to increased traffic safety is relatively high. This in turn raises the question of the suitable direction of evasion – into or against the direction of motion of the obstacle object? Objectively judging evasive steering against the direction of motion of the obstacle object would be the latest possible safety measure for geometrical collision avoidance. However, steering against the direction of the collision object may not be intuitively comprehensible for the driver, which in return may negatively affect driver acceptance.

If neither braking nor evasion is able to avoid the collision, a collision mitigation maneuver remains as last alternative.

3 Situation Interpretation

Each of the intersection assistance systems described above requires an interpretation of the driver's intention to either stop the vehicle ahead of a certain point (e.g., stop bar, line of visibility, lane markings, etc.) or to pass this point without stopping and enter/cross the intersection.

In case of a STOP sign as right-of-way regulation, this interpretation is relatively easy since the STOP sign mandates that the driver stops his vehicle at the clearly marked STOP bar. This rule is independent of potentially privileged third-party vehicles in crossing traffic. Thus, the stopping point is well defined for both the driver and the system. The same is true for the necessity to come to a complete halt at this point. By combining different indicators (such as a correlation of velocity and distance to the stopping bar as well as the operation of brake and accelerator pedal by the driver), the question on whether or not the driver is going to run over a STOP sign unintentionally can be answered relatively early while approaching the intersection (Meitinger et al. 2004). Regarding the driver interpretation of the driver's intention a similar relation is expected for traffic light assistance systems (Kosch and Ehmanns 2006).

In case of turn-into/crossing assistance systems, this interpretation of the driver's intention is more complicated. Since the necessity of a complete halt depends on the presence of privileged vehicles, a driver decision on whether or not to stop the vehicle ahead of the intersection/junction is taken in a relatively late phase of the intersection approach, especially if the visibility of crossing traffic is obscured, e.g., by buildings or parking vehicles. As a result, a cross-traffic assistance system needs to take into account the possibility that a driver revokes an earlier decision and cancels a braking.

Thus, a cross-traffic assistance system needs to continuously assess potential changes of an interpreted driver decision and react accordingly. This is particularly challenging in situations, in which the driver reduces the vehicle speed while approaching the intersection to gain time to better assess crossing traffic.

Driver behavior studies show that at yield sign-regulated intersections, drivers usually start reducing the vehicle speed before passing the last possible point for a warning intervention. By implication the absence of speed reduction can be used as indicator for a potentially absent driver reaction to an intersection. Also, this indicator is available at a time when a warning is in many cases still sufficient to prevent the vehicle from entering the intersection. If however the driver decelerates the vehicle and only later decides to abort this deceleration or even accelerate to cross the intersection instead of stopping the vehicle (e.g., due to driver failure in assessing privileged crossing traffic on the major road), this usually is visible after the last possible warning point. In this case the change in driver intention cannot be identified in time for a warning intervention, and the vehicle can only be prevented from entering the intersection by an automatic emergency braking (Mages 2009). The same is true for turn-into accidents as well, but increased degrees of freedom regarding the stopping position have to be considered.

In addition to the identification of driver intention, turn-into/crossing and turnover assistance systems require an interpretation or third-party vehicles regarding the right-of-way regulation to estimate the probability of a potential collision. To quantify the danger of a collision requires not only a driver behavior prediction (ego vehicle) and a prediction of the movement of other road users but also an abstracted, mathematical description of the problem. Two potential approaches are illustrated in this chapter.

One way to assess the probability of potential collisions with other road users is to represent the situation based on 3D trajectories for all vehicles involved as used in the prototype implementation of a turnover assistance system (Meitinger et al. 2006). In an intersection-fixed coordinate system, the predicted path of each vehicle is described in two dimensions (x and y) by geometrical coordinates representing the location of the vehicles on the road surface with a third dimension representing the estimated time of a vehicle to reach this x/y position. This leads to a so-called 3D cloud for each vehicle, representing its future path. A collision is possible if two 3D clouds of different vehicles overlap.

One alternate illustration of same problem is a time difference map as used for a prototype turn-into/crossing assistance system (Mages 2009). Instead of the actual time to reach a certain point, the time difference map shows the time difference between two potentially colliding vehicles when occupying the corresponding x-/y-coordinates in the third coordinate axis. If the time difference $|\Delta||\Delta t|$ of any point in the map deceeds the minimum time difference that is typically accepted by drivers in general, a warning can be issued. If the time difference is zero, both vehicles are predicted to occupy the same position simultaneously, so an automatic emergency braking is activated to avoid an imminent collision.

To assess different action strategies when assisting privileged vehicles with the right of way in intersection accidents as discussed in Sect. 2.5, the additional question arises, which of the road users involved can avoid a collision last. One criterion to

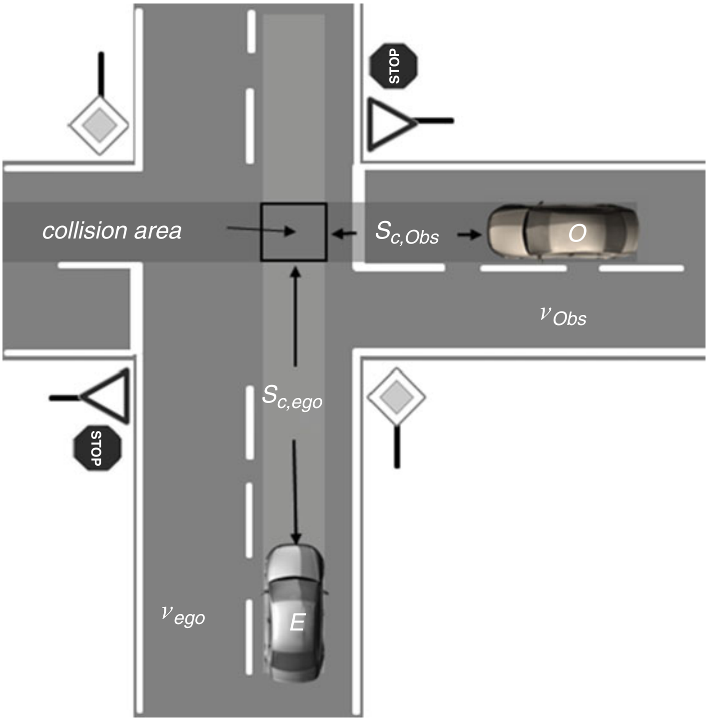


Fig. 3 Collision zone in the intersection

address this question is the comparison of the t_{tB} (time-to-brake), i.e., the remaining time to the last possible brake intervention $t_{tB,ego}$ and $t_{tB,obs}$ as illustrated in Fig. 3. The correlation is described by Eqs. 1 and 2 with s_k representing the current distance to the collision area and s_b representing the braking distance at the highest possible deceleration D_{max}. The t_{tB} corresponds with the time to collision (TTC, representing the time until a collision occurs if the velocity remains constant) while considering the individual distance to the predicted collision zone.

$$t_{tB,obs} = \frac{s_{k,obs} - s_{b,obs}}{v_{obs}} \text{ with } s_{b,obs} = \frac{v_{obs}^2}{2 \times D_{max}} \tag{1}$$

$$t_{tB,ego} = \frac{s_{k,ego} - s_{b,ego}}{v_{ego}} \text{ with } s_{b,ego} = \frac{v_{ego}^2}{2 \cdot D_{max}} \tag{2}$$

The vehicle with a higher t_{tB} value can still successfully avoid a collision by braking when this is no longer possible for the potential collision partner. The analyses of the TTC alone would not suffice in this case, as the still tolerable position P_{CA} of both vehicles differs (except for a corner to corner collision as illustrated in Fig. 4 left) from the collision position P_{Coll}.

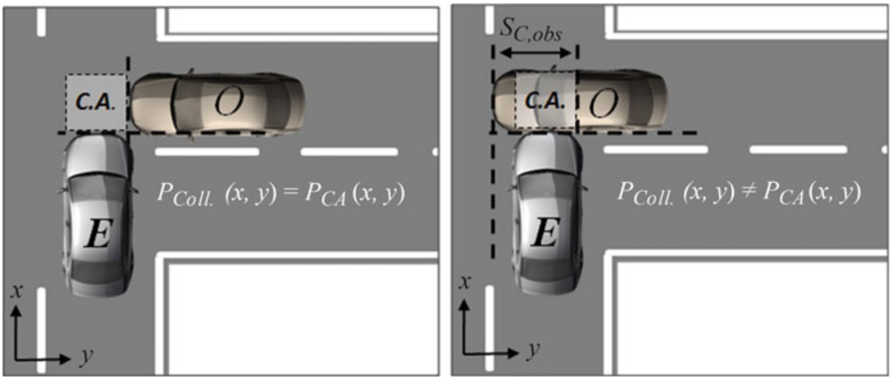


Fig. 4 Constellation of two colliding/not colliding vehicles

If the condition $t_{tB,obs} < 0$ is true during the intersection approach and if there is no braking or evasive maneuver by the potential collision partner, the vehicle that is mandated to stop/give way can no longer avoid a collision. At this point in time, it is suggested to assess whether there is an action strategy for the privileged vehicle that still allows it to avoid a collision. If the condition $t_{tB,ego} \geq 0$ is true, emergency braking may suffice; otherwise, an evasive steering maneuver may be the only option to avoid or, depending on the collision constellation, mitigate an imminent collision.

4 Warning and Intervention Strategies

Except for rear-end collisions in longitudinal traffic, most of the collisions between multiple vehicles result from right-of-way violation, commonly caused by a driver fault when assessing the right-of-way regulation.

Thus the obvious approach to avoid intersection accidents is to assist the driver responsible for the right of way regulation in a way that either supports the driver's assessment of the right of way regulation or minimizes the harmfulness of a wrong driver reaction to the current right of way regulation by active safety measures. In the context of this chapter, this type of intersection assistance is referred to as "assistance for the vehicle mandated to yield." To maximize the overall benefit of an intersection assistant on road traffic safety, the following subchapter describes which safety measures are available for "assistance for the prioritized vehicle."

4.1 Assistance for the Vehicle Mandated to Yield

It should be obvious that in cases with a very limited time available between the first possible detection of an imminent collision and the latest possible (driver) reaction to avoid such a collision, the effect of sheer driver information is also very limited,

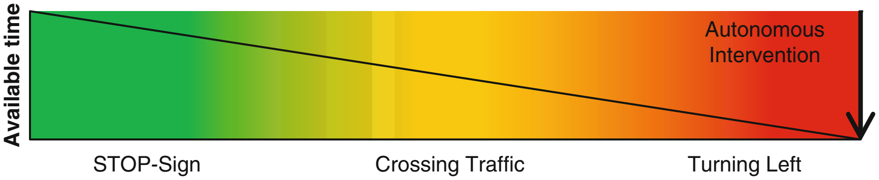


Fig. 5 Potential HMI solution for intersection assistance systems using a HUD (Institut für Straßenverkehr 1998)

since the driver still requires time to assess the information and to react accordingly. That is why – depending on the type of intersection accident – an automatic emergency braking may be the only active safety measure capable of collision avoidance, since the time for a driver reaction is not available. Figure 5 illustrates different intersection assistance systems as presented earlier, sorted by the time/space available for active safety measures to defuse a critical situation after detecting the criticality of the situation, and suggests a human-machine interface (HMI) to assist the driver with maximum efficiency.

Whenever the time available for driver intervention is very limited, driver information may not suffice and specific driver warning or even autonomous system intervention may become necessary. Studies in a driving simulator showed that a driver warning for intersection assistance was mostly accepted by the driver, if in retrospect the driver mistake was obvious. Therefore, to gain maximum driver acceptance, a system intervention (be it warning or automatic emergency braking) should only be issued, if the driver does not intend to avoid a collision. This requirement is also supported by legislatory aspects.

This leads to the so-called warning dilemma, an effect already discussed on earlier chapters. The warning dilemma describes the conflict between the efficiency of a warning and the estimated number of unnecessary warnings in real-world traffic. This conflict results from the fact that for an effective warning, the timing needs to take into account the time necessary for a driver reaction. Thus, to avoid a collision by driver warning, the warning has to be issued at a time, when he can still avoid the collision by himself. The option to avoid a collision by driver warning without negative effect on driver acceptance does only exist, if a driver reaction is detectable in most situations before the time of the last possible warning. Thus, if such a driver reaction is absent at the last possible warning point, the driver behavior can be considered atypical, which – if it occurs in combination with severe criticality of a collision – can be considered a sufficient indicator to issue a driver warning.

To determine whether or not the warning dilemma as mentioned above is a critical factor for intersection assistance systems and thus whether a driver warning is an effective safety measure to avoid intersection collisions, an analysis of "typical" driver behavior in intersections is required. Due to the complexity of intersection traffic scenarios, no general answer to this question can be given, but for specific scenarios a positive driver acceptance is deemed feasible under certain conditions. For turn-into/crossing assistance systems, for example, the driver intention to stop and give way is only visible in time, if the vehicle speed while approaching the intersection exceeds a certain threshold (Mages 2009). Below this velocity the reliability of driver intention detection is reduced dramatically, which would lead to either an unacceptable number of false-positive warnings or severely reduced efficiency of warnings.

If limiting factors like this are known, they can be considered when choosing an appropriate warning – and intervention strategy. One option is to enhance the driver warning by an active brake jerk or automatic emergency braking with only partial brake pressure (Mages 2009). This not only leads to a reduction of vehicle speed during the driver's reaction time, which in return effectively increases the time available for a driver reaction, it may also reduce the average reaction time required by the driver due to haptic feedback. As a result, the warning can be issued later, which may increase the reliability of a driver intention detection and therefore driver acceptance.

There are intersection accident scenarios, in which even the use of a brake jerk or partial emergency braking does not suffice to avoid a collision, sometimes independent of the right-of-way regulation (though some right-of-way regulation may lead to more corresponding scenarios than others). One example is the case of a vehicle starting from a complete halt at the intersection.

In the case of crossing/turning into intersections with one vehicle standing at the line of sight directly ahead of the intersection, there is no space and thus no time available for a driver reaction after the vehicle starts moving. In case of STOP sign-regulated intersections, the driver is mandated to stop at a clearly marked line, which usually is not identical with the line of sight, so in the case of a vehicle starting off after stopping at a STOP sign, the vehicle may not be as close to the potential collision zone and thus the situation may not be as critical. For turning-over accidents, a driver warning is in many cases not sufficient, even if the vehicle is still driving (Sect. 2.4).

In these scenarios only an autonomous emergency braking as safety measure can effectively avoid the vehicle from entering the conflict zone. If the vehicle is standing directly at the intersection with no space between itself and the potential conflict zone, it even seems justified to prevent the vehicle from starting into the intersection by disabling the accelerator pedal. One possible implementation is to combine this deactivation of the accelerator pedal with a warning (mostly to inform the driver about the measure) and to keep the accelerator pedal disabled until the driver presses and releases the brake pedal, thus acknowledging the necessity of a vehicle stop. This measure was implemented in a prototype intersection assistance system in a driver simulator with positive feedback regarding both efficiency and driver acceptance (Mages 2009).

One additional scenario, in which an intersection accident usually cannot be avoided by either driver warning or partial autonomous braking occurs, when a driver obliged to give way before crossing/turning into an intersection decelerates while approaching the intersection, e.g., to gain time to assess privileged traffic. If the driver decides to abort this deceleration due to a false interpretation of third-party vehicles in crossing traffic and accelerate into the intersection, this intention usually cannot be detected early enough to avoid a collision by warning as discussed in Sect. 3. Since driver failure in recognizing/assessing privileged traffic is a frequent cause of turn-into/crossing accidents, the potential effect of an intersection turn-into/crossing assistance system using only a driver warning or partial brake intervention is severely reduced in comparison to a system capable of full autonomous emergency braking (Mages 2009).

If entering the intersection/junctions as a potential conflict zone cannot be avoided, it may be an option to introduce a so-called tolerance zone. This approach has been used in a prototype turnover assistance system. While crossing the center lane marking that separates the own lane from opposing traffic could be avoided for only 80 % of the cases, it was still possible to avoid the vehicle from entering the opposing lane for more than 0.2 m in 95 % of the cases. Comparing the lane width and vehicle width, it seems justified to assume that an overlap with the opposing lane of 0.2 m or less can still be considered collision avoidance in most cases (Meitinger et al. 2006).

Due to current legal regulations and also because of the Convention on Road Traffic (more commonly known as the Vienna Convention), all safety measures mentioned above have in common that the driver has to be able to overrule or deactivate the systems at any time. One example to fulfill this requirement would be by using the kick-down position of the accelerator pedal to deactivate any automatic brake intervention. This way the driver can overrule the system without the use of additional control elements.

4.2 Intersection Assistance for the Prioritized Vehicle

For acceptance reasons any type of intervention in the prioritized vehicle (including a warning) should be issued in the latest possible moment, when otherwise the collision can no longer be avoided. As a consequence the time budget for collision avoidance is very short. While driver warning will not be effective to avoid a collision, the specific implementation of an HMI in the prioritized vehicle has to provide information about the active intervention.

When the waiting duty vehicle can no longer avoid a collision by braking ($t_{tB,obs} = 0$, see Sect. 2.5), the prioritized vehicle may in some scenarios still be able to intervene successfully by emergency braking or emergency evasive steering, depending on the situation. Assessing this means answering the question, whether a collision avoidance at $t_{tB,obs} = 0$ is still possible for the prioritized vehicle. In longitudinal traffic scenarios the necessary evasion offset is to be assumed as constant. In Sect. 2.5 it has been derived that the optimization of the distance-related offset, which in turn follows the last possible time of a successful intervention, can be realized by a combined braking/evasion maneuver.

In intersection scenarios the lateral offset required to evade an obstacle object in crossing traffic varies over time due to the intersecting trajectories. The larger the effective deceleration of the host vehicle during the maneuver and the longer the distance of travel to the “new” collision zone (compared to the initial driving path straight ahead), the later the host vehicle will arrive at the height of the potential collision area. Due to the velocity component of the obstacle object perpendicular to the host vehicles motion, this in turn leads to an additional distance of travel covered by the obstacle vehicle, which in turn decreases (in case of evasive steering against the direction of the crossing object) or increases (in case of steering with the direction of the crossing object) the necessary lateral offset to successfully avoid the collision. To assess whether an evasion maneuver is feasible in a given situation, a system first needs to estimate which evasive trajectories are available and at which time the corresponding positions will be reached. The next step is putting trajectories of the host vehicle and the obstacle object into relation to one another.

To maximize the scenarios addressable by such a system, evasive maneuvers should make best use of all space available, and furthermore, while accelerating has to be rejected for safety functionalities, the velocity of the evading host vehicle should be as constant as possible in the specific situation, to minimize the additional time required for the maneuver. In addition to the available lateral offset according to lateral vehicle dynamic, further possible collision objects have to be considered to assess the “ideal” escape trajectory. A simplified approach for estimating available trajectories according to vehicle dynamic limitations is the two-time integration over time of the effective target lateral acceleration $a_{y,\mathrm{set}}$. This maximum possible lateral acceleration depends on the existing coefficient of friction μ. The transition conditions for determining the duration of the individual phases of the maneuver (steering, maximum lateral acceleration, counter-steering t_{CS}, maximum lateral acceleration) result from the constraints of the following equations (see also Fig. 6):

$$\int_{t_0}^{t_{\mathrm{end}}} a_y(t')dt' = 0 \tag{3}$$

$$\int_{t_0}^{t_{\mathrm{end}}} v_y(t')dt' = y_{\max} \tag{4}$$

$$t_{\mathrm{end}} = t_{\mathrm{resp}} + t_{a_y,\max,1} + t_{\mathrm{trans}} + t_{a_y,\max,2} \tag{5}$$

One example of different types of action is shown in Fig. 7. An intervention is only able to avoid a collision as far as the two trajectories do not meet or intersect. In the example, only the evasion maneuver may avoid an imminent collision. If collision avoidance is still possible by either braking or evasive steering, emergency braking should be chosen as safety measure due to the higher complexity of an evasive steering maneuver.

If neither braking nor evasion is sufficient to avoid the collision, mitigation maneuvers remain as a last option. Obviously the severity of accidents can be mitigated by a reduction of the velocity vector Δv and with it the kinetic energy of

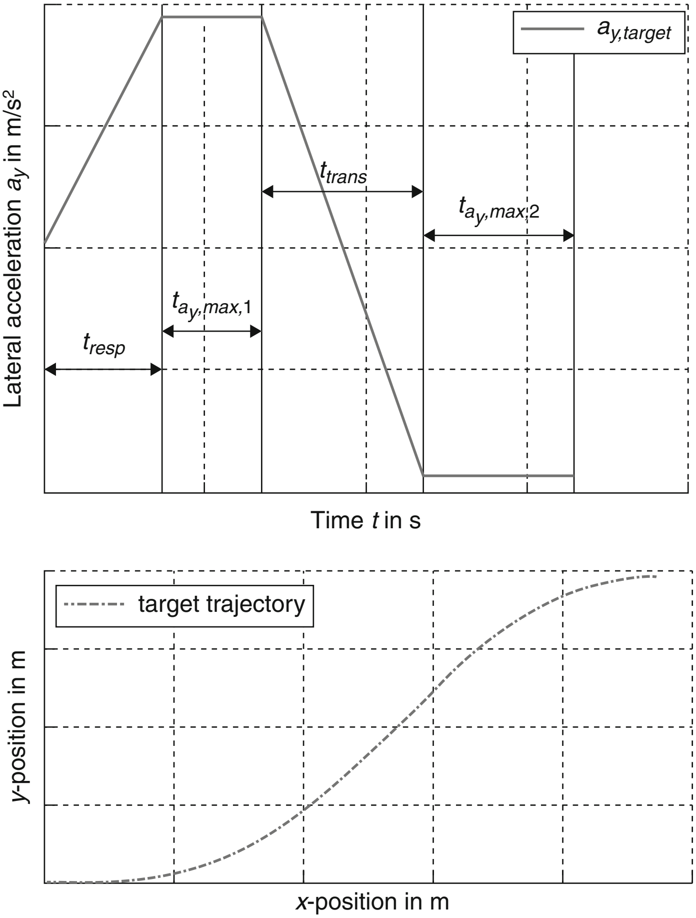


Fig. 6 Example of time course of lateral acceleration while evading and the following trajectory

the crashing vehicles. However one decisive factor for the consequences of intersection collisions is the constellation of colliding vehicles. Accidents with crossing traffic often result in one vehicle being struck at or close to the passenger cabin. Such collisions lead to more severe or fatal injuries than collisions, in which none of the vehicles involved is hit at the passenger cabin. Thus, to reduce accident severity of an imminent collision, one possible strategy is to minimize the collision velocity while simultaneously shifting the hit point as far as possible away from the passenger compartments for both vehicles. Both requirements can be fulfilled with a defined braking maneuver (Heck et al. 2013) or a combined steering/braking maneuver (Stoff and Liers 2013). The advantage of a combined maneuver is that it may also reduce the collision angle between the two collision partners due to the already discussed direction of evasion (into the direction of motion of the obstacle object), which in turn leads to a further reduction of the relative velocity.

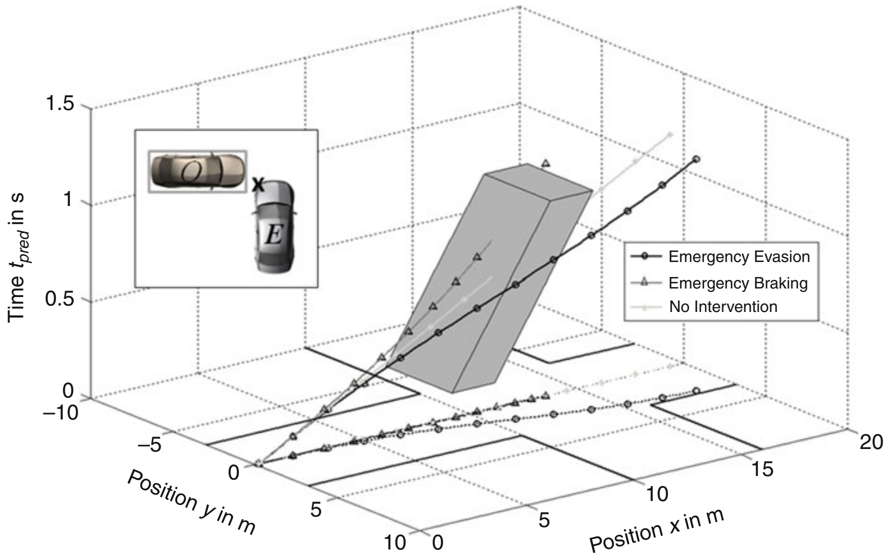


Fig. 7 Example for assessing the potential of different measures to avoid a collision at intersections

One possibility for the realization of such a maneuver is presented in Stoff and Liers (2013). This approach is based on the consideration of a band of trajectories with varied acceleration vector $\vec{a}_i$, described by its absolute value $\left|\vec{a}_i\right| = \mu \cdot g$ and the respective angle γ_i, according to Eqs. 3, 4, 5, and 6:

$$x_{pr}(t,\gamma) = \int_{t_0}^{t} v_{pr}(t',\gamma_i) \cdot \cos\left(\psi_{\kappa,pr}(t',\gamma_i)\right) dt' \tag{6}$$

$$y_{pr}(t,\gamma) = \int_{t_0}^{t} v_{pr}(t',\gamma_i) \cdot \sin\left(\psi_{\kappa,pr}(t',\gamma_i)\right) dt' \tag{7}$$

$$v_{pr}(t,\gamma) = v_0 + a_{x,\mathrm{set}}(\gamma_i)(t - t_0) \tag{8}$$

$$\psi_{\kappa,pr}(t,\gamma) = \int \frac{a_{y,\mathrm{set}}(\gamma_i)}{v_{pr}(t',\gamma_i)} dt' \tag{9}$$

In these equations the angle γ_i is varied between $\gamma_0 = 0°$ (only braking) and $\gamma_{\max} = 90°$ (only steering). Depending on the direction of motion of the obstacle object, the available trajectories of the front corner of the host vehicle, relevant for the collision, $P_{\mathrm{ego},\gamma,t}$, for varied γ_i within the prediction horizon t_{pred} could be compared with the predicted target point of collision $P_{\mathrm{obs,soll},t}$ at the obstacle object (see Fig. 8). The minimum of this comparative analysis following Eq. 7 describes the

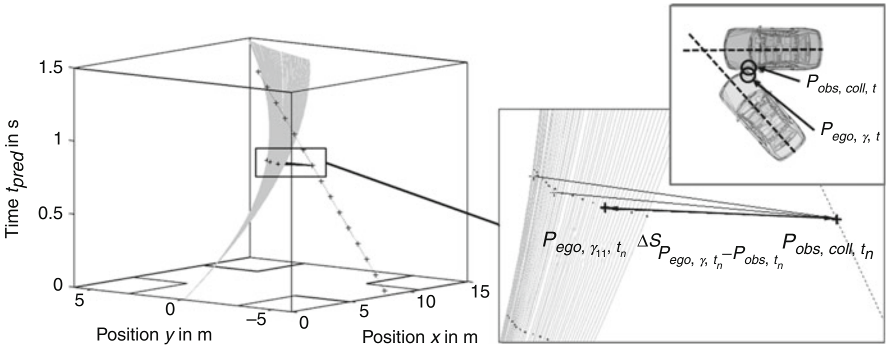


Fig. 8 Trajectory prediction – comparison of reachable sets with target point of collision over time

minimal realizable Euclidean distance to the target point of collision that can be realized with the band of trajectories within the prediction horizon. As far as this value does not exceed a certain limiting value $\Delta s_{\max}$, it determines in accordance to the corresponding $\gamma = \gamma_{CM}$, converted in the related acceleration vector $\vec{a}_{\gamma_{CM}}$, the target setting for trajectory control, as well as the predicted collision time $t = t_{CM}$.

$$\mathrm{Min} \sum_{\gamma=\gamma_0}^{\gamma_{\max}} \sum_{t=t_0}^{t_{\mathrm{pred}}} \left(P_{\mathrm{ego},y,t} - P_{\mathrm{obs,soll},t}\right) = \Delta s_{\gamma_{CM},t_{CM}} \tag{10}$$

The simplified approach known from evasion maneuver in longitudinal traffic, based on the transfer of Kamm's circle into the so-called spatial domain, does not fit the problem discussed here, as changes in the course angle are not considered. This, in the case of an evasion maneuver without counter-steering, leads to increasing deviations from the planning with increasing duration of the maneuver (see ▶ Chap. 46, "Fundamentals of Collision Protection Systems"). This deviation by far exceeds the tolerance for the target area (area of the obstacle vehicle in front of the front axle) for a collision-mitigating maneuver presented here, so that a desired collision constellation could not be ensured in this way.

5 Challenges

The accidentology data shows that intersections and junctions allow for a relatively high increase in road traffic safety, if addressed by active safety systems. This is mostly due to the fact that so far there is no designated intersection assistance system in series production. The latest version of the "Brake Assist System plus" or "BASplus," introduced in the Mercedes-Benz S-Class (W222), demonstrates a first step in this direction. This system detects crossing pedestrians and (unlike similar systems before, ▶ Chap. 46, "Fundamentals of Collision Protection Systems") crossing traffic. Based on the sensor detection, a driver warning is issued, and, if a brake activation by the driver is detected, the brake pressure is increased to assist the driver in avoiding the collision. The manufacturer markets the system as brake assist system with specific functions for pedestrians and crossing traffic (Daimler 2013).

One potential reason preventing the introduction of intersection assistance systems into series production so far is the relatively complex traffic situation in junctions and intersections. Thus, intersection traffic scenarios result in particularly demanding requirements for active safety sensors, not all of which can be fulfilled with today's series sensors.

A basic STOP sign assistance function for instance only requires the information about the current right-of-way regulation and the vehicle's distance to the designated stopping line, while a cooperative intersection assistance system as described in Sect. 2.3 poses additional informational requirements regarding, e.g., possibly prioritized third-party vehicles and intersection geometry. Intersection-specific data may be provided by advanced digital maps (Weiss and Dietmayer 2007) with the well-known issue that these maps have to be kept up to date.

If no communication technology is available within the vehicle, detecting third-party vehicles becomes particularly difficult in cases with obscured line of sight, which would pose a serious limitation for crossing and left turn assistance systems. Vehicle-to-vehicle communication technology is independent from whether or not the line of sight is obscured and thus can compensate for the lack of information provided by self-sufficient surround sensors (i.e., RADAR, LIDAR, or video) if the sensor field of view is limited by, e.g., obstruction or parking vehicles.

It should be noted that the potential of communication technology for collision avoidance in intersections highly depends on the fitment rate of this technology in the field. Unlike self-sufficient sensors, communication technology requires a third-party vehicle equipped with a compatible communication solution to be beneficial in a critical situation. A visible reduction of the number of intersection collisions due to communication technology can only be expected, if a relevant portion of vehicles (or intersections) is fitted with such systems. Among other aspects of communication technology, the research project simTD is looking into the potential of this technology for intersection assistance (simTD Deliverable 2009).

The available accuracy of measured data leads to additional limitations regarding the accident types, which can be addressed successfully by intersection assistance systems. The driver intention detection of a prototype turn-into/cross assistance system is a good example to see the influence of sensor accuracy (Mages 2009). Figure 9 illustrates velocity over intersection distance for a sporty driver while approaching the intersection with the intention to stop, the braking distance of a vehicle assuming a constant deceleration ($a_x = -8$ m/s^2), and the stopping distance of the same vehicle. The stopping distance has been simplified as sum of braking distance and the distance traveled with a partial emergency braking during the reaction time of the driver ($T_R = 1$ s; $a_x = -2$ m/s^2).

The illustration shows that with lower speed, the deceleration curve of the sporty driver gets close to and in fact crosses the curve of the stopping distance.

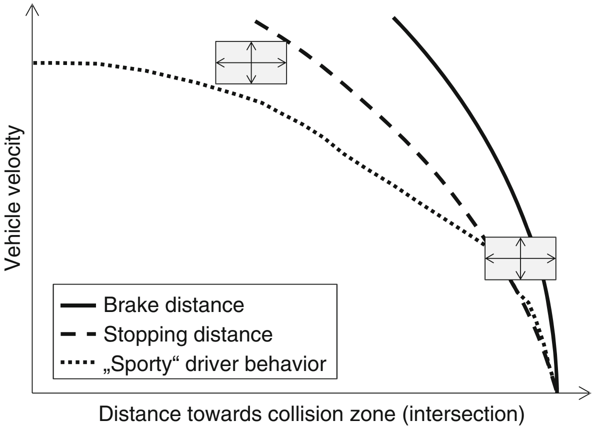


Fig. 9 Effect of sensor inaccuracy on potential warning/intervention thresholds of a turn-into/crossing assistant

This means that the warning dilemma as mentioned in Sect. 3 is only manageable at speeds higher than the vehicle velocity at this point of intersection. It can be seen that inaccuracies in distance and vehicle velocity (in Fig. 9 represented by a rectangle for fictional sensors) have a direct impact on this velocity. Thus, the expected inaccuracies of all required information have to be considered when defining the appropriate intervention strategies like warning criteria for different accident types. In the case of Fig. 9, this means that the minimum velocity, below which a warning does not suffice to avoid a collision since the driver intention cannot be assessed in time, will increase if sensor accuracy is worsened. A similar correlation exists for the assessment of third-party vehicles: if the speed of potential collision partners is reduced, the requirements for sensor accuracy increase (Mages 2009). This is why the sensor accuracy requirements are higher for intersections in inner-city areas due to the slower vehicle speeds. Nevertheless, inner-city areas account for the majority of all intersection collisions and thus should not be left out in the development of such systems.

One general condition for all active safety systems is driver acceptance. A reduction of accident numbers in intersections is only possible, if drivers are willing to use intersection assistance systems. It should be ensured in an early phase of the development process that the driver accepts the intended system functions. For some of the intersection assistance functions described in this chapter, the HMI has been evaluated in the driving simulator of BMW Group (Gradenegger et al. 2006) with real test persons. The results of these tests allow a first estimation regarding the possible safety benefit. In addition there are real vehicles equipped with different types of intersection assistance systems as result of the recently ended research projects AKTIV and INTERSAFE-2 (AKTIV 2010; Meinecke et al. 2009).

Each of the assistance functions described in this chapter only addresses a specific part of intersection accidents. Instead of introducing separate systems to the market, it seems reasonable to integrate all intersection assistance functions into one intersection assistance system. From a technical point of view, overlapping sensor requirements result in synergy regarding the necessary hardware components. In addition this allows the use of one consistent human-machine interface (HMI), which may have a positive effect on driver acceptance compared to isolated applications. First approaches in this direction can be seen as result of the EU-funded research project PReVENT (Hopstock and Klanner 2007).

One potential stimulant for the development and introduction of intersection assistance systems for mass production might be the consumer test organization Euro NCAP. The announced enhancement of the vehicle safety five-star rating for future years among other target actions addresses junction and intersection scenarios for 2020 (Miller et. al. 2015). Thus, future assessment tests may well include intersection assistance systems.

References

AKTIV (2010) AKTIV homepage. www.aktiv-online.org. Accessed 26 June 2015

Branz W, Öchsle F (2005) Intersection assistance – collision avoidance system for turns across opposing lanes of traffic. In: Proceedings of: 5th European congress on ITS, Hannover

Chovan J, Tijerina L, Everson J, Pierowicz J, Hendricks D (1994) Examination of intersection, left turn across path crashes and potential IVHS countermeasures. Potential IVHS Countermeasures, Cambridge

Daimler AG (2013) Mercedes-Benz S-Class Advertising Brochure, Stuttgart

Donath M et al (2007) Intersection decision support: an overview – final report. University of Minnesota

Garber N, Miller JS, Abel R et al (2007) The impact of red light cameras (photo-red enforcement) on crashes in Virginia, Virginia transportation research council, final report, Charlottesville

GIDAS (2010) German in-depth accident study. Dresden & Hannover

Gradenegger B et al (2006) Untersuchung des Linksabbiegeassistenten, des Querverkehrsassistenten, des Ampelassistenten und des potentiellen Nutzens eines Workload-Management-Systems [Analysis of a left turn assistant, a cross traffic assistant, a traffic light assistant and the potential benefit of a workload management system]. Final report, Würzburger Institut für Verkehrswissenschaften, Würzburg

Heck P et al (2013) Collision mitigation for crossing traffic in urban scenarios. In: Proceedings of IEEE intelligent vehicles symposium, Gold Coast

Hoppe M, Zobel R, Schlag B (2006) Identifikation von Einflussgrößen auf Verkehrsunfälle als Grundlage für die Beurteilung von Fahrerassistenzsystemen am Beispiel von Kreuzungsunfällen [Identification of influencing factors on traffic accidents with regard to an assessment of advanced driver assistance systems, using the example of intersection accidents]. In: Proceedings of Fahrer im 21. Jahrhundert, Braunschweig

Hopstock M et al. (2007) Advanced systems for intersection safety within the BMW dynamic driving simulator. In: Proceedings of PReVENT in action, Versailles

Hopstock M, Klanner F (2007) Intersection safety – just a vision? BMW activities for active and preventive safety at intersections. In: Proceedings of car safety, Berlin

Institut für Straßenverkehr (1998) Unfalltypenkatalog [Accident type catalogue]. Cologne

Intersafe (2005) D40.4 “INTERSAFE requirements”. PReVENT SP Deliverable, Brussels

Klanner F (2008) Entwicklung eines kommunikationsbasierten Querverkehrsassistenten im Fahrzeug [Development of communication-based cross-traffic assistance in the vehicle], VDI-report series 12, Nr. 685, Düsseldorf

Kosch T, Ehmanns D (2006) Entwicklung von Kreuzungsassistenzsystemen und Funktionalitätserweiterungen durch den Einsatz von Kommunikationstechnologien [Development of intersection assistance systems using communication technology]. In: Proceedings of Aktive Sicherheit durch Fahrerassistenzsysteme, Munich

Kraftfahrt-Bundesamt (2012) Bestand an Kraftfahrzeugen und Kraftfahrzeuganhängern nach Fahrzeugalter [Vehicle age in Germany], FZ 15

Mages M (2009) Top-Down-Funktionsentwicklung eines Einbiege- und Kreuzenassistenten [Top-down development of a turn-into- and crossing assistant]. VDI report, series 12, Nr. 694, Düsseldorf

Meinecke MM et al (2009) User needs and operational requirements for a cooperative intersection safety system, Intersafe2 Deliverable 3.1

Meitinger KH et al (2004) Systematische Top-Down-Entwicklung von Kreuzungsassistenzsystemen [Systematic top-down development of intersection assistance systems]. VDI-report Nr. 1864

Meitinger KH, Heißing B, Ehmanns D (2006) Linksabbiegeassistenz – Beispiel für die Top-Down-Entwicklung eines aktiven Sicherheitssystems [Left turn assistance – an example for top-down development of an active safety system]. In: Proceedings of Aktive Sicherheit durch Fahrerassistenzsysteme, Munich

Miller A et al (2015) Euro NCAP 2020 ROADMAP, rev. 1.0, Brussels. http://euroncap.blob.core.windows.net/media/16470/ratings-group-report-2015-version-10-with-appendix.pdf. Accessed 26 June 2015

Pierowicz J et al (2000) Intersection collision avoidance using IST countermeasures, NHTSA DOT HS 809 171, final report

simTD (2009) simTD Deliverable D21.4 – Spezifikation der Kommunikationsprotokolle [Communication protocol specification], Sindelfingen

Statistisches Bundesamt (2010) Verkehr – Verkehrsunfälle 2009 [Traffic accidents in 2009]; Fachserie 8/Reihe 7. Wiesbaden

Stoff A, Liers H (2013) Ausweichfunktionalität für Kreuzungsszenarien zur Unfallfolgenlinderung durch Optimierung der Crash-Kompatibilität [Evasive steering to optimize crash compatibility in intersection accident scenarios]. In: Proceedings of 9. VDI-Tagung Fahrzeugsicherheit – Sicherheit 2.0, Berlin

Suzuki T, Benmimoun A, Chen J (2007) Development of an intersection assistant. Technical report, Denso Automotive (vol 12 No 1)

Weiss T, Dietmayer KCJ (2007) Automatic detection of traffic infrastructure objects for the rapid generation of detailed digital maps using laser scanners. In: Proceedings of IEEE intelligent vehicles symposium, Istanbul

Yan X, Radwan E, Klee H, Guo D (2005) Driver behavior during yellow change interval. In: Proceedings of DSC North America, Orlando

Traffic Jam Assistance and Automation

51

Stefan Lüke, Oliver Fochler, Thomas Schaller, and
Uwe Regensburger

Contents

S. Lüke (✉)
Continental Division Chassis & Safety, Advanced Engineering, Frankfurt am Main, Germany
e-mail: stefan.lueke@continental-corporation.com

O. Fochler
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: oliver.fochler@continental-corporation.com

T. Schaller
BMW Group, München, Germany
e-mail: thomas.sa.schaller@bmw.de

U. Regensburger
Daimler AG, Sindelfingen, Germany
e-mail: uwe.regensburger@daimler.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_52

Abstract

Traffic jams are situations where a high degree of automation could give a large benefit to customers. In addition, the relatively simple situation of a traffic jam means that a high degree of automation can be expected in the near future. This chapter explores in detail the motivation, the conditions, and the versions of assistance and automation systems designed to assist in traffic jams. Different levels of automation for traffic-jam assistance and automation systems will be discussed, such as the design of traffic-jam assistance systems and of automated following systems. Moreover, HMI concepts of state-of-the-art systems are presented, and their implications on controllability and take-over scenarios are discussed. Finally, the legal situation and therefore marketability aspects are regarded.

1 Introduction

There isn't a driver who does not find traffic jams irritating and annoying. They usually occur unexpectedly. The additional time that they consume on the way to work, to shopping, to friends, or on vacation causes a great deal of dissatisfaction, stress, and aggression.

Traffic jams are the kind of situations where a high degree of automation could confer a great deal of benefit on customers when a situation cannot be avoided. Moreover, the relatively simple situation of a traffic jam means that a high degree of automation can be expected in the near future. The following will explore in detail the motivation, the conditions, and the versions of assistance and automation systems that are designed to deal with traffic jams.

1.1 Motivation

Today's assistance systems for stop-and-go traffic assume control only over longitudinal motion and, thus, facilitate a driver's task only partially. By adding assistance or automation features to deal with lateral motion, it will be possible to facilitate a driver's task still further. Counterbalancing a high degree of automation are legal questions, the cost of systems, questions of liability, and additional risks posed by automation.

Adaptive Cruise Control (ACC) systems – mostly RADAR assisted – that assume control over longitudinal motion have established themselves to the extent that nearly every carmaker now offers them. Recent years have seen the increasing introduction of Full Speed Range Adaptive Cruise Control (FSRA) systems (see ▶ Chap. 45, "Adaptive Cruise Control") that govern a car's speed by bringing it to a full stop and even setting it into motion again.

The enhanced benefits of such systems mean that customers will probably make more use of them. Increased use of such regulating systems in turn magnifies the benefits of flowing traffic such as less CO_2 emission.

1.2 Acceptance

Due to their increased prevalence, a growing number of users have become acquainted with the way in which ACC systems assume control of longitudinal motion. However, drivers must permanently pay attention to lateral motion when they are stuck in traffic or in stop-and-go traffic. Studies have shown that most drivers would therefore prefer systems that could assume control over lateral motion in stop-and-go situations. In the forefront of customer expectations are relief from monotonous tasks in traffic and also the enhanced sense of safety while possibly performing other activities (Schaller et al. 2008). One can therefore assume that expanding the pure FSRA solution to include a lateral motion function will heighten the attractiveness and the chances of success of these systems.

1.3 Definitions

In order to categorize different forms of systems designed to assist drivers in traffic jams both technically and legally, it makes sense to consider them not individually but to assess them by means of a more generalized category. A report by the German Federal Highway Research Institute (BASt) contains such a catalogue of general assistance or automated systems arranged according to categories (Gasser et al. 2012). ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems" of this book details the legal aspects. Of importance for the discussion here are the definitions of the various design levels and their implications for a driver's responsibility for driving in a traffic jam.

Assistance systems assume control, within certain limits, over either the longitudinal or the lateral motion of a car. This presupposes that the driver monitors the system at all times and is prepared to reassume responsibility for driving. Partially automated systems assume control of longitudinal and lateral motion in certain scenarios; the driver, however, must constantly monitor the system here as well and be prepared to reassume control immediately. One way of achieving this is to require hands-on driving. Hands-off recognition (i.e., off the steering wheel) coupled with a deactivation strategy should prevent drivers from turning their attention away from traffic for longer periods of time. A major step is the transition to highly automated systems that allow drivers extra time to resume driving responsibility in specific situations, thus freeing them from the need to constantly monitor the system. The highest level of development is fully automated systems that completely free drivers from the need to monitor the system in defined circumstances.

A practical aspect is the question as to whether, and how long, drivers should be able to leave their hands off the steering wheel. Even partially automated systems permit this already as long as the – by definition constantly vigilant – driver is able to handle unexpected situations (Gasser et al. 2012). Naturally, it would be much more convenient if drivers were able to remove their hands from the steering wheel in stop-and-go traffic for longer periods of time and devote their attention to other activities.

2 Information About a Car's Immediate Environment

Before dealing with the different types of assistance and automated systems and going into their technical aspects, we wish to first tackle the question as to what information about the surroundings should be made available to the system and how should it be collected.

Information on lane structures such as markings provide the basis for any traffic-jam assistance function on the one hand and information on other vehicles in the immediate environment on the other hand. Camera-based systems (cf. ▶ Chap. 20, "Fundamentals of Machine Vision") usually detect lane markings. RADAR systems (see ▶ Chap. 17, "Automotive RADAR") and/or camera systems with object-recognition algorithms (see ▶ Chap. 24, "Representation of Fused Environment Data") provide data on the position and movements of vehicles in the vicinity, among other things.

Knowledge of the lane markings and the position and motion of the vehicle immediately ahead are initially enough to provide basic traffic-jam assistance. Thus, many of today's cars equipped with front RADAR for ACC systems (see ▶ Chap. 45, "Adaptive Cruise Control") and cameras for Lane Departure Warning or Lane Keeping Support (see ▶ Chap. 48, "Lateral Guidance Assistance") already possess the basic sensors needed. Any system governing lateral motion must orient itself to both lane markings and the vehicle in front because the vehicle in front may cover lane markings wholly or in part, especially in heavy traffic (Fig. 1).

Information on traffic to the sides and to the rear of a vehicle is indispensable for providing more complete assistance to a driver. Side-mounted and rearview cameras (surround-view systems) could monitor cars traveling alongside or lane markings alongside a car. Short- or mid-range radars could keep track of vehicles traveling alongside. Due to the relatively low speeds encountered in heavy traffic scenarios, it is possible to dispense with rear-mounted short- or mid-range RADAR.

GPS could be used at lower levels of automation to limit functions to safe scenarios (such as highway driving) that conform to predetermined areas of application. Map data could also be useful to improve system availability, however. If the complexity of the system under review grows to include highly automated functions, the inclusion of GPS positioning data and highly accurate mapping data may become necessary to provide the redundancy demanded by the number and curvature of lane markings. This information plays a decisive role in applying functions to urban scenarios in particular. Comparisons with landmarks captured above and beyond the road surface, such as bridge pillars or traffic signs, are a logical extension of such systems.

Fig. 1 An example of a traffic jam on the German Autobahn

3 System Design Variants

3.1 Stop-and-Go Assistant with Longitudinal Regulation Only

FSRA can be considered a functional and technical basis for higher forms of traffic-jam assistants and automation (see Fig. 2), because it covers all the longitudinal regulation of a vehicle.

Although FSRA provides longitudinal guidance, a driver must permanently perform lateral guidance him-/herself. The system thus remains technically and legally an assistance system by which a driver never relinquishes responsibility.

3.2 Traffic-Jam Assistant (Monitors Car in Front and Lane-Holding Assistant)

In contrast to the longitudinal guidance provided by an FSRA assistance system, traffic-jam assistants provide lateral guidance as well. They do so by generating steering force for the purpose of keeping a car on a predefined path.

Two different kinds of information go into the calculation of such an intended trajectory as detailed in Sect. 2. One concerns the position and, thus, the movement of nearby vehicles, while the other concerns lane markings. In order for path controlling to work, the intended trajectory must consist of data on deviation y_0 (eccentricity to the middle of the lane), orientation θ_Δ (heading) relative to the trajectory, and its curvature κ_T (cf. Fig. 3). As this is the input for lateral guidance as

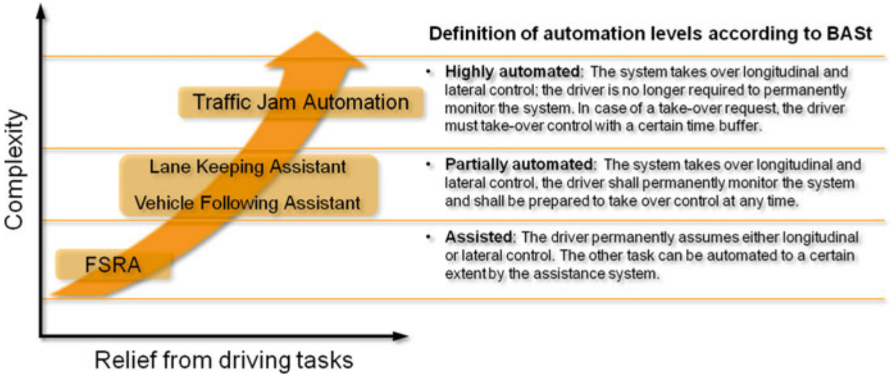
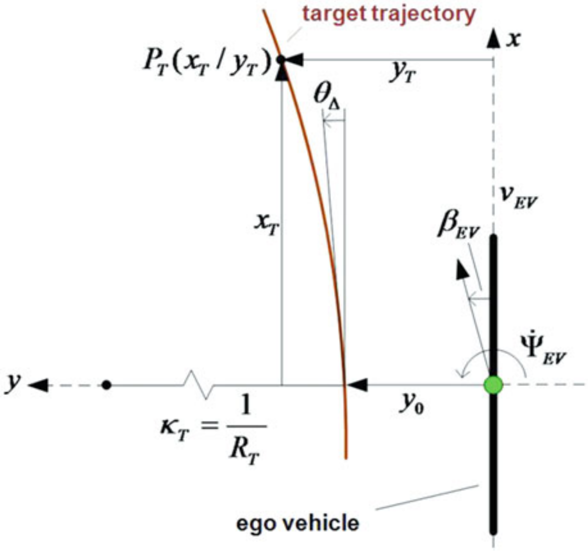


Fig. 2 Degrees of complexity and size of driver-assistance systems

Fig. 3 Data for describing trajectory guidance



shown in Fig. 4, these values should be observed through a Kalman filter. The following illustrates how to design this filter.

The following equations can serve to illustrate the movement of a vehicle along a trajectory, taking into account the vehicle's own movement (vehicle speed v_{EV}, slip angle β_{EV}, and yaw rate $\dot{\Psi}_{EV}$):

$$\begin{pmatrix} \dot{y}_0 \\ \dot{\theta}_\Delta \\ \dot{\kappa}_T \end{pmatrix} = \begin{pmatrix} 0 & v_{EV} & 0 \\ 0 & 0 & v_{EV} \\ 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} y_0 \\ \theta_\Delta \\ \kappa_T \end{pmatrix} + \begin{pmatrix} -v_{EV} & 0 \\ 0 & -1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \beta_{EV} \\ \dot{\Psi}_{EV} \end{pmatrix} \tag{1}$$

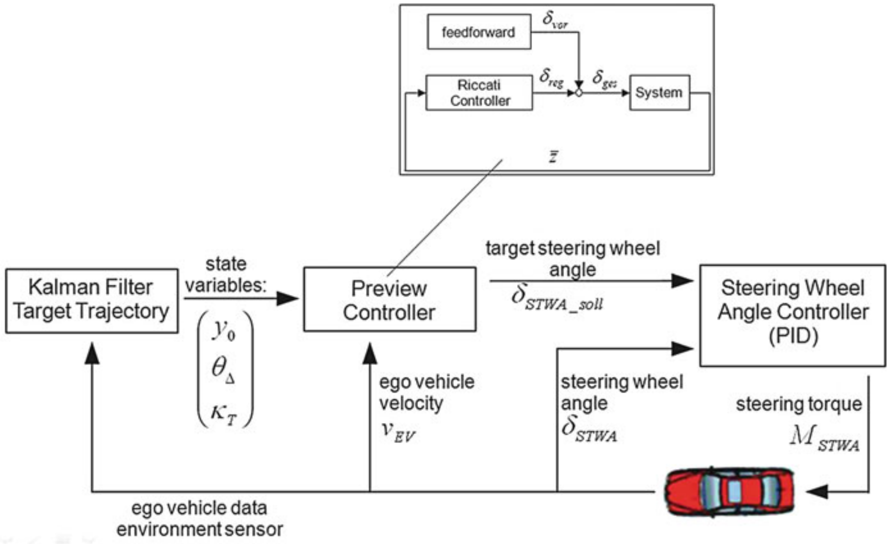


Fig. 4 Closed-loop system for lateral guidance

In order to take the relative position of the vehicle in front $P_{FC}(x_{FC}/y_{FC})$ into account, we assume that the vehicle in front will follow the intended trajectory, whereby $x_T = x_{FC}$ and $y_T = y_{FC}$:

$$y_{FC} = y_0 + x_{FC} \cdot \theta_\Delta + \frac{1}{2} \cdot x_{FC}{}^2 \cdot \kappa_T \tag{2}$$

Camera information on the relative position of lane markings and, thus, the distance to the middle of the lane, y_{0_BV}, the orientation θ_{Δ_BV} to the lines, and the curvature of the lane $\kappa_{T_{BV}}$ can serve directly to compute a lateral target trajectory to the middle of the lane (the midpoint between two lane markings). The BV index illustrates that the information on lane markings comes from image processing.

This results in the equation:

$$\begin{pmatrix} y_{FC} \\ y_{0\_BV} \\ \theta_{\Delta\_BV} \\ \kappa_{T\_BV} \end{pmatrix} = \begin{pmatrix} 1 & x_{FC} & \frac{1}{2} x_{FC}^2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} y_0 \\ \theta_\Delta \\ \kappa_T \end{pmatrix} \tag{3}$$

Depending on the expected quality of the information going into the calculation of the target trajectory, the covariance matrix Q can serve to achieve a weighting between the vehicle ahead and lane markings.

The vector calculated and, hence, the relative target trajectory can then be used to trigger a lateral adjustment as shown in Fig. 4. This is how the system calculates a steering adjustment, which steers the car laterally back to the target line.

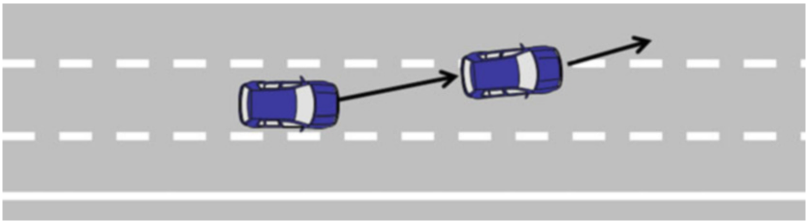


Fig. 5 Follow assistance system

First-generation traffic-jam assistance systems utilize information on the relative position of the vehicle ahead in cases in which sensors cannot adequately detect lane markings. The proximity of other vehicles that is typical at traffic-jam speeds often obscures the vision of ADAS cameras usually mounted behind the windshield. At the same time, however, information about the vehicle ahead is much more available and much more stable, especially in heavy traffic situations. The traffic-jam assistant monitors the vehicle in front (see Fig. 5). The system behaves in a manner that is totally understandable for a driver: it follows the car in front. Should the car in front change lanes, the driver would have to resume complete control of lateral movement because the assistant in this version would cause his/her car to follow the other car into the new lane. The information gleaned from detection of lane markings (eccentricity to the middle of the lane, orientation to the trajectory, and curvature of the trajectory) serves, if available, on the one hand, to improve quality of regulation by taking account of global orientation and curvature information and, on the other hand, to generate a prompt command to resume control and switch off lateral assistance in case the car in front crosses the lane markings to change lanes. In this situation, it is better to follow along the lane markings (in the middle of the lane) than to simply switch off lateral assistance, as long as the markings are clearly recognized long enough in advance. The purpose of the assistance here is to keep a car in lane until a driver resumes control.

From a driver's point of view, it is worth striving for lateral assistance within a lane based on lane-marking recognition. This means that development work on traffic-jam assistants will go more in the direction of lane-holding assistance in the future. Greater availability and more stable recognition of lane markings will be necessary for this purpose. Refined camera sensors that are already in use in parking-assistance systems that contain image-processing algorithms (based on information obtained from cameras mounted on outside rearview mirrors and the rear of a car) can achieve this. Furthermore, expanded sensors on the side and rear of a car can provide information on nearby vehicles to permit a car to flow along in traffic even though lane information may not be available over the short term (Schaller et al. 2008).

Although we must assume that a lane-holding assistance system (see Fig. 6) will meet a driver's expectations, it is more of a partially automated system. The fact

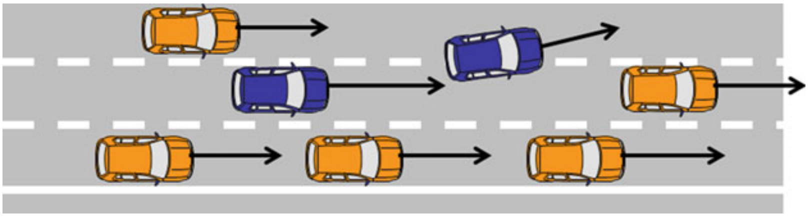


Fig. 6 Lane-holding assistant

Fig. 7 An automated drive in traffic (Source: Continental AG)

that it cannot be considered a highly automated system is due to limitations imposed on sensors by the possibility of false recognition or unavailability. A driver must constantly monitor a lane-holding assistant.

3.3 Automatic Following in a Given Speed Range

The version of a traffic-jam assistant described above meets the expectations of users, but it does require a driver to constantly monitor the system. The value of this system thus primarily lies in its ability to relieve drivers, but they will formally not be able to devote their attention to other things. A highly automated lane-following system (see Fig. 7) that conforms to the BASt listing (see Sect. 1.3) would be the next logical step. The fact that it would permit moderate attention to things other than driving would confer palpable value on such a lane-following system.

In spite of comparable functionality, the demands on such a system are much higher than on a lane-holding assistant. Since a driver is no longer required to constantly monitor the system, there must be assurance that the system does not cause a car to change lanes unintentionally. This results in greater demands on the availability of lane information.

Moreover, a driver must have sufficient time to resume driving if the system reaches its limits – in contrast to an assistance system that follows a car ahead. In the case in point, this would be the maximum speed for an automatic lane-following system, but it could also be in case the system fails. The system must be highly available in order not to negate the convenience it confers on a driver by requiring him/her to resume control too often. Combining highly precise map data with a model of the road would be one way to heighten the availability of the system if, for example, lane markings are missing. Since information on a car ahead is no long part of the equation necessary for calculating a target trajectory, the equation presented in Eq. 3 becomes simpler:

$$\begin{pmatrix} y_{0\_l} \\ \theta_{\Delta\_l} \\ \kappa_{T\_l} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_0 \\ \theta_\Delta \\ \kappa_T \end{pmatrix} \tag{4}$$

The l index indicates that the values for calculating the target trajectory may stem from different sources such as a frontal camera, parking camera, or highly accurate digital maps.

This places great demands on redundancy concepts for capturing surroundings, for functions, and for activating systems. The system must be designed in such a way that no critical situations arise that the car cannot handle itself within the defined take-over time.

The system's speed limit should be set – as in the case of assistance systems – in such a way so as to permit a car to "flow along" in typical stop-and-go situations. At the same time, the speed should not be so high as to place unnecessary demands on the system. Studies of traffic jams have revealed that a maximum speed of 50 km/h permits drivers to comfortably avail themselves of the system (Sandkühler 2002).

4 Interaction of Driver and System

One important aspect of creating such systems – also of assistance-system versions – will be shaping the human-machine interface (HMI). A driver should have a clear understanding of the system status at all times so as to be able to react intuitively in an emergency situation. Systems for recognizing a driver's status (see ► Chap. 37, Driver Condition Detection") will permit dynamic interaction between car and driver in the future. The point in time for a driver to resume control of a car will depend on how much attention the driver is currently paying.

4.1 Human-Machine Interface (HMI)

In contrast to a system that governs purely longitudinal motion (FSRA), adding lateral assistance places additional demands on the HMI. It represents a new mode that has to be available only in certain situations such as a traffic jam or stop-and-go traffic.

4.1.1 BMW Human-Machine Interface

The following is a presentation of the BMW i3 traffic-jam assistant. The purpose of the assistant is to relieve drivers of the need to move longitudinally or laterally when stuck in traffic. The system that BMW offers needs only a mono camera to direct longitudinal and lateral movement. BMW has been offering this system at a comparatively low price in all its cars since 2013.

The steering wheel and the lengthened side bars depict active longitudinal and lateral motion support in traffic here, as in Fig. 8a. If the function is activated and certain other conditions are fulfilled (e.g., map data verifies the highway), but the vehicle speed exceeds the functional range of 0–60 km/h, only the side bars appear (Fig. 8b). ACC is active; lateral guidance is in standby. If the speed drops below 60 km/h, lateral guidance kicks in without any user interaction. If a system limit is exceeded (e.g., if the system detects a hands-off situation) and the lateral guidance system is thus deactivated, an optical signal (red blinking steering wheel in Fig. 8a) and audible information for the driver are generated.

4.1.2 Daimler Human-Machine Interface

Daimler introduced DISTRONIC PLUS with steering assistant and stop-and-go pilot as part of its Mercedes-Benz Intelligent Drive in 2013 in the new S-Class (Daimler 2013) and the facelifted E-Class, the first time that it offered a comprehensive driver-assistance system for longitudinal and lateral motion over the entire range of speed. At the lower speed range, the system follows a vehicle ahead and also orients itself to any lane markings it recognizes (stop-and-go pilot). As the speed increases, the system only reacts to lane markings, whereby it does take account of other traffic and any road restrictions it detects. The lateral guidance system does not switch off at boundary speeds. Instead, it transits seamlessly to a lane-centering system. Lateral guidance support for the entire range of speed works only if longitudinal guidance (DISTRONIC PLUS) has been activated. There is a button which can be used to activate it separately (see Fig. 9). A gray steering-wheel symbol appears in the instrument cluster in addition to the LED display. The steering wheel's color changes to green as soon as lateral guidance becomes active.

4.2 Handover and Controllability

Any technical system designed to guide motor vehicles laterally and longitudinally has to have a way to hand over control of the vehicle back to a driver in case of an

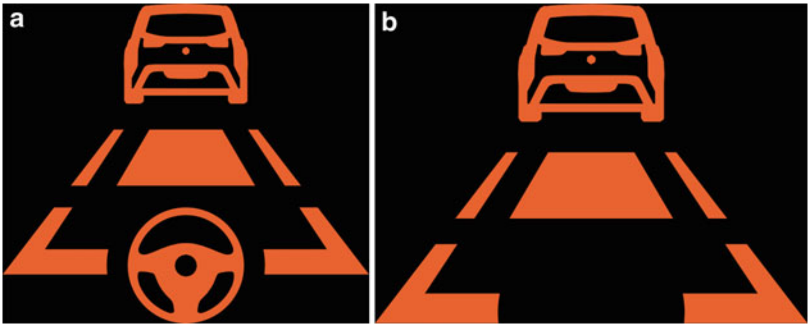
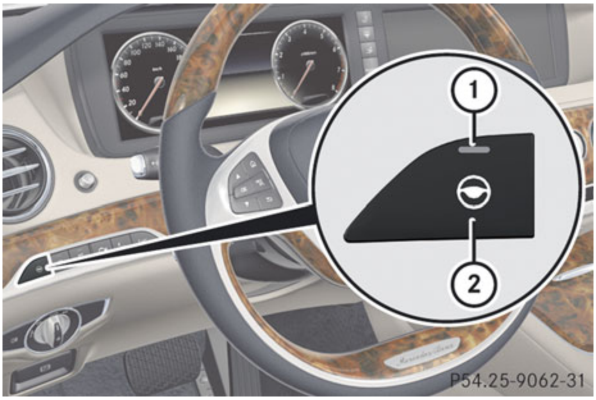


Fig. 8 Traffic-jam assistant (**a**) active and (**b**) available (Source: owner's manual for the BMW i3, 2013)

Fig. 9 Button (*2*) with LED control light (*1*) to switch the steering assistant on and off (Source: owner's manual for the Mercedes-Benz S-Class (BR 223), 2013)



error or a traffic situation that its specifications do not anticipate (system limits). One variable among the various versions of systems is the time that the system takes to transfer control back to a driver.

The need for a handover can occur suddenly in the case of partially automated driver-assistance systems. That is the reason why these systems are designed so that a driver must be present and must constantly monitor the trip so that he/she can intervene when necessary. It is known that a high degree of trust in a system can lead to a driver reacting slower to system limits (Niederée and Vollrath 2009). Trust in the system in turn depends heavily on the perceived reliability of the system (Niederée and Vollrath 2009). Unavailability or experience with system errors, on the other hand, reduce driver acceptance. As the degree of support a system lends increases, the system will continually monitor a driver's attention via such things as the hands-off recognition shown in Sect. 1.3.

Highly automated systems free drivers from the need to constantly monitor traffic within a defined scenario, allowing them to take their hands off the wheel while they pursue other activities (within reason). The length of time that a driver can remove his/her hands from the wheel is a function of the extent and the quality of the scenario covered. When the system hands control of the car back to the driver, the question arises as to how quickly and how well this will occur and what influence the type of side activity will have on the handover.

Zeeb and Schrauf (2014) distinguish between two aspects of the driver resuming control. (1)"Formal" resumption, including the initial intervention. It becomes possible as soon as a driver is ready in the sense of hands on the wheel and eyes fixed on the road. (2) Resumption which is adequate to avoid an accident. It involves cognitive recognition plus the choice of an adequate reaction. The time for adequate resumption depends on whether a driver must completely reorient him-/herself or whether he/she still has a valid mental model of the traffic situation and only has to update it. A study of a critical traffic situation reveals that drivers will take between 1.6 and 2.3 s once they have been requested to resume control, before hitting the brakes.

Aside from the time it takes for a driver to resume control, it is necessary to consider the quality of that resumption and the control that goes with it when it comes to highly automated driving. Experience regarding control is still thin. Studies up till now have addressed normal driving and have examined such things as disturbances in electronic power steering and their controllability (Neukum et al. 2009, 2010).

Damböck et al. (2012) have noted no significant differences in the quality of how a situation is handled exist after a period of 6–8 s following a handover, compared to a group of normal drivers. The group examined performed intense manual visual and cognitive side activities in noncritical traffic situations. The study addressed tasks associated with stabilizing, steering, and navigating. In none of the three variations does the study represent a critical traffic situation. Instead, drivers could use all of the relatively long time made available to them. The influence of a variation in the side activity as observed by Petermann-Stock et al. (2013) led to maximum resumption lags between 2.4 and 8.8 s, regardless of the age or sex of the people tested. Eight different studies by Giesler and Müller (2013) brought to light similar results. The median time for resumption was 2.7 s, while the maximum was likewise 8.8 s.

In general, one can say that the studies of highly automated driving that have been conducted thus far have revealed time lags before resumption of up to around 9 s. In noncritical situations, therefore, the time lag was quite long, whereas in critical traffic situations, very short reaction times of around 2 s were observed. The urgency of the situation obviously exerts a great influence on the time needed to resume control, but it can be to the detriment of quality.

Also interesting is the question how a system reacts if a driver does not respond to a return of control. Deactivating the function is certainly the technically easiest solution here. This is the usual practice with partially automated functions. Depending on the situation, gradually shutting down the function is preferable to abruptly deactivating it. Whether it is better to deactivate the function after the time for resumption has expired or whether it is preferable to perform a minimal risk maneuver is still an open question which surely will depend on the scenario.

Studies focusing on a driver's resuming responsibility for driving, especially in heavy traffic situations, are not known. Due to the low speed, it is not anticipated that drivers will "overreact." Furthermore, it is no problem to go for the safe status of "stop" making it easy to tap it as a solution in critical traffic situations.

Fully automated systems represent the highest degree of automation. They relieve drivers of the necessity of monitoring traffic for a long time. What requirements result for procedures aimed at abandoning fully automated driving, and how long such a handover really takes, still gives rise to a multitude of questions. One need only imagine the case of a system handing over control to a driver who is napping, to put the question into perspective.

4.3 Marketability

4.3.1 The Legal Situation

The same legal challenges and problems of liability present themselves for traffic-jam assistants (and especially for traffic-jam automation) as they do for other assistance systems or automation scenarios. Whereas accidents caused by counterparties should not pose any problem, there is a gray zone between carmakers and customers when it comes to damage caused by technical failures or damage caused by a driver's own fault, for example, by ignoring a command to resume control of the car.

These aspects must still be clarified before introducing and marketing highly automated systems for stop-and-go traffic. The highly automated traffic-jam assistance system is currently not permitted for speeds above 10 km/h (ECE 2006). Refer to ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems" for a more detailed discussion of these topics.

4.3.2 Analysis of Marketability

Compared to implementing highly automated functions at high speeds, the specific automation of traffic-jam scenarios has several advantages that simplify its technical and commercial feasibility greatly. Automating longitudinal and lateral movement at typical stop-and-go speeds of up to around 50 km/h requires less sensor range compared to automation at higher speeds. The system also does not rely on a complicated and expensive back end. The effort needed to ensure function in case of a failure (such as if an actuator fails) diminishes, again due to the short braking distances involved. Stop-and-go traffic on highways, in particular, is relatively easy to describe and comprehend. There are no intersections and no traffic lights (except in the case of tunnels), nor does one normally have to deal with pedestrians or cyclists. The curves on highways, moreover, exhibit a generous radius. These factors reduce the complexity of the scenario and also greatly reduce the demands on sensors and on models of the surroundings. The speeds and positions of surrounding vehicles and the lane markings are enough to regulate the function. A more wide-ranging and detailed detection of maneuvering space is, at least for the simpler versions of the system (see Sect. 3), unnecessary. Whether it will be possible to implement the technical safety concept using the components found in today’s automobiles (sensors and actuators) due to the low speed and the possible effects of a system failure is the subject of debate at the current time.

If it were just a matter of one car and the ones immediately surrounding it, a stop-and-go situation on a highway is hardly distinguishable from traffic situations on other roads or in urban environments (see Fig. 1). Urban traffic, however, presents special challenges for automated systems. In addition to intersections, turn lanes, and traffic lights, there are also nonmotorized participants. A driver in urban traffic must reckon with pedestrians or cyclists moving among cars at any time. Any system that is going to provide safe automation in this kind of environment is going to have to be able to recognize nonmotorized participants reliably and at any time. The system must be able to recognize intersections and other complex scenarios early enough so that drivers have sufficient time to take over when a system reaches its limits. These points place greater demands on sensors, models of the surroundings, and analysis than a highway situation. The result is that a system designed to automate a traffic-jam situation on a highway is not necessarily transferrable to an urban environment. The sensors for a system designed for use on a highway are insufficient for recognizing reliably whether a vehicle is in urban traffic or stuck in a traffic jam on a highway. One possible solution would be to enlist the aid of map data from a common navigation system to assist in differentiating.

This could give rise to a market gap. Customers who have had a positive experience with longitudinal and lateral guidance in traffic situations on highways would like to see expanded functionality, such as on other roads or in an urban environment.

5 Final Remarks

FSRA systems are becoming more popular even in smaller cars. Systems that govern solely the longitudinal motion of a car have achieved a level of quality and availability (see ▶ Chap. 45, “Adaptive Cruise Control”). Systems that govern lateral motion have now established themselves as well (see ▶ Chap. 48, “Lateral Guidance Assistance”). It is now time to take the next step and combine longitudinal and lateral guidance systems into a new, comprehensive system. This could take the form of classic partial automation. However, a highly automated system that permits drivers to pursue non-driving activities would represent a really large step forward in terms of benefits and acceptance. An automatic assistant for keeping in lane while stuck in traffic at relatively low speeds could represent an initial system. The disadvantage of such a system is that it reaches its limits when a certain top speed is attained, something that can happen more or less frequently, depending on how traffic is flowing. This will quickly instill the desire among customers for a comprehensive system covering the whole range of speed. It is not yet possible to establish highly automated systems due to legal and technical reasons. Mercedes-Benz and BMW have already taken an initial step in making longitudinal and lateral assistance systems a reality in 2013 as they introduced lateral guidance functions. This is certainly just the beginning of a long way to automated driving, and the market will witness the introduction of many more systems.

References

Daimler AG (2013) Die Fahrassistenzsysteme: Helfer im Hintergrund (Driver assistance systems: helpers in the background), press release, Stuttgart/Toronto, 02.07.2013

Damböck D, Farid M, Tönert L, Bengler K (2012) Übernahmezeiten beim hochautomatisierten Fahren (Driver overtaking time within highly automated driving). In: 5. Tagung Fahrerassistenz, Munich

Gasser T et al (2012) Rechtsfolgen zunehmender Fahrzeugautomatisierung (Legal aspects of increasing vehicle automation). Report of the German Federal Highway Research Institute no F 83

Giesler B, Müller T (2013) Opportunities and challenges on the route to piloted driving. In: 4th international Munich chassis symposium, Munich

Neukum A et al (2009) Fahrer-Fahrzeug-Interaktion bei fehlerhaften Eingriffen eines EPS-Lenksystems (Driver-vehicle interaction in situations of false interventions of an ESC steering system). VDI reports no 2085, pp 107–124

Neukum A et al (2010) Einflussfaktor Fahrzeug – Zur Übertragbarkeit von Aussagen über die Wirkung von Zusatzlenkmomenten (Influencing factor vehicle – portability of statements over the effect of additional steering torque). VDI reports no 2104, pp 361–374

Niederée U, Vollrath M (2009) Systemausfälle bei Längsführungsassistenten – Sind bessere Systeme schlechter? (System failures in longitudinal control systems – are better systems worse?). In: 8. Berliner Werkstatt Mensch-Maschine-Systeme, vol 22

Petermann-Stock I, Hackenberg L, Muhr T, Mergl C (2013) Wie lange braucht der Fahrer? Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt (How long does the driver take? A analysis of driver-overtaking times from different side tasks during highly automated driving in a traffic jam). In: 6. Tagung Fahrerassistenz: Der Weg zum automatischen Fahren, Munich

Regulation No 79 of the Economic Commission for Europe of the United Nations (2006) Uniform provisions concerning the approval of vehicles with regard to steering equipment, Revision 2

Sandkühler D (2002) Analyse von Stausituationen für die Entwicklung eines Stauassistenten im Rahmen von INVENT (Analysis of traffic jam situations for the development of a traffic jam assist within INVENT), INVENT final report, Forschungsgesellschaft Kraftfahrwesen mbH, Aachen

Schaller T, Schiehlen J, Gradenegger B (2008) Stauassistenz – Unterstützung des Fahrers in der Quer- und Längsführung: Systementwicklung und Kundenakzeptanz (Traffic jam assistance – assisting the driver in lateral and longitudinal control: system development and customer acceptance). In: 3. Tagung Aktive Sicherheit durch Fahrerassistenz

Zeeb K, Schrauf M (2014) Re- vs. Neuorientierung: Situationsgerechtes Blickverhalten beim hochautomatisierten Fahren (Situation-dependant gaze behavior within highly automated driving). In AAET 2014: Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel, Braunschweig, in press

Road Assistance for Commercial Vehicles

52

Karlheinz Dörner, Walter Schwertberger, and Eberhard Hipp

Contents

Abstract

The chapter path guidance assistance for commercial vehicles looks at the differences between commercial vehicles and cars and highlights the specific requirements and solutions for commercial vehicles. The requirements concern both the driver and the technology. The requirements for commercial vehicles are considerably stronger than for cars. High mileages that significantly exceed 100,000 km per year and correspondingly high annual operating hours as well as heavy duty operation are the reason for this. The operating hours are recorded using tachographs which log the driver's time at the wheel and rest periods to provide unbroken monitoring. Assistance and safety systems for commercial vehicles are also prescribed in addition to this device. The necessity of these systems is highlighted with the aid of accident scenarios. The chapter discusses

---

K. Dörner (✉)
Engineering Research Electronics (ERE), MAN Truck & Bus AG, München, Germany
e-mail: karlheinz.doerner@man.eu

W. Schwertberger • E. Hipp
MAN Truck & Bus AG, München, Germany
e-mail: walter.schwertberger@man.eu; eberhard.hipp@web.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_53

lane departure warning (LDW) systems, automatic cruise control (ACC), and the advanced emergency braking system (AEBS). In addition to reliability and safety, fuel consumption is a key criterion for operating commercial vehicles. Consequently, efficient, anticipatory driving strategies have been developed whose function is explained in this chapter. The outlook highlights further developments, such as lane change assist systems, turning assist systems, and platooning.

To complement the previous chapters about lane guidance assistance, this section will go into the special characteristics of lane guidance assistance for commercial vehicles. Heavy trucks, for example, semitrailer tractors, and buses for passenger transport are referred to when using the term "commercial vehicles."

From a statistical perspective, coaches count among the safest means of transport on the roads. However, if there is an accident, there is a significantly higher damage potential due to the considerably larger number of passengers when compared to a car, which has an average occupancy of 1.2 people. With regard to mass in motion, heavy commercial vehicles also have a higher damage potential for accidents than cars due to the kinetic energy in an accident. This applies in particular to transport of dangerous goods.

Passive safety measures for heavy commercial vehicles quickly reach their physical limitations. In contrast to this, active safety systems specifically for commercial vehicles contribute significantly to a further increase in road safety and minimize the results of accidents. This chapter provides an overview of the lane guidance assistance systems available on the market and their specific features for commercial vehicles.

Driver assistance systems make a valuable contribution to increasing road safety. The scientific analysis of accidents involving heavy commercial vehicles conducted together by the Allianz Zentrum für Technik (Allianz Center for Technology) and MAN Nutzfahrzeuge came to these results in the context of the "Safe Truck" project (Daschner and Gwehenberger 2005). Technologies for active, anticipatory safety systems were developed in this project, which was funded by the BMBF (German Federal Ministry of Education and Research). Implemented in commercial vehicles in the future, they should prevent accidents or minimize their consequences.

1 Requirements for Drivers of Commercial Vehicles

The drivers of the commercial vehicles referred to here are professional drivers, as opposed to the drivers of cars. On the one hand, this means that these drivers' workplaces are at the steering wheel of a commercial vehicle; this is also why the term "driver's workplace" is common. The drivers usually drive for approximately 9 h each workday. This illustrates how important ergonomic design is for the driver's workplace and that, for example, an air-conditioning system cannot be

Table 1 Summary of the Commission Regulation (EC) No. 561/2006 with regard to driving and rest periods (it should be noted that this is only an informative compilation for this manual and that drivers must observe all provisions in the respective valid regulation)

Daily driving time	Maximum of 9 h Increase to 10 h is permitted twice per week
Weekly driving times	Maximum of 56 h per week Maximum of 90 h in two successive weeks
Driving time break	At least 45 min after 4.5 h of driving time Division into one period of 15 min followed by one period of 30 min is permitted
Daily rest period	At least 11 h Shortening to 9 h is permitted (three times between 2 weekly rest periods) Division into two periods is possible, but then at least 12 h of rest time must be observed daily; first three, then 9 h of rest time must be taken If there is more than one driver, at least 9 h within a period of 30 h
Weekly rest period	At least 45 h including one daytime rest period Shortening to 24 h is possible, but the following must be met within 2 weeks: Two rest periods of 45 h One rest period of 45 h plus one rest period of at least 24 h (compensation within 3 weeks is required) Weekly rest period must be taken after six 24-h periods

viewed as a luxury for the driver; rather, it maintains the daily condition of the driver and thus contributes to driving safety.

On the other hand, the truck is also simultaneously the living and sleeper compartment for many professional drivers. This is a significant characteristic of living compartment design for trucks that are used in long-haul transport, because only a well-rested driver can safely and competently handle the task of driving. So, in addition to a high-quality bed, good noise dampening is also a considerable factor. Lay-bys are often arranged in a way that requires drivers to park their trucks with the cab facing the motorway. For the driver, the combination of living space and workplace is essential because driving time and rest time are precisely defined (see Table 1).

The drivers' driving time, rest periods, and driving speeds are registered in digital trip recorders (tachographs). With the help of these tachographs, the driver's driving activities can be monitored (see Fig. 1). Due to tight schedules, modern just-in-sequence concepts, and continuously increasing freight traffic volume, the drivers of today are under considerable pressure. Because parking areas and lay-bys for trucks were not expanded in the past to correspond to the rise in traffic, it is not easy for truck drivers to find an available parking area at a suitable time. A further issue is the risk posed by decreasing attention after driving for several hours at constant, steady speeds. Critical situations arise when tired drivers search for a parking area and, due to the lack of parking possibilities, are forced to continue driving or are woken from sleep by the police and prompted to drive off because, out of necessity, they park their vehicle outside of permitted parking areas.

Fig. 1 Tachograph (*left* digital, *right* analog)

Drivers have a great deal of responsibility during goods transport. In addition to on-time collection and delivery (with very short loading times and few breaks for rest), sufficient cargo securing and safe transport to the destination are also relevant in this process. The drivers must safely move trucks and semitrailer combinations with up to 40 t of gross train weight at speeds of up to 80 km/h on the roads. Specific knowledge is required and must be applied to safely lash tall loads onto the truck bed. Mistakes can lead to dangerous situations.

The ability to drive fully loaded trucks on routes with inclines or downward slopes requires both experience and technical driving skills from the driver. Anticipatory shifting is just as necessary as the correct use of continuous and service brakes. In modern commercial vehicles, the driver is supported by automated manually operated gearboxes and maximum speed control functions. In addition to technical driving ability, haulage contractors require an economical driving style and use special analysis tools. With their help, the economic efficiency of drivers' driving style is evaluated.

The results are used, in part, to give drivers a salary-based incentive to drive economically. Because of this, drivers are often subject to direct competitive pressure from their colleagues.

In comparison with car drivers, truck drivers are subject to further boundary conditions: a number of traffic signs are only relevant for commercial vehicles and not for cars. This essentially has to do with the larger dimensions, larger mass, wider turning radius, types of goods transported, and, in comparison with cars, lower specific performance. When talking about specific performance, we refer to the engine power in relation to the gross train weight of the vehicle. Truck drivers have to consciously perceive all of these traffic signs. High property damages can arise if he does not do this, as in the case of passing under bridges (see D, Fig. 2).

Narrow carriageways and local areas also require the driver's utmost attention. In the process, he must take into account that other road users may possibly be unfamiliar with the handling of trucks. For example, if the driver takes a tight right turn that has several lanes, he must pay attention to the vehicles driving to his left. A right turn is only possible if the driver can move over into the left lane to drive into the turn. Also, the visibility within close range of the truck, especially on the right

Fig. 2 Truck accident due to failure to observe the clearance height (Source: Karlsfeld fire service)

side of the cab, has significantly larger areas that cannot be seen directly when compared to cars (see Sect. 2). This is why several mirrors are compulsory for trucks.

In addition, there are a number of truck-specific controls that cars do not have, which will not be further discussed here.

Since the introduction of the directive concerning the initial qualification and periodic training for drivers of road vehicles (*Berufskraftfahrer-Qualifikations-Gesetz* – BKrFQG), professional drivers who transport more than eight people or vehicles weighing more than 3.5 t are obligated to receive further training. The objective is to improve safety on the roads, reduce the environmental impact, and prevent unfair competition within the EU's transport industry. The advanced training consists of 35 training hours and must be repeated every 5 years.

2 Significant Differences Between Trucks and Cars

Cars and trucks are different, both in their economic significance and vehicle technology. The latter applies in particular to drive and brake technology, dimensions, and mass, but also for equipment with safety and assistance systems.

From an economic perspective, acquisition of a truck, as opposed to a car, is always an investment asset. The truck must generate profit for the haulage contractor. This is why life-cycle costs of a truck are so important. In addition to low acquisition costs, low operating costs, high operating hours, high availability, large service intervals, fast service, durability, and high resale value are the decisive points. Many trucks change hands for the first time after 2–4 years of operation. Up to that point, a vehicle has driven about one million kilometers in long-haul transport. This is the equivalent of 200,000–250,000 km mileage per year. The next owner uses the truck for an additional two million kilometers.

The comparison of operating hours between cars and trucks illustrates the higher strain that a commercial vehicle must withstand: If a truck runs for 30,000 operating hours in 10 years, a car runs for 3,000 h in the same period of time. In addition, the service life of the trailer is significantly longer by 20–30 years. At times, this aspect slows down innovation due to the interfaces between the tractor unit and semi-trailer, for example, for the equipment of truck-trailer combinations including semitrailers with ESP or modern brake systems.

Just like the technical durability, investments in assistance and safety systems must also be calculated and contribute to economic profits for the haulage contractor. This is the crucial difference when compared to cars for the successful market introduction of driver assistance systems in commercial vehicles.

Back to the technology: essentially, the significant difference is between truck and car vehicle dimensions and vehicle masses. The maximum permitted dimensions for tractor trailers, semitrailers, and articulated trains are precisely defined and may only be exceeded with special permits. For example, a Euro tractor trailer may only have an articulated train that is 18.75 m long to 4.0 m high, and it may only be 2.55 m wide excluding an exterior mirror. For trucks in Germany, speeds of 80 km/h are allowed on motorways and 60 km/h are allowed on main roads. Powerful car engines are only limited by manufacturers starting at 250 km/h.

In addition to the maximum mass of 40 t, a minimum motorization of 6 HP per tonne is defined by law. In current practice, this is a very low value that is usually substantially exceeded to ensure that trucks make quick headway on inclines. However, the longitudinal dynamics for trucks are significantly lower than for cars. A 40 t commercial vehicle equipped with a 480 HP engine has 12 HP per tonne. In comparison, a 1.5 t midsized car motorized with 12 HP per tonne only has an engine output of 18 HP.

Due to the variety of goods to be transported, trucks, as opposed to cars, have an abundance of bodies, such as box bodies or refrigerated bodies. The different loads a truck transports influence its mass and center of gravity and thus affect the quality of the driving dynamics. For this reason, different manufacturers have developed diverse procedures to identify the respective loads and total vehicle masses. This data is used in internal vehicle control systems (e.g., ESP, cruise control, adaptive cruise control), but also displayed directly to the driver. In this way, he can recognize and avoid overloading as well as adjust his driving style to the load. Up to now, the task of calculating the center of gravity has not been decisively resolved. However, this is crucial to the driving dynamics, because the tipping point depends on the vertical position of the center of gravity. This value is crucial for determining the maximum speed at which the vehicle can take a turn. It is also included in the algorithms that calculate the restoring forces required for damping of the electronic damping system while the correspondingly equipped truck is being driven. The electronically controlled damping system automatically adjusts the damping strength in the truck to the respective load condition, driving situation, and road conditions within milliseconds and enables efficient active roll stabilization.

Several brake systems are available for sufficiently and safely braking trucks. Truck service brakes today are, as a rule, electronically controlled dual-circuit pneumatic brake systems. If the electronic system fails, the pneumatics of the brake system is directly controlled with the brake pedal. In addition, trucks are equipped with different continuous brake systems. Continuous brakes are wear-free, unlike service brakes. Different variants of engine brakes and retarders exist as continuous brakes. The market offers retarder solutions both for engines as well as for gearbox inputs and gearbox outputs. When designing longitudinal control systems, the different types of continuous brakes are to be observed because they demonstrate very different braking and regulation behavior (e.g., with regard to discontinuity, delay times, dependency on gear used, and driving speed).

Gearboxes for trucks usually have up to 16 gears for manually operated gearboxes and up to 12 gears for automated gearboxes. In contrast to the torque converters common in automated gearboxes in cars, trucks with automated gearboxes do not have torque converters; rather, they have an input-side friction coupling that is electronically controlled. The electronic control takes over the tasks of gear and clutch operation from the driver.

The economic boundary conditions, the qualities of the driving dynamics, and the technical data clearly illustrate that safety for trucks is subject to different boundary conditions than cars. For drivers, these factors represent both a great strain due to continuous operating hours in long-haul transport as well as higher stress while maneuvering vehicles that weigh up to 40 t and are 2.55 m wide. A series of electronic safety and assistance systems are available for trucks and buses/coaches to relieve drivers today, such as the electronic stability program (ESP), adaptive cruise control (ACC), emergency brake assist (EBA), or the lane departure warning (LDW) system.

An additional strain for truck drivers is limited visibility. Although the road traffic regulations for goods transport vehicles > 7.5 t require two large main exterior rearview mirrors on both sides of the vehicle, respectively, one wide angle and one proximity exterior mirror as well as a front mirror, the visibility to the rear and on the sides is limited. To improve the visibility for blind spots – depending on the configuration with mirrors and sensors, up to nine blind spots can exist for the semitrailer combination – in the future, different technical solutions should be available: Video cameras at the rear will transmit their images to a monitor in the cab to provide an overview of the space behind the semitrailer. Sensors monitor the distance and relative speed of objects to the side of the vehicle (see Sect. 7).

In coming years, truck drivers will have to accomplish safe maneuvering of their own vehicles in increasingly rising traffic volume. The Institute for Mobility Research in Berlin has projected an approximate 80 % increase in freight traffic capacity in Europe by the year 2025 [pT.-In Germany alone, a doubling of transit volume on the east–west axis is forecast by 2025]. The infrastructure cannot grow at the same pace, which means further demands on vehicle technology and the drivers. If the safety standard we already have is to be maintained or bettered, efforts in all areas of safety are essential at all levels – from the infrastructure through the vehicle to the individual road user.

3 Accident Scenarios

An extensive analysis of accident statistics typically precedes the development of assistance systems. Here, the number and distribution of accidents are checked for respective types of accidents as well as the number of accidents during the last 15–20 years. To the extent possible, a detailed analysis is conducted of the course of events during accidents. In the statistics, the traffic volume is taken into account by comparing the number of accidents with the transport performance in goods transport (see D, Fig. 3).

The haulage capacity of road freight transport increased 82 % from 252.3 to 460 billion tonne-kilometers from 1992 to 2011 (Niewöhner et al. 2004). Despite rising operating hours, the number of accidents involving commercial vehicles that lead to serious personal damages with dead or critically injured road users decreased significantly in the same period of time. In 1992, exactly 1883 fatal accidents were recorded. In 2011, only 889 were recorded. This represents a reduction of 53 %. With 13,345 victims of accidents in 1992 and 7835 in 2011, the number of seriously injured road users represents a reduction of 41 %.

To differentiate the accidents involving commercial vehicles that led to death or serious injury of road users, the Statistisches Bundesamt (German Federal Statistical Office) distinguishes between nine categories (see D, Fig. 4). At 29.5 %, the most common type of accident in 2011 was a rear-end collision with a vehicle in front. A further 17.5 % of accidents can be traced back to collisions with oncoming traffic. This is followed by accidents at junctions with 15.9 %. In 12.2 % of all accidents, the vehicles come off the roadway to the right or left. Collisions with vehicles at the side were less common (9.7 %). Accidents with a stationary vehicle (5.0 %), pedestrians, or cyclists (4.0 %) or another obstacle in the carriageway (0.8 %) are also considerably more infrequent.

When analyzing the data, the actual cause of the accident is to be differentiated from the type of accident. Rear-end collisions can usually be traced back to a safety margin that is too small and nonadapted speed. With 16.8 % (margin) and 10.8 % (nonadapted speed), these are the most common reasons for accidents involving goods transport vehicles (Niewöhner et al. 2004). The high kinetic energy of trucks usually leads to serious accidents: if a truck that weighs 40 t and is driving 90 km/h collides with a stationary obstacle without braking, the energy released is equivalent to approximately 3500 Wh. For a car that weighs 2 t, it would only be approximately 400 Wh at 100 km/h.

When coming off the roadway, two scenarios can be essentially differentiated: coming off due to driving dynamics or slow drifting. Coming off the roadway due to driving dynamics is a typical result of taking turns at speeds that are too high, sudden lane changes because of an obstacle, or slippery roadways. They can often be traced back to an error of judgment with regard to the driving situation by the driver. On the other hand, slow drifting from the roadway is often caused by inattention or fatigue on the part of the driver, as the result of a distraction or dull trips on monotone routes, for example.

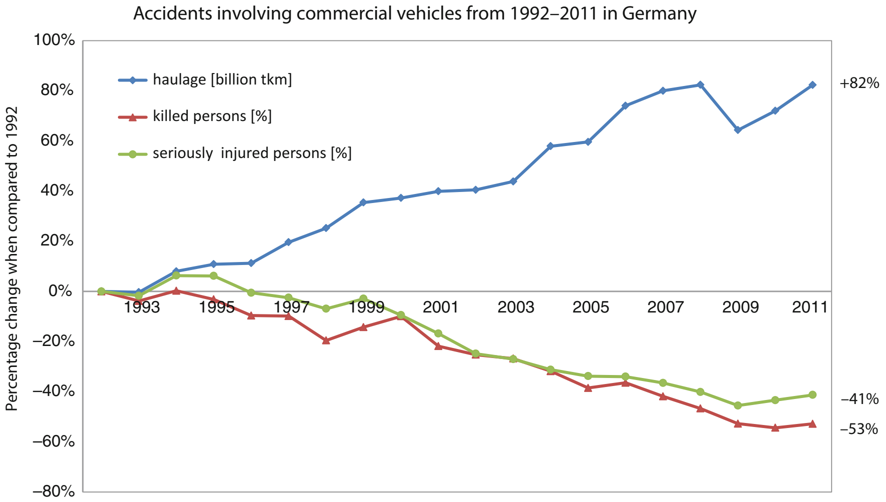


Fig. 3 Transport performance of goods transport vehicles in comparison with accidents leading to death and seriously injured road users in Germany (StBA 2011)

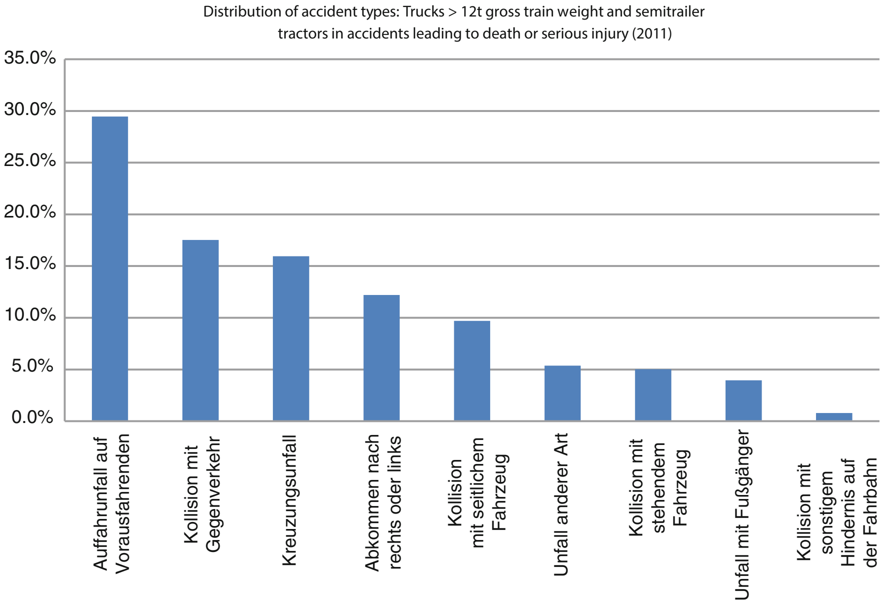


Fig. 4 Distribution of types of accidents leading to death or serious injuries for road users in Germany in 2008. Only collisions involving semitrailer tractors and trucks with a gross train weight greater than 12 tonnes are represented (StBA 2011)

According to the evaluation by the Statistisches Bundesamt (German Federal Statistical Office) with regard to human error on the part of drivers of goods transport vehicles, the most common errors at junctions are errors when turning (17.4 %) and neglecting right of way (7.8 %) (Niewöhner et al. 2004).

Accidents with stationary vehicles and with pedestrians and cyclists can normally be traced back to limited visibility in front of the cab and to the side. A shared study conducted by DEKRA Automobil GmbH (a service provider for the motor vehicle sector) and the Bundesanstalt für Straßenwesen (German Federal Highway Research Institute) analyzed approximately 120 accidents that took place in towns involving trucks (>3.5 t) taking right turns and pedestrians or cyclists (Gwehenberger et al. 2003): In 88 % of cases, initial contact in accidents between commercial vehicles and pedestrians or cyclists is to the side of or immediately in front of the cab. In a further 7 % of all accidents, initial contact took place between the cab and the rear wheel of the tractor.

Control systems for driving dynamics such as the electronic stability program (ESP) and driver assistance systems with sensors that monitor the surroundings such as adaptive cruise control (ACC), emergency brake assist (EBA), or lane departure warning (LDW) can significantly reduce serious truck accidents. A study performed by the Gesamtverband der Deutschen Versicherungswirtschaft (German Insurance Association) together with the Knorr-Bremse Systeme für Nutzfahrzeuge GmbH (Knorr Brake Company) and Munich Technical University investigated the potential of ESP in the case of 850 serious commercial vehicle accidents (Gwehenberger et al. 2003).

With the use of ESP, 9 % of these accidents could be avoided. With regard to single-vehicle truck accidents caused by driving dynamics, approximately 44 % could be avoided with ESP. A joint study analyzing the potential of LDW and ACC by MAN Nutzfahrzeuge AG and the Allianz Zentrum für Technik GmbH (the Allianz Center for Technology) (Daschner and Gwehenberger 2005) also had clear results: If all trucks in Germany were equipped with the adaptive cruise control available today, 71 % of serious end-on collisions involving trucks on German highways could be avoided, and approximately 30 % of serious end-on collisions on all main roads in Germany could be avoided (see Sect. 4). If all trucks warned their drivers with a lane departure warning before they came off the roadway and drivers could take corrective action, 49 % of accidents involving vehicles coming off the roadway to the right or left could be avoided (see Sect. 5).

4 Adaptive Cruise Control (ACC) for Commercial Vehicles

Adaptive cruise control (ACC) is an assistance system that automatically adjusts the driving speed to the speed of cars in front and maintains a distance set by the driver. On a free road, the system operates like a normal cruise control.

Adaptive cruise control initiates both the cruise control and Bremsomat brake control. The cruise control automatically controls the speed of the vehicle using fuel supply in the engine. In this way, the vehicle can maintain a speed entered by the

Fig. 5 Installation situation of an ACC sensor on the truck

driver. However, it is possible for the vehicle to accelerate beyond the desired speed even without fuel supply when driving down a slope due to downhill force. If this is not desired or only to a certain degree, it is possible for the truck driver to activate the Bremsomat function. This automatically triggers the retarder or engine brake if the desired speed or one of the adjustable offsets above the desired speed are exceeded so that a predetermined speed is maintained, even on downward gradients. There are convenient solutions for setting the desired speed, which, for example, create a set value from the current speed if the driver releases the brake after adjusting the speed by braking on a downward gradient.

The adaptive cruise control expands the aforementioned cruise control and Bremsomat functions by measuring the distance and relative speed of vehicles driving in front. This makes it possible to automatically match the speed of the vehicle in front and maintain an adjustable desired distance. The adjustable desired distances depend on the speed. They therefore correspond to an adjustable time interval, which can still be supplemented by constant minimum distances if necessary.

The sensors of the ACC are based on high-frequency RADAR. The RADAR system is usually installed in the bottom section of the front apron and records vehicles driving in front (Fig. 5). The same RADAR sensors are used here that are used in cars (see Sect. 3). However, with regard to the tracking and specific truck boundary conditions, sensor adjustments are required, e.g., with regard to pitch behavior of the truck, driving dynamics parameters, driving in caravans behind trailers with fluttering rear canopies, vibrations typical for trucks, 24 V voltage supply, etc.

The adaptive cruise control intervenes in the interaction of engine components and brake systems for distance and cruise control. In contrast to ACC systems in cars, different brake systems must be triggered for vehicle deceleration in trucks. First, the wear-free continuous brakes such as the engine brake and retarder are always triggered. In the process, their transmission behavior is to be observed, which in part demonstrates a stepped response characteristic and large response delays. To compensate for these undesired effects, there are solutions that provide a brief intermediate quick triggering of the service brakes, creating constant brake torque with a faster response time.

If the brake output of the continuous brake is not sufficient to decelerate the vehicle according to the controller specifications, the service brakes are also triggered. However, this triggering must be limited with regard to the brake energy released as heat to prevent overheating of the service brakes. In this respect, service brakes are only triggered for braking adjustments if the speed of the vehicle must be reduced quickly. As a result, the maximum triggered vehicle deceleration with current adaptive cruise control systems is approximately -3 m/s^2. If a vehicle must be decelerated over a longer period on a downward gradient, this should be done using only the continuous brakes and not with the service brakes to prevent overheating. To this end, the vehicle must be decelerated to a slower speed using the service brakes if necessary, alternating with switching to lower gears so that the continuous brake performance is sufficient.

Depending on the default – set by either the driver or the system manufacturer – ACC systems function with or without driver takeover requests. This can signal to the driver that adaptive cruise control is triggering the maximum deceleration of, e. g., -3 m/s^2, but that this is insufficient in the current driving situation. The driver is then prompted to brake more than the ACC system can. There are also ACC systems with rear-end collision warnings that are, in part, also active when the adaptive cruise control is turned off. This should signal the driver that there is immediate danger of a rear-end collision accident and prompt him to brake.

ACC systems are deactivated by pressing a control (e.g., button or lever) or by displacing the brake pedal. On the other hand, displacing the accelerator leads to an override of the ACC systems. The driver can use this to avoid ACC braking behind a truck that slows down at the beginning of an incline, for example. Because today's ACC systems still cannot provide route previews, only the driver can recognize situations in which braking for a vehicle in front, e.g., due to the beginning of an incline, is inexpedient. Furthermore, the override also serves to decrease distance before overtaking or for faster acceleration.

The following points are particularly relevant for the adaptive cruise control:

- On German highways, trucks must, by law, maintain a minimum distance of 50 m at a speed of 50 km/h. This must be maintained by at least one of the selectable distance levels.
- If the distance drops below the desired distance, e.g., due to a merging vehicle, differential speeds of two to four km/h are usually set to increase the distance again. This differential speed is to be present from the start for overtaking vehicles so that the truck can constantly continue driving with adaptive cruise control. A "passing back effect," which is sometimes feared by laypeople, does not occur.
- If the differential speed of merging vehicles is faster than the entered value for creating the desired distance, an undesired "pulling effect" can occur during controlled driving. In other words, the truck accelerates behind the merging

vehicle. However, because the merging vehicle either has to soon reduce its speed or exit the lane again due to slower vehicles in front (so-called through drivers, e.g., at motorway exits and entrances), these situations can correspondingly be calculated in the ACC system and taken into account to avoid "pulling."

- In addition to the distance and speed of the vehicle in front, its acceleration is also significant for adaptive cruise control. The acceleration can be derived from the speed, although the gearshifts can cause brief, considerable changes in acceleration. Consideration of acceleration is essential if, for example, a slower car merges into the lane in front of the truck at a motorway entrance. Without acceleration of the merging vehicle, the truck would have to brake. However, if there is sufficient acceleration, the differential speed becomes positive before the distance becomes critical. In this case, the truck can continue driving constantly.
- In the distance control strategy, vehicles that are in front of the vehicle in front or vehicles in neighboring lanes can also be included, in addition to the vehicle directly in front.
- The distance regulation behavior represents a compromise between maintaining the desired distance and economical driving style. Exact maintenance of the desired distance would mean that, if necessary, the brake systems would have to react immediately to the deceleration of vehicles in front. This contradicts economical driving style, which strives to use the brakes as little as possible.

Today's ACC systems for commercial vehicles are designed for driving on motorways and well-developed main roads. The driver must deactivate the system on main roads that are not as well developed, on secondary roads, and in urban traffic. The regulation of distance and driving speed with adaptive cruise control occurs starting at a minimum speed prescribed by the manufacturer. A typical value for this is 25 km/h. If the speed falls below this minimum speed, the driver must take over longitudinal control again. There are some ACC systems available that brake to a standstill and can also be used in stop-and-go traffic. Automatic restarting in stop-and-go traffic only occurs if a standstill period of, e.g., 2 s is not exceeded. Otherwise, the driver has to operate the controls to drive off. Many current ACC systems do not yet react to stationary objects – even vehicles at the end of a traffic jam. Vehicles that are driving very slowly may also be interpreted as being stationary objects and are not recognized as vehicles driving in front. These are typical situations where the driver has to intervene.

Adaptive cruise control must satisfy different requirements in trucks and coaches. Trucks often drive in longer queues with a constant speed of 80 km/h. For truck drivers, adaptive cruise control is primarily a convenient function that first and foremost makes their job easier when traffic is largely constant. Driving times are usually long and this helps maintain driver performance for a longer period of time. The automatic distance control increases traffic safety and sudden braking situations due to distances that are too narrow or driver inattention are avoided. For this reason, adaptive cruise control is the preferred order from freight specialists for trucks transporting dangerous goods in particular, not least due to corresponding requirements from shippers.

In contrast, a coach driving at an average speed of 100 km/h can easily overtake a truck but is usually slower than a car. In comparison with trucks, coaches do not usually drive in caravans with regulated distances. However, if the bus approaches a slower vehicle, the adaptive cruise control initiates speed throttling so that a safe distance to the vehicle in front is maintained. This is why the safety aspect is in the foreground for coaches.

The effectiveness of ACC systems for avoiding accidents was investigated by the Allianz Zentrum für Technik GmbH (Allianz Center for Technology) in the context of the “Safe Truck” project, which was sponsored by the BMBF (German Federal Ministry of Education and Research) (Daschner and Gwehenberger 2005). Out of 583 analyzed accidents, 127 were relevant for adaptive cruise control, in other words, rear-end collisions on the same carriageway. Accidents in urban traffic and on secondary roads as well as accidents with stationary obstacles are also included therein. The potential was consequently analyzed for five scenarios that represent different development stages of ACC systems. Furthermore, the scenarios were differentiated by whether the driver intervenes or not. In all scenarios, a maximum vehicle deceleration of -2 m/s^2 with adaptive cruise control was assumed:

- ACC system that only regulates above a minimum speed
  - Without driver intervention (scenario 0)
  - With driver intervention with maximum delay (6 m/s^2) after 2 s (scenario 1)
- ACC system that controls up until standstill and is also suitable for local urban traffic
  - Without driver intervention (scenario 2)
  - With driver intervention with maximum delay (6 m/s^2) after 2 s (scenario 3)
- ACC system that controls up until standstill is suitable for local urban traffic and also recognizes stationary vehicles (scenario 4)

The study is based on 127 ACC-related accidents involving commercial vehicles. Based on reconstructions of well-documented accidents, the prevention potential of accidents in the individual scenarios was analyzed and projected onto the individual categories (see Fig. 6 – ACC potential):

- If all trucks were equipped with the ACC systems available today, approximately 6 % of all serious accidents involving commercial vehicles could be avoided without requiring driver intervention in the brake system. If the driver intervenes in the brake system within 2 s after the adaptive cruise control intervenes with the maximum possible delay, 7 % could be avoided.
- If all trucks were equipped with ACC systems that regulate up until standstill and are also suitable for local urban traffic, 8 % of all serious accidents involving commercial vehicles could be avoided without the driver braking. If the driver also reacts with full braking within 2 s after the adaptive cruise control intervenes, the prevention potential increases to 17 %.
- If all trucks were equipped with ACC systems that react to stationary vehicles with additional sensors, 21 % of all serious accidents involving commercial vehicles could be avoided.

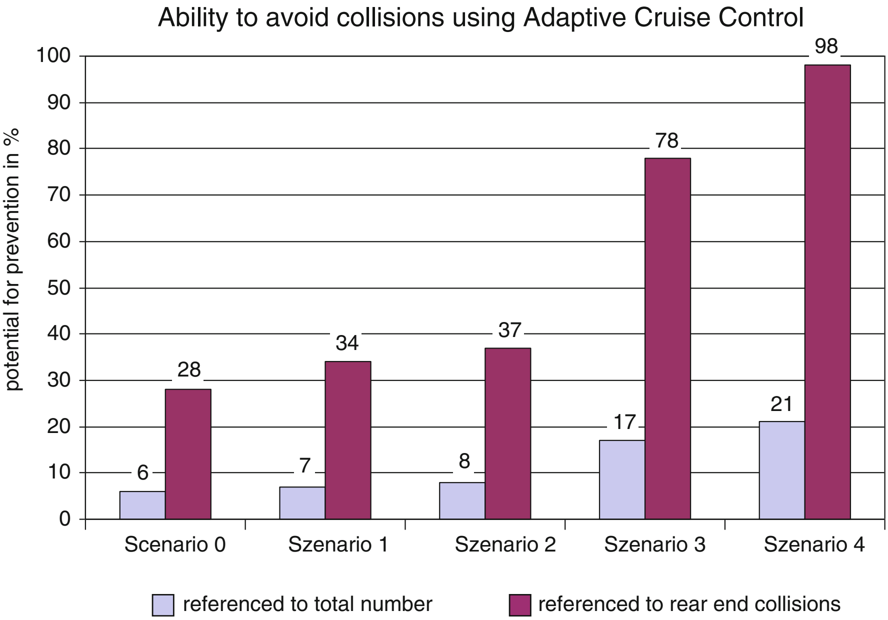


Fig. 6 Ability to avoid collisions using adaptive cruise control (Daschner and Gwehenberger 2005)

Because today's ACC systems are only designed for use on motorways and well-developed main roads, the potential for avoiding accidents was considered separately for this environment. With respect to rear-end collisions involving commercial vehicles on motorways, the results imply that 71 % of these accidents could be avoided if all trucks were equipped with the ACC systems available today. Suppose that the driver recognizes the adaptive cruise control intervention as a haptic warning and initiates full braking, even after 2 s, then 86 % of all rear-end collisions involving trucks on motorways could be prevented.

5 Lane Departure Warning for Commercial Vehicles

Lane departure warning (LDW) for a commercial vehicle monitors whether the vehicle remains in the carriageway and warns the driver if he unintentionally leaves the marked lane. In particular, the system supports the driver on long and monotonous routes if his attention decreases or if he is distracted. An unintentional lane change can be avoided by warning the driver so that single-vehicle accidents due to drifting from the carriageway or collisions with vehicles in neighboring lanes or on hard shoulders can be avoided. Lane departure warning has been offered for commercial vehicles since 2001 and is designed for use on motorways and well-developed main roads.

Fig. 7 Installation situation for a camera in a truck for detection of lane markings [Source: MAN Truck & Bus AG]

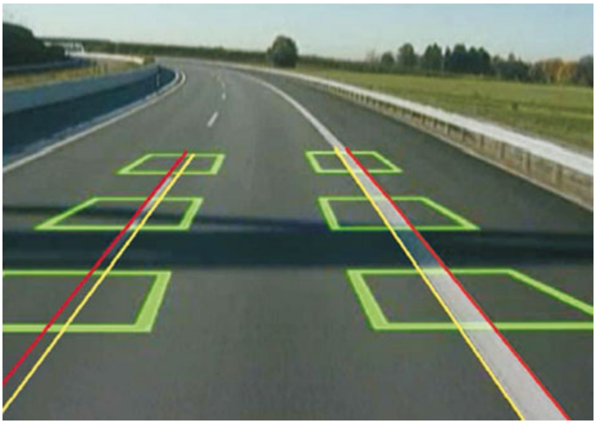
In accordance with the EU regulation (Gwehenberger et al. 2003), equipment with lane departure warning is required for registration of new truck types > 3.5 t and new bus/coach types with more than eight seats starting November 2013. Starting November 2015, lane departure warning for all new registrations of trucks > 3.5 t and buses with more than eight seats is required. Only certain vehicle types, for which this equipment does not make sense, are excluded from this regulation. The systems available for trucks today detect these lane markings using a camera that is installed inside the cab on the windscreen as close to the center as possible (Fig. 7). It is also possible to use mounting positions that are not in the center as long as corresponding adjustments to the parameters are performed on the evaluation algorithm. The camera should be located in the area of the windscreen wiper. A camera like this has a more favorable angle of view to the road surface in a truck than in a car due to the higher mounting position. On the other hand, rolling and pitching of the cab make evaluation of the detected image more difficult.

One of the most common methods used for detecting lane markings is the search for light–dark transitions on the road surface. This is why the cameras used are black-and-white cameras. The sensors can only exactly detect the lane markings when the contrast is high enough, in other words, if the markings can be clearly recognized and are as straight as possible. The illumination from the vehicle's headlights is sufficient for recognition of lane markings in the dark.

Although the camera constantly detects the path of the lane, the analyzing algorithms do not check the entire image. To save computing power, only the outer areas of the road are evaluated with the help of search windows (Fig. 8).

If the system recognizes that the vehicle is approaching the lane marking or even drives over it without activation of the turn signal, a warning is emitted. The warning can occur in the haptic form of a steering wheel vibration, for example,

Fig. 8 Detection of lane markings by the lane departure warning [Source: MAN Truck & Bus AG]



or use a side-specific acoustic signal (e.g., in the form of a simulated rumble strip). The only warnings worth considering for coaches are warnings that are solely perceived by the driver and not by the passengers. This is why there are systems available for coaches that warn the driver using a side-specific vibration in the driver's seat. This avoids upsetting the passengers.

The conditions for triggering a warning can vary depending on the manufacturer, but must be sufficient for the requirements stipulated in the EU regulation. For example, a warning can be triggered depending on the driving speed when driving over the inside or the outside of the lane marking. The transverse speed at which the vehicle exits the lane can also be taken into account. Beneath a minimum vehicle speed, e.g., 60 km/h, today's systems normally do not issue a warning. Systems like those that are currently available on the market do not actively intervene in steering; rather, they only warn the driver.

To prevent false warnings, the sensors for detecting lane markings have strict limits. As a rule, a warning is not issued in the following situations: if the windscreen is very dirty in the area near the sensor, a snowy, dirty, or repaired carriageway; if there are several markings next to and in front of each other – as they primarily occur at entrances and exits near construction areas; and if the carriageway is wet. If there are grooves filled with rainwater on the carriageway or snow lines on the road, there is a risk that these structures could be recognized as lane markings. Due to the strong contrast between light snow or the reflective surface of water and dark asphalt, the black-and-white images from the video camera cannot be precisely evaluated. Research is currently working on improved sensors.

While the first system generation required markings on both sides for the assistance function to work properly, development is now going in the direction of also having a function for one-sided lane markings and, if possible, only sufficient contrast to the lane boundary must be present.

To check the effectiveness of lane departure warning, Allianz Zentrum für Technik GmbH (Allianz Center for Technology) evaluated 583 truck accidents from their database. Of these, 44 were relevant with regard to leaving the lane unintentionally. During the analysis performed in the context of the BMBF

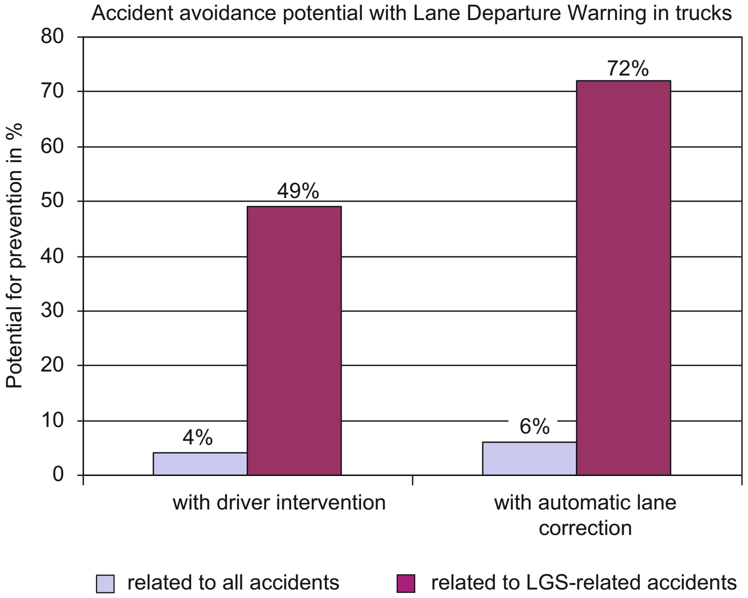


Fig. 9 Accident prevention potential with lane departure warning in trucks (Daschner and Gwehenberger 2005)

(German Federal Ministry of Education and Research) "Safe Truck" project, two system characteristics with different ranges in function were considered (Daschner and Gwehenberger 2005):

- Lane departure warning systems available today with driver warning starting at a driving speed of 60 km/h and an assumed intervention in steering on the part of the driver after 1 s of reaction time.
- An expanded system that is also designed for speeds starting at 60 km/h but, in addition, also performs an automatic correction if the lane is exited.

The results of the study show that 49 % of all accidents involving commercial vehicles that leave the carriageway could be prevented if all commercial vehicles were equipped with lane departure warning. If, in the future, automatic correction for returning to the carriageway is performed, then even 72 % of these accidents could be prevented (Fig. 9).

6 Emergency Brake Systems

Assistance systems that automatically initiate full braking have now become established in the commercial vehicle market. According to the EU regulation (VERORDNUNG Nr. 661 2009), emergency brake systems are a prerequisite for

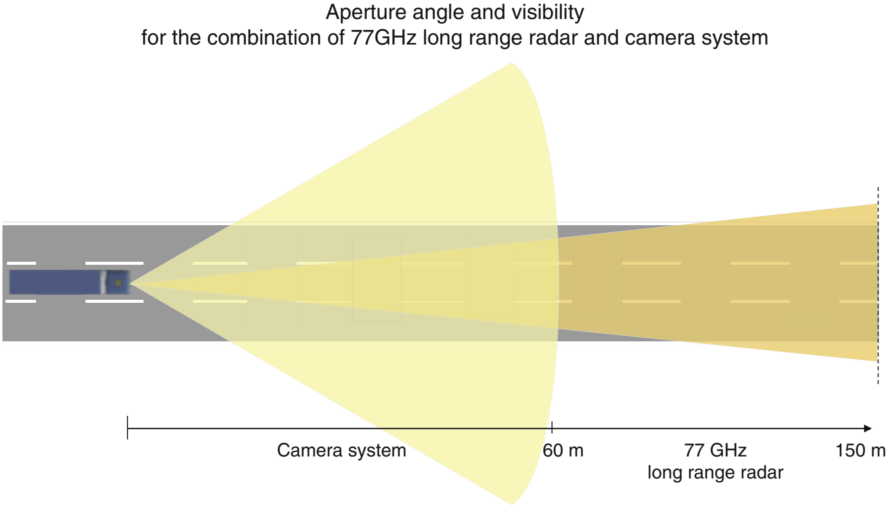


Fig. 10 Aperture angle and visibility for the combination of 77GHz long-range RADAR and camera system [Source: MAN Truck & Bus AG]

new type approvals for trucks > 3.5 t and buses with more than eight seats since November 2013. Starting in November 2015, emergency brake systems are a prerequisite for all new approvals of trucks > 3.5 t and buses with more than eight seats.

These systems urgently warn the driver if there is acute danger of a rear-end collision and, if necessary, automatically initiate full braking if the rear-end collision is inevitable. This prevents rear-end collisions and significantly reduces the severity of accidents if a collision is not avoidable. To do this, the emergency brake assistance must be designed for all traffic situations; this means that it must not trigger unnecessary full braking in any traffic situation – after all, the driver cannot be expected to, e.g., turn the system off in time before entering inner-city traffic if it is not designed for that and would initiate false braking.

The basis of today's emergency brake systems are high-frequency RADAR sensors like the ones used in ACC systems. They detect traffic situations ahead. Evaluation of the traffic situation is performed with special algorithms that generate system reactions with one or more steps. In the process, the challenge is to ensure that false braking is not initiated and that critical traffic situations are correctly recognized (Fig. 10).

If the sensor detects an obstacle and recognizes at the same time that the distance is closing and the driver is not reducing the speed, the emergency brake system intervenes in the driving. First, the driver is made aware of the danger with a visual signal in the center display and an acoustic warning signal. If the assistance system still does not register a reaction from the driver – like an intervention in the brake system or a steering maneuver – partial braking with vehicle deceleration of approximately $-2\ \mathrm{m/s^2}$ is performed. If the risk of a collision increases, the system initiates full braking with vehicle deceleration of approximately $-6\ \text{m/s}^2$. If braking occurs, the brake lights are triggered to warn traffic behind the truck and avoid resulting accidents. The particular objective of this function is to avoid full-speed collisions with slower vehicles and late braking by the driver.

Recognition of stationary obstacles for which emergency braking is required is much more difficult than recognizing moving obstacles. This was also taken into account in the implementing provision for the introduction of emergency brake systems. The regulation stipulates that, assuming a truck speed of 80 km/h with a vehicle in front traveling at 30 km/h, emergency braking must be performed to avoid an accident. This means that the deceleration of the truck must be at least 50 km/h. In contrast, the speed must only be decreased by 10 km/h for a stationary obstacle.

Because an active emergency brake system directly intervenes in vehicle control, as opposed to an emergency brake warning, the system and the development process must correspond to increased safety requirements.

The interpretation of data from the Statistisches Bundesamt (German Federal Statistical Office) about accidents involving goods transport vehicles on the roads in 2011 (Niewöhner et al. 2004) illustrates the kind of potential that emergency brake systems have: In nearly 17 % of all accidents caused by goods transport vehicles, errors with regard to the distance to vehicles in front were the most common cause of accidents. In addition, the high kinetic energy involved in rear-end collisions with commercial vehicles usually leads to serious accidents: Active and warning emergency brake systems are in a position to mitigate these critical situations.

7 Predictive Driving

An important criterion for operation of commercial vehicles is their fuel consumption. This can be considerably reduced with predictive driving. Two approaches are followed here.

On the one hand, the driving strategy can take the data from the route ahead from digital maps and take these into account in the driveline control. There are systems on the market that evaluate the incline profile ahead and integrate it into the driving strategy. In the process, the objective is always to let the vehicle roll without fuel injection whenever possible or to open the driveline in specific situations to save fuel. Typical situations, in which rolling out makes sense, are existing before a downward gradient where the truck has to brake. While a normal cruise control maintains a constant speed until the downward gradient begins and the vehicle consumes fuel in the process, a predictive cruise control takes the downward gradients ahead into account and reduces the fuel injection in time so that the vehicle rolls to the downward gradient, thus slowing down slightly before the downward gradient begins. The control calculates ahead of time when the fuel injection is reduced to maintain parameters from the manufacturer or a speed reduction set by the driver before a downward gradient. The amount of the tolerated speed reduction influences the fuel saving proportionally and creates a compromise between efficiency and acceptance. Ultimately, the truck may not become a traffic obstruction due to its parameterized speed reduction before a downward gradient, and the behavior must also be acceptable for the truck driver so that he does not override the system. In doing so, he would work against the desired increase in efficiency. Driver override is possible at any time, of course. Because a truck operated in a predictive manner enters downward gradients at slower speeds, it can also brake later which, in turn, provides the benefits of less wear and improved heat balance.

While a normal cruise control maintains a constant speed up to the end of a downward gradient and slows down trucks on the downward gradient using the continuous brake, a predictive cruise control takes into account the approaching end of the downward gradient and triggers the continuous brake at the right time so that the truck gains a bit of speed and thus gains momentum. This is an advantage if it comes to an adjacent incline or also a subsequent level in which the engine torque must be built up slowly in order to maintain the desired speed. The resulting speed increase at the end of the downward gradient can be calculated exactly by the controls ahead of time and represents, in turn, a compromise between acceptance and efficiency, in which the legal boundary conditions must also be observed with regard to permitted maximum speed. When building up momentum at the end of a downward gradient, the time loss caused by rolling at the beginning of the downward gradient is also compensated.

In addition to the predictive interaction of engine components, the route preview is also taken into account in the gearbox control. In this way, for example, switching gears before and on an incline can be improved when compared with previous, non-predictive systems.

Modern commercial vehicles not only include predictive automated driving strategies, they also include onboard driver training, which also shares knowledge to provide drivers with eco-training in accordance with the directive concerning the initial qualification and periodic training for drivers of road vehicles. The onboard driver training differs depending on the manufacturer, but they all have the same goal: to teach the truck driver an anticipatory, efficient, and material-friendly driving style. To do this, the system analyzes how the handling can be improved with regard to fuel consumption and minimization of wear and gives the driver corresponding information on the vehicle display.

8 Development for the Future

Current driver assistance systems support drivers in exactly defined traffic situations. Lane departure warning monitors the vehicle position within the carriageway while adaptive cruise control maintains the correct speed and distance to the vehicle in front. Each assistant works independently as an individual system. Safety assistants in the future, on the other hand, will cooperate with each other and merge into integrated systems.

In the future, adaptive cruise control systems will have stop-and-go functionality. Image processing will also be used increasingly in applications for commercial vehicles.

Lane departure warning will also be further developed into versatile lateral guidance systems. These could intervene in lateral guidance if the driver does not react to the lane departure warning. Active intervention like this, e.g., with systems that superimpose torque in the steering or in the form of targeted independent wheel braking, is possible.

In the future, lane change assistants could tell the driver whether everything is clear to overtake or change lanes quickly because of an obstacle. If the driver activates the turn signal and the system detects a vehicle approaching from behind, he is warned with a red signal in the exterior mirror and a corresponding display in the center display, for example. Combined with a camera for lane detection, the system can also emit a warning if the driver changes lanes without activating the turn signal. In this case, a lane departure warning can also occur depending on the risk of a lateral collision or an automatic correction in lateral guidance can be performed.

Turning assistance systems are currently under development to protect pedestrians and cyclists in the vicinity – immediately in front of and to the side of the truck. To do this, sensors detect the surroundings in front of and to the side of the truck. The driver can then be warned if there is a risk of collision with cyclists or pedestrians.

A cooperative form of lane guidance was under development until the end of 2009 in the KONVOI project, which was sponsored by the BMWi (Federal Ministry for Economic Affairs and Energy). Universities, freight forwarding companies, and corporate research departments from the commercial vehicle industry evaluated the "convoy truck" traffic system on motorways in real traffic under everyday conditions. The project is technologically based on sensors, actuators, communication technology, and algorithms for longitudinal and lateral guidance that were developed in past national and European projects like Prometheus, INVENT, and Chauffeur. With the help of driver assistance systems, trucks are electronically coupled to each other. Longitudinal and lateral guidance systems regulate the distance to vehicles in front as well as the vehicle position within the carriageway. An organization assistant integrates the potential convoy participants and helps the drivers while creating a convoy. In the project, possibilities for optimizing traffic flow and improved utilization of existing infrastructure were investigated. Furthermore, it could be demonstrated that driving in a convoy can provide both improved safety and fuel efficiency.

The idea of the convoy was pursued within the SARTRE project (September 2009 to December 2012), which was sponsored by the EU. Here, the goal was to realize a mixed convoy of trucks and cars.

Additional possibilities for the improvement of traffic safety and traffic flow present themselves with prospective vehicle–vehicle communication and vehicle–infrastructure communication. Due to the expected quantities, these developments are primarily driven by the car industry to begin with but will also find application in commercial vehicles. In this way, anticipatory and safe driving will be made possible far beyond the driver's field of vision.

References

Daschner D, Gwehenberger J (2005) Wirkungspotenzial von Adaptive Cruise Control und Land Guard System bei schweren Nutzfahrzeugen. Allianz Zentrum für Technik GmbH, Bericht Nr. F05–912, im Auftrag von MAN für das BMBFProjekt Safe Truck, München, Report article: [Potential of adaptive cruise control and lane guard system reducing heavy truck accidents, report nr. F05-912, by order of MAN for funded BMBF project Safe Truck, Munich]

Gwehenberger J, Langwieder K, Heißing B, Gebhart C, Schramm H (2003) Unfallvermeidungspotenzial durch ESP bei Lastkraftwagen. ATZ Automobiltechnische Zeitschrift 105, 504–510, Report article: [potential of accident avoidance by ESP oft rucks]

ifmo (Institut für Mobilitätsforschung) (Hrsg) (2005) Zukunft der Mobilität – Szenarien für das Jahr 2025. Erste Fortschreibung. Eigenverlag, Berlin, Report article: [ifmo institute for mobility research: future of mobility – scenarios 2025]

Niewöhner W, Berg A, Nicklisch F (2004) Innerortsunfälle mit rechtsabbiegenden Lastkraftwagen und ungeschützten Verkehrsteilnehmern. DEKRA/VDI Symposium Sicherheit von Nutzfahrzeugen, Neumünster. Report article: [urban accidents of right turning trucks and vulnerable road users, DEKRA/VDI symposium safety of trucks]

StBA, Statistisches Bundesamt (2011) Verkehr. Unfälle von Güterkraftfahrzeugen im Straßenverkehr. Statistisches Bundesamt, Wiesbaden, Report article: [census bureau: traffic, accidents of freight transports in road traffic]

VERORDNUNG (EG) Nr. 661/2009 DES EUROPÄISCHEN PARLAMENTS UND DES RATES vom 13. Juli [decree of European Parliament and European Council of July 13th 2009]

Assistance Systems in Agricultural Engineering

53

Marco Reinards, Georg Kormann, and Udo Scheff

Contents

Abstract

Agricultural vehicles are complex machines which are used in a wide range of applications. To improve system performance and support operators a number of driver assistance systems have been developed and introduced in the recent past. This article describes a selection of driver assistance systems for agricultural tractors. The main focus is on handling assistance systems, process assistance systems, and automated steering systems which are state of the art in the

M. Reinards (✉) • U. Scheff
John Deere Werk Mannheim, John Deere GmbH & Co. KG, Mannheim, Germany
e-mail: reinardsmarco@johndeere.com; scheffudo@johndeere.com

G. Kormann
Advanced Engineering, John Deere GmbH & Co. KG, Intelligent Solutions Group, Kaiserslautern, Germany
e-mail: kormanngeorg@johndeere.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_54

agricultural industry today. The article closes with an outlook on highly automated vehicles, which is one focus area right now in the industry to further automate working processes.

1 Introduction

In the past 50 years the maximum travel speed of agricultural tractors has almost tripled. At the same time the vehicle weights and installed engine power has grown which again has influenced the vehicle size. The increased weight in combination with higher road and field speeds drive requirements in terms of vehicle dynamics which are addressed by the development and implementation of enhanced chassis concepts (Moitzi 1997).

Agricultural tractors are similar to on-road vehicles. The movement on a given surface can be controlled by the operator in lateral, longitudinal, and vertical direction within the given physical limits. In this general description of drive dynamics it is assumed that all external forces as torques, with the exception of gravity and aerodynamic forces are transferred through the surface contact area between tire and driving surface and with that demand the movement of the vehicle (Braess and Seifert 2000).

2 Assistance Systems

The transport of passengers and goods is the primary functional task of vehicle motion for on-road vehicles. Tractors, respectively agricultural vehicles and construction equipment, usually have to accomplish several additional functions simultaneously – such as provision and control of mechanical, hydraulic, or electrical power, cargo handling performance, and traction performance. These additional requirements arising from the integration into the process of agricultural production have a significant impact on the design of the agricultural vehicles. This also includes the design of the human-machine interface in order to assist the driver in vehicle guidance and process monitoring.

In general, two types of driver assistance systems can be identified:

- Handling assistance systems,
- Process assistance systems.

Handling assistance systems emphasize vehicle motion control – similar to on-road vehicles – with limited integration into an agricultural process. In contrast, process-assistance systems focus on automation solutions to support the driver during the execution of tasks throughout agricultural production.

Fig. 1 Size comparison between tractor and on-road vehicles

2.1 Handling Assistance Systems

Figure 1 displays a size comparison between a 150 horsepower standard tractor, a mid-size passenger car, and an 18 t truck.

Even though this comparison only depicts a small portion of the universal application spectrum of a tractor, two main differences in the design characteristics of the vehicles become apparent:

- Loading respectively ballasting outside the wheelbase,
- Very large tires on short wheel base.

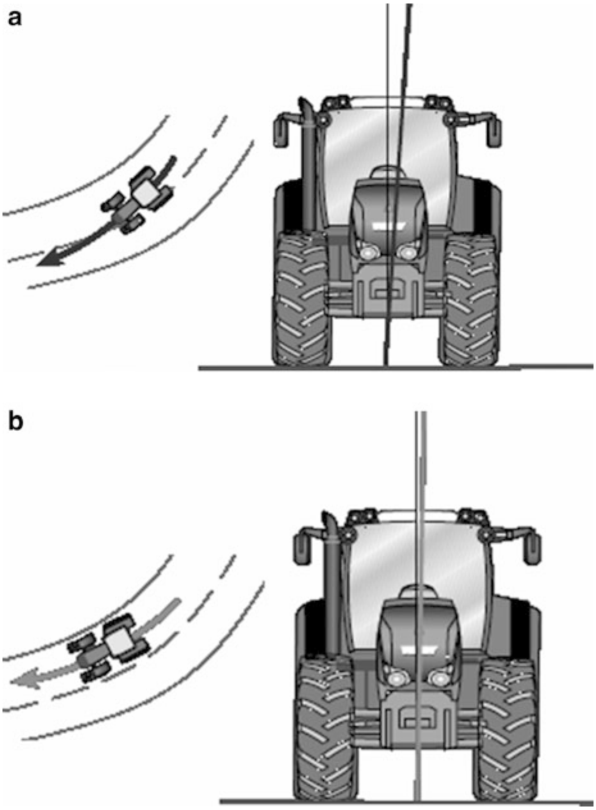
In addition to these obvious differences, the design concept of a standard tractor suspension with a rear axle that is rigidly mounted to the chassis and an oscillating beam type front axle has a significant influence on the longitudinal, vertical, and also the lateral dynamics of tractors.

Lateral Dynamics All roll stabilization of the vehicle is realized by the rear axle, more specifically the rear wheels, because the rear axle is rigidly mounted to the chassis and the front axle can oscillate. This leads to over- or understeering vehicle characteristics dependent on the ballasting situation, especially during transport applications. Aside from improvements in tire technology, steering system tuning, and the introduction of suspended front axles, three important systems for driver assistance systems of lateral vehicle dynamics were introduced over the past years:

- Front axles suspension with switchable roll stabilization
- Switchable steering ratios,
- Steer-by-wire steering systems with lateral dynamics control and variable steering ratio.

The utilization of front axles with hydro-pneumatic suspensions offers various possibilities for hydraulically coupling the suspension cylinders and also

Fig. 2 With and without Fendt stability control



possibilities to design front axle suspensions with or without hydraulic roll stabilization. The general concept of a front axle suspension in tractors consists of two hydraulic cylinders with ring room and piston room cavities connected in parallel manner. This allows oil volumes to freely move during oscillation of the front axle without obstruction of axle motion. For concepts with hydraulic roll stabilization various designs, for example cross linking or decoupling of piston and ring room, can be implemented. The additional oil volume displacement at the hydro-pneumatic suspension can be used for stiffness and damping purposes (Bauer 2008).

To improve the vehicle dynamics during road transport the tractor manufacturer Fendt introduced with the development of the 936 Vario the Fendt Stability Control (FSC), which is dependent on vehicle speed adds supplementary roll stabilization at the front axle (see Fig. 2). For tractor models with a maximum design wheel speed of 60 km/h, the connection between the suspension cylinders is changed for speeds higher than 20 km/h and with that a roll-stabilization and a changed vertical spring stiffness is imposed. By this switch of suspension characteristics the handling behavior is changed without driver interaction in order to improve the vehicle handling especially at higher speeds (AGCO Fendt 2007).

In addition to switching the suspension characteristic to vary the roll stabilization, manipulation of the steering ratio offers another technical option to support the driver at vehicle handling. Tractors respectively agricultural vehicles as well as construction equipment mainly utilize so-called auxiliary powered or power steering systems without a mechanical steering linkage. This mechanically disconnected concept, where the mechanical linkage is replace by hydraulic or electric components, additional options for control and therefore assistance in vehicle handling are generated (Dudzinski 2005; Hesse 2008).

In addition to the hydraulic power supply and the steering cylinder, the hydrostatic steering unit – which usually consists of a proportional, mechanical rotary slide valve, and metering unit – the essential element for the transformation of the steering wheel movement into a wheel movement. As standard design for tractors a steering ratio of about 14:1 is selected which corresponds to four to five turns of steering wheel motion for steering from steering stop to steering stop. In this case, the maximum steering angle at the wheel can vary depending on the equipment, tires, and application. The steering ratio, which is determined by the volume flow, provides a particularly good possibility for a controlled volume flow gain as a function of steering wheel movement by electro-hydraulic systems in parallel. The independent flow gain and thus independent of the driver's steering input motion is described in more detail in the section Process Assistance Systems (Sect. 3). Figure 3 shows the parallel arrangement of an exemplary electrohydraulic proportional valve to a steering unit – also referred to as electro-hydraulic summation steering system.

Through a parallel arrangement of the mechanical and electrical flow metering basically variable steering ratios can be realized. In real applications, systems such as the Fendt Vario Active Steering are offered. Through a by the driver pre-activated reduction to half of the steering wheel motion respectively a doubling of the steering ratio is enabled depending on the driving speed. This allows in particular for shunting and cargo handling a stress reduction for the driver (Wiedermann 2012).

A logical development of the electro-hydraulic parallel systems for flow metering is the introduction of a steer-by-wire system with complete integration of electrical transmission elements in a power steering system. With the aim to improve the steering effort and the vehicle handling and thus to reduce driver fatigue, John Deere introduced a complete steer-by-wire steering system in some tractor series. With this under the name active command steering (ACS) introduced steering system; the following characteristics of the lateral dynamics of assistance have been implemented (Schick and Kearney 2010):

- Dynamic steering angle control,
- Variable steering ratio,
- Prevented steering play and steering creep,
- Variable steering effort.

At the dynamic steering angle control yaw rate sensor measures the yaw motion of the vehicle. The system can automatically make small steering adjustments and

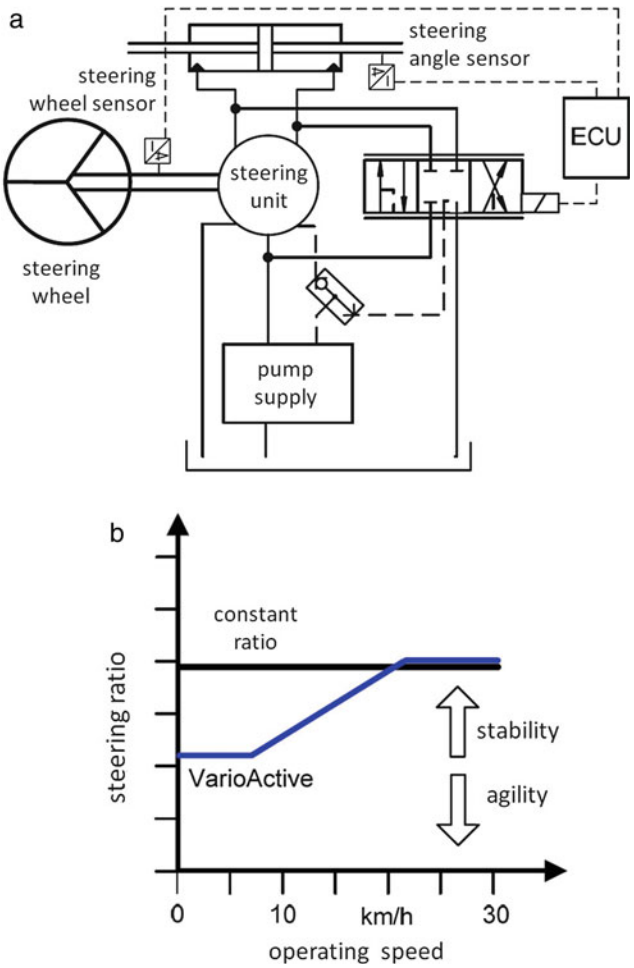


Fig. 3 (**a**) Shunt circuit of steer unit and electro-hydraulic proportional valve (**b**) Relation between steering ratio and operating speed at VarioActive

generate very good lane keep or tracking (Fig. 4). Thus, the vehicle handling is improved in rough terrain on one hand and on the other hand in transport application over-steering due to rapid steering motion is avoided.

Similar to the previously described systems with combined mechanical and electrical volume flow control, the use of the steer-by-wire concept allows a variable steering ratio. At the concept, executed by John Deere, a vehicle speed dependent and continuously adapting steering ratio is implemented. This concept requires at low vehicle speeds about three and a half revolutions at the steering wheel for the entire steering range and at transport operation speeds about five turns of the steering wheel.

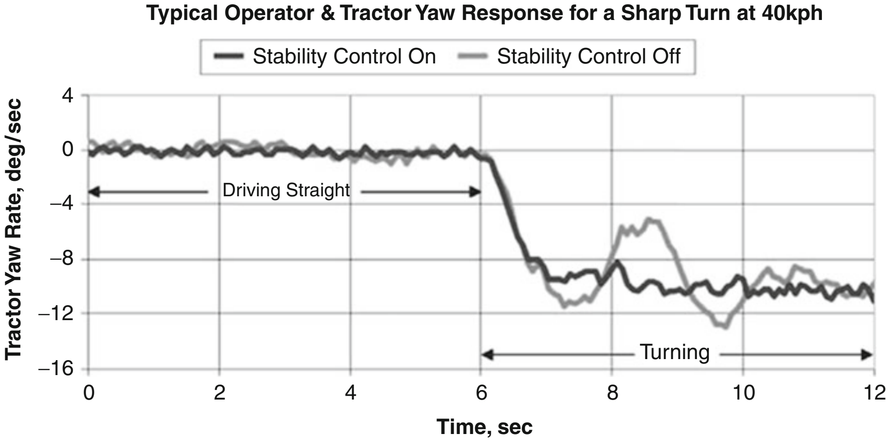


Fig. 4 Tractor yaw response with and without stability control

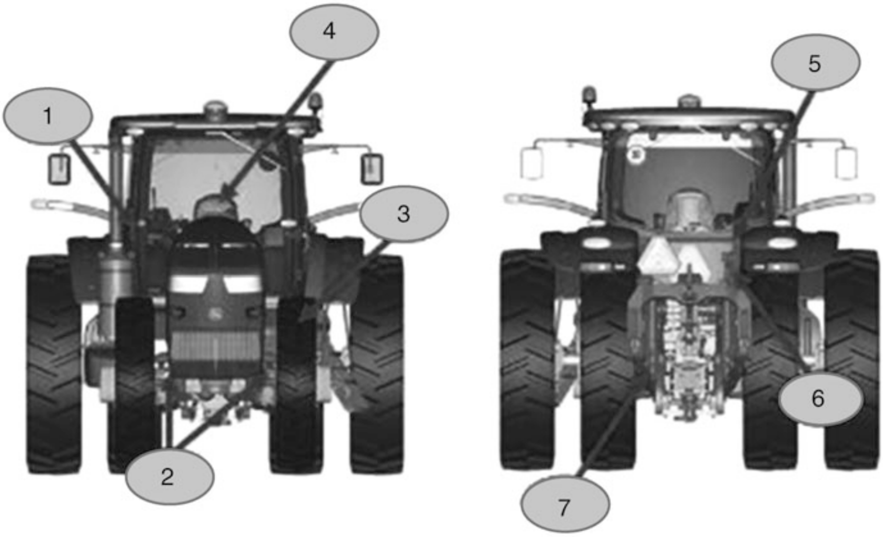


Fig. 5 Overview of components within the ACS system

The use of steering angle sensor units with controlled damping eliminates the typical disadvantages of hydrostatic steering, e.g., steering play and steering creep which are caused by the use of traditional rotary valves. At the same time a variable steering effort can be realized and the driver can get a feedback adjusted to vehicle speed and vehicle operation. Figure 5 shows an overview of the individual components of the steer-by-wire concept.

With the yaw rate sensor (1), the yaw rate of the tractor is sensed and, together with the signals from the steering angle sensor (2) at the right and the left wheel (3) the vehicle speed, is used to control the steering angle as a function of the driver on the steering wheel angle sensor (4) generated steering motion. The ACS-system is designed as a reliable system with two parallel circuits, each with independent control units (5) and hydraulic valves (6). For the hydraulic supply an additional electric booster pump (7) is installed as a fallback in parallel to the main hydraulic supply; through the use of the vehicle battery a fallback for the alternator is provided for the electric power supply (Schick and Kearney 2010).

3 Process Assistance Systems

Both process assistance systems as well as automation solutions are well established in today's agricultural machinery environment. Looking at self-propelled machinery, options like closed loop wheel speed control or integrated automatic guidance systems have proven their value in terms of increased efficiency of the machinery and the overall processes (Balke 2006). Tractors can be equipped with smart drive strategies and headland management systems again to improve overall efficiency. These options have become state of the art in the recent past. Tractors featuring infinitely variable transmissions provide the possibility to automatically adjust wheel speed depending on engine load to either increase process performance or improve process efficiency by running the combustion engine permanently at the most efficient operating speed. Headland management systems focus on the operator by reducing the amount of interactions between machine and operator required when using a tractor-implement-system through automation. A pre-programmed sequence of events can be executed with a push of one button which greatly reduced stress for the operator. Also available in the market of agricultural machinery are implement-internal automation solutions, which do not interact with the pulling tractor. These solutions are tailored to the specific application. This applies for example to the unloading sequence of a silage wagon, where a sequence of events has to be worked through in order to unload the cargo from the trailer (Anonymous 2008). In the following sections, different process-assistance systems for tractors are explained in more detail.

3.1 Tractor-Implement-System Automation

Tractors are usually being used in combination with implements to execute the work. The tractor can be seen as a universal power unit, which can be combined with different implements having different requirements to the power unit. Therefore tractors provide different interfaces which allow the combination with a wide variety of implements. This includes mechanical interfaces to physically connect the implement to the tractor to pull or carry the implement. In addition there are interfaces to transfer power from the tractor to the implement like hydraulic power, mechanical power or electrical power to support functions on the implement side.

The idea of tractor-implement-system automation (also known as TIM/tractor implement management) is to view the combination of a tractor and an implement as one system. This allows optimization of the combination in terms of efficiency and improves the fit of the systems to a specific application. This approach requires infrastructure design changes on the tractor. The enabler is a bi-directional communication interface between tractor and implement using a CAN interface according to the ISO standard ISO11783/ISOBUS (ISO 2012) with an extended functionality. This interface enables the development and implementation of automated processes, which include both tractor and implement. This holistic approach enables significant improvements in terms of productivity compared to the traditional approach which was focusing on either tractor or implement specifically.

3.2 System Architecture

Based on this extended CAN communication interface, the attached implement can access the following functions on the current John Deere tractor models of the 6R/7R78R series:

- Wheel speed control (if tractor is equipped with an infinitely variable transmission)
- Vehicle acceleration/deceleration (if tractor is equipped with an infinitely variable transmission)
- Electronically controlled hydraulic selective control valves (flow rate and timing)
- Power take-off gear selection
- Power take-off disengagement

The communication interface does not allow full and unrestricted access to the vehicle communication bus-system. Figure 6 shows the breakup between the open bus-system for the implement and the proprietary bus-system on the vehicle side. The electronic control unit named “TECU” acts as the gateway between the two systems. While in automation mode, the implement can request actions from the tractor via the TECU; the TECU checks the request and transmits it to the vehicle bus-system.

An important extension of the standardized communication protocol according to ISO11783 is the security layer, which ensures two things: Firstly, the security layer manages the access rights and ensures that only those implements which have been approved by the manufacturers can access the tractor bus-system and tractor functions. Secondly, the security layer also manages the access level such that implements can only access tractor functions, which have been approved by the manufacturers for this specific application. This also included for example the wheel speed range, in which the implement can request speeds from the tractor during operation.

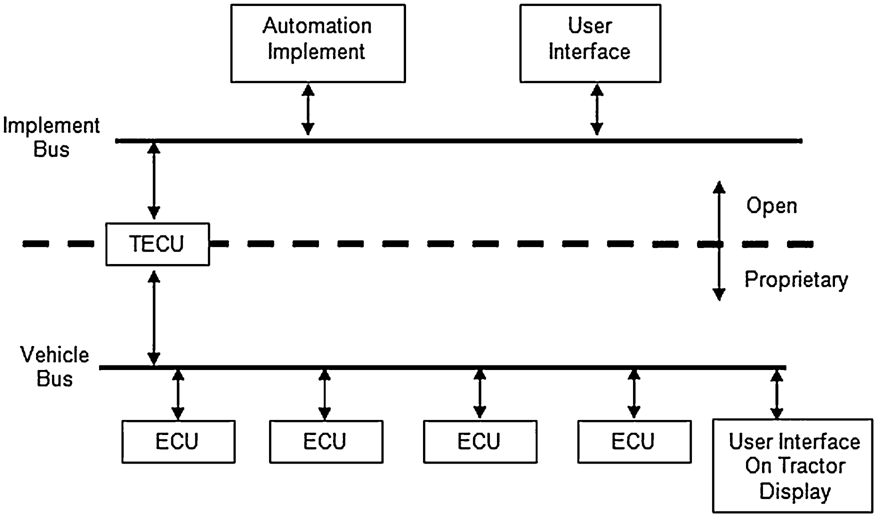


Fig. 6 Bus system with TECU gateway

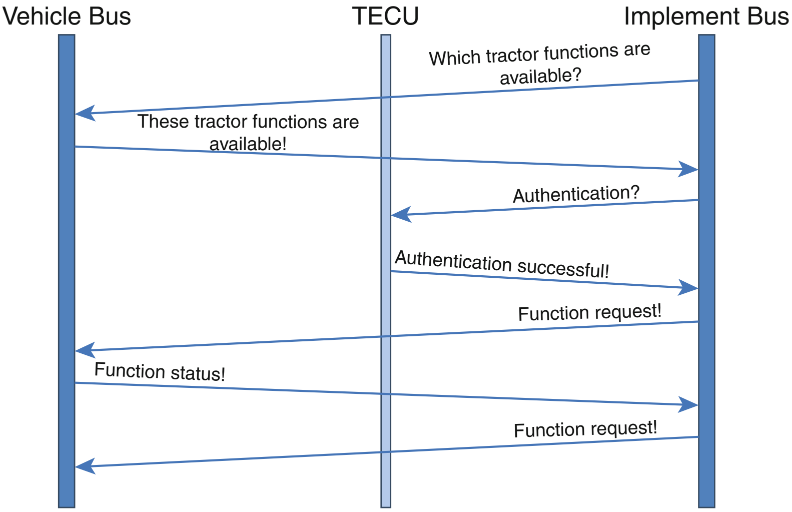


Fig. 7 Initial handshake between tractor and implement

Figure 7 shows the initial handshake between tractor and implement, when the communication is being established for the first time by plugging tractor and implement physically together. In this phase the implement identifies itself to the tractor and the authentication runs.

The implement can send requests for allowed functions to the tractor after a successful authentication. These functions will be executed as long as the operator

Fig. 8 Tractor with roundbaler ejecting a bale

on the tractor site intervenes by manually controlling one of the automated functions. For example the operator reduces manually the wheel speed of the tractor my moving the drive lever while the implement is requesting a specific wheel speed from the tractor. In this case the tractor-implement-automation system is deactivated and the system goes back into manual control mode.

3.3 Tractor-Roundbaler-Automation

In 2009 an automated system composed of a tractor and a roundbaler was introduced in the market as one of the first commercially available solutions including TIM. A roundbaler (Fig. 8) is being used to compress materials like straw or hay into cylindrical bales for transport and storage purposes. The process of baling using this type of machine is a discontinuous process which requires a lot of interaction between operator and machine (Thielicke 2005). Therefore it has a lot of potential for automation.

The following steps have been automated in this system (see also Fig. 9):

- Decelerate the tractor-implement systems down to standstill, as soon as the pre-set bale diameter is reached in the roundbaler

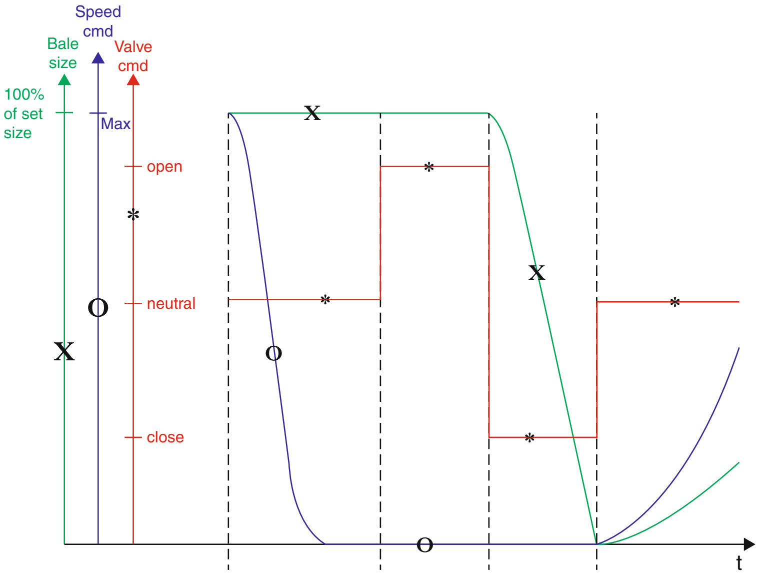


Fig. 9 Automated baling process

- Initiate wrapping process of the bale, as soon as the tractor-implement system is at a standstill
- Open tailgate of the baler by actuating a selective control valve on the tractor, as soon as the wrapping process is done
- Close tailgate of the baler by actuating a selective control valve on the tractor, as soon as the bale has been ejected.

After termination of the automated sequence, the operator has to initiate tractor motion by physically moving a drive lever in order to start the next baling cycle.

4 Automated Steering Systems

Agricultural crop production is always using certain work patterns within a field. Those patterns are achieved by defined tracks and also synchronized planting patterns (Fig. 10). The most enhanced usage of predefined and fixed patterns is utilized for the so called controlled traffic farming. This allows reduction of soil compaction to very limited areas. Hereby the main idea is to have vehicles and implements running only on virtually predefined driving tracks (AGCO Fendt 2011).

Within this agricultural crop production process an increase of efficiency can mainly be reached by changing the following variables:

- Highly developed crop hybrids
- Maximum usage of water and nutrients in the soil

Fig. 10 Potato ridges and tramlines for controlled traffic farming

- Optimized plant positioning inside fields
- Intelligent plant fertilizing and crop protection strategies
- Mature transport logistics and work sequences
- Vehicle systems with more horsepower and larger working width used over more hours/day

The last four variables can be positively influenced by guidance and automation solutions for farm machinery. Usage of those machines can be characterized by many repeated sequences done by operators. Dust, varying solar attitude as well as large working width (up to 60 m/197 ft) and large vehicle length make this work very challenging. This results in faster operator fatigue. Those circumstances are predestined for utilization of automated vehicle systems. A typical work sequence of the operator within one track looks like:

1. Turn vehicle to desired direction
2. Activate multiple work functions for implements in front and rear of vehicle
3. Set desired vehicle speed and engine rpm
4. Continuous adjustment of vehicle heading
5. Continuous monitoring and adjustment of working processes
6. Constant readjustment of tool settings
7. Speed correction for turning maneuver
8. Deactivation of work functions

Steps four to six make up the main action during field work. Thus the main stress relief for the operator can be achieved in those areas. This will also result in a higher work quality over long working periods. As a result of this analysis the following operator assistance systems are defined with their functional requirements:

1. Automatic guidance
   Repeat driving of field tracks with an accuracy of ±5 cm.
2. Obstacle detection
   Potential obstacles in the field like trees, ditches, stones, power poles, shall be detected. A special challenge for that will be that obstacles could be covered underneath the soil or plant canopy.
3. Automatic set point tracing for tools
   Height/depth control of tools as well as automatic adjustment of overloading devices and rate control for fertilizer input increase quality of work.
4. Predefined sequences for process functions
   For a tractor with front- and rear implement functionality needs to be switched when reaching a field boundary and both implements need to be raised/lowered when reaching the work-boundary.
5. Automatic turn maneuvers
   Vehicles and implements need to automatically turn around (most of the time 180°) within minimum radius.

The following sections describe several operator assistance systems addressing those automation needs. The goal of operating agricultural vehicles is an exact placement of tools that are mounted on the implement in most cases. Therefore steering of the implement is required. This can either be done by steering of the towing vehicle or by steering of towing vehicle and implement. At first different options for steering of towing vehicles will be explained followed by actively steered implement. After that different automation solutions for vehicle and fleet automation will be reconsidered.

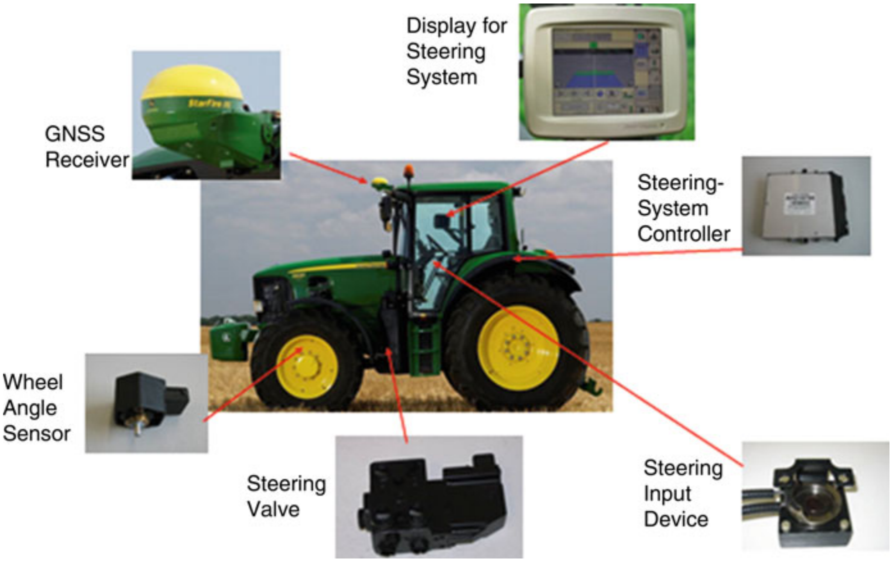


Fig. 11 Components of an automatic steering system John Deere AutoTrac$^{\text{TM}}$

4.1 Steering Assist Systems for Agricultural Vehicles

Those systems are categorized into manual steering assist systems and automated steering assist systems. Manual systems comprise a GNSS (global navigation satellite system) – receiver and a display unit, that indicates to the operator visually and/or acoustically how to manually steer in order to follow a predefined track. Automated steering systems take over steering from the human who uses a display to set up tracks, the system will follow afterwards. Components used for those assist systems are shown in Fig. 11. Therefore a wheel angle sensor determines the current orientation of the steerable wheels. The GNSS receiver determines the position of the vehicle in world coordinate system, the vehicle speed and the direction. The integrated IMU (inertial measurement unit) of the GNSS receiver senses roll, pitch, and yaw angle. To achieve a steering accuracy of ±5 cm or even better and absolute repeatability of tracks agricultural applications normally utilize RTK- (real time kinematic) GNSS systems. Therefore correction signals are transferred to the vehicle either by radio or cell-modem. The steering system controller contains the inner loop of the control system. The inner loop controls the steering valve to get the wheel into the desired position (Fig. 12). Furthermore this controller is monitoring the steering input device detecting manual interaction of the operator with the steering wheel that stops the automatic steering. The display unit contains the track definition, the outer control loop and the user interface.

Besides the integrated steering systems there are also universal steering systems available as retrofit solutions. In this case an electric motor is either frictionally or force-fit engaged to the steering column. In this case there is no direct interaction

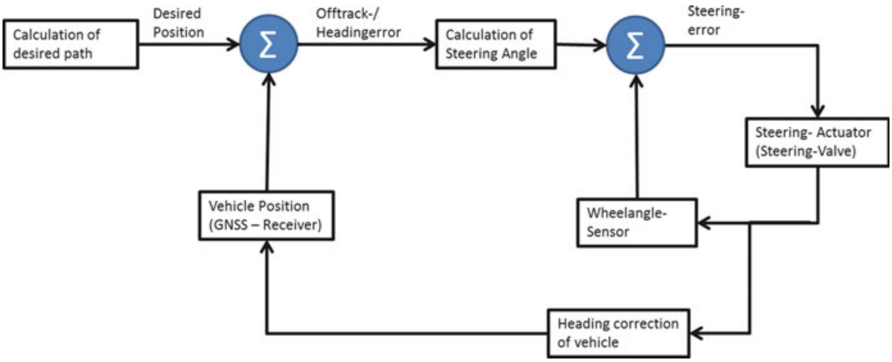


Fig. 12 Control loop of an automatic steering system

with steering hydraulics. Examples of those retrofit solutions are John Deere AutoTrac Universal and Trimble EZ-Steer (Trimble Navigation Limited 2014).

Automatic steering systems allow users to drive the following contours:

- Parallel straight lines that are defined by two points or one point and heading,
- Parallel curves defined by manual driving of the first track and storing of the contours,
- Circle tracks set up through centerpoint and radius or multiple points on a circle,
- Designed tracks defined by path planning software or previous recording.

As soon as an automatic steering system gets activated, the control system trys to get the vehicle as soon as possible to the closest defined track. Therefore the lateral offset as well as the heading error are continuously calculated. Both values are entered into a PI-control system as shown in Fig. 12. The calculation of a kinematic vehicle model can be done based on a bicycle model of a tractor (Fig. 14).

4.2 Steering Assist Systems for Implements

As stated at the beginning the exact positioning of the tools is the main focus for automatic guidance systems. Therefore controlling the vehicle is only sufficient if the tools are rigidly mounted to the vehicle and the vehicle is following the desired track without a tracking angle error. This rigid connection between vehicle and tools is currently not possible for large tractor-implement combinations. Pulled implements usually have at least one hinge between tractor and tools. Thus additional steering concepts are needed to get the tool to the desired position with the desired orientation. The following chapters describe two principle attempts to make this happen.

4.2.1 Passive Implement Steering

Passive implement steering solutions alter the path of the tractor to the extent that the implement finally follows its desired track. This means the tractor has to

Fig. 13 Tractor on side hill slope with passive implement guidance John Deere iGuide^{TM}

significantly deviate from the implement's track in curves and on side hill slopes (Bowman 2009). This steering method can be used on an open-loop approach calculating the implement position only based on geometry data. The accuracy of this solution is limited to the ability of the implement to follow a theoretically calculated path. For a closed loop approach a second GNSS-receiver is mounted to the implement. This allows determining the accurate position of vehicle and implement and adjusting controls therefore (Fig. 13). In this figure the white lines describe the targeted trajectory of the implement and the yellow line the trajectory a tractor needs to follow in order to keep the implement on a desired track. Here it is very obvious that a steering maneuver of the tractor causes a change of the implement position with quite some delay.

Furthermore this kind of compensation is only possible, if there is no standing crop in the field. Otherwise the tractor would need to roll over crop in order to get the implement onto the desired track. Thus this method can only be used for broad acre farming. As soon as there are potato ridges, tramlines or controlled traffic tractor and implement must follow exactly the same track. In addition to that Fig. 13 shows clearly that the orientation of tools is not the same as the working direction. The next section will describe a solution that addresses those challenges.

4.2.2 Active Implement Steering

High value crops like bedded crops and vegetables require a steering solution that allows avoidance of damage to the crops and the ground based irrigation systems

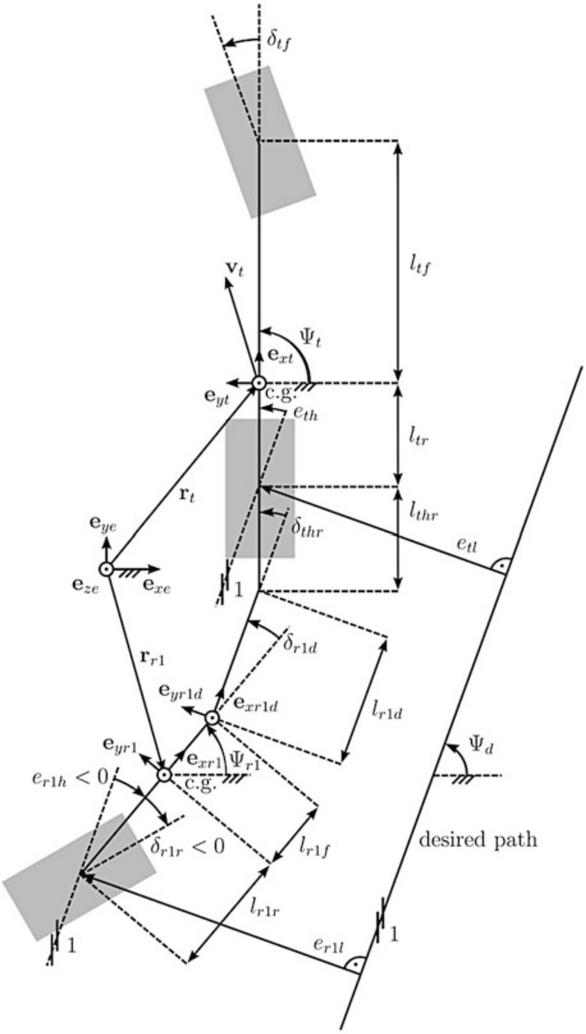


Fig. 14 Tractor and steerable implement bicycle model showing earth-fixed reference frame (e_{xt}, e_{yt}, e_{zt}), tractor-fixed reference frame (e_{xt}, e_{yt}, e_{zt}) at tractor's center of gravity (c.g.), implement fixed reference frame (e_{xr1}, e_{yr1}, e_{zr1}) at implement's c.g., tractor front wheel steering angle δ_{tf}, implement wheel steering angle δ_{r1r}, drawbar steering angle δ_{r1d}, hitch angle δ_{thr}, orientation of the desired path Ψ_d, tractor lateral error e_{t1}, tractor heading error e_{th}, implement lateral error e_{r1l}, implement heading error e_{r1h}, and all geometric dimensions required

caused by tools or tractor wheels. Therefore highly accurate repeatability of tractor and implement tracks is required. An active implement steering solution allows this. In this setup tractor and implement each have a GNSS receiver and both units are independently steerable. The communication between tractor and implement uses a CAN protocol according to ISO 11783 (ISO 2012). In order to not only control the implement position but also the orientation of tools multiple steering actuators are needed on the implement. Werner et al. (2012) describes a potential combination of steerable wheels combined with a steered drawbar. The bicycle model of a tractor and a steerable implement are shown in Fig. 14. A control system for two degrees of freedom allows in this case manipulation of tools in lateral direction as well as in orientation/heading.

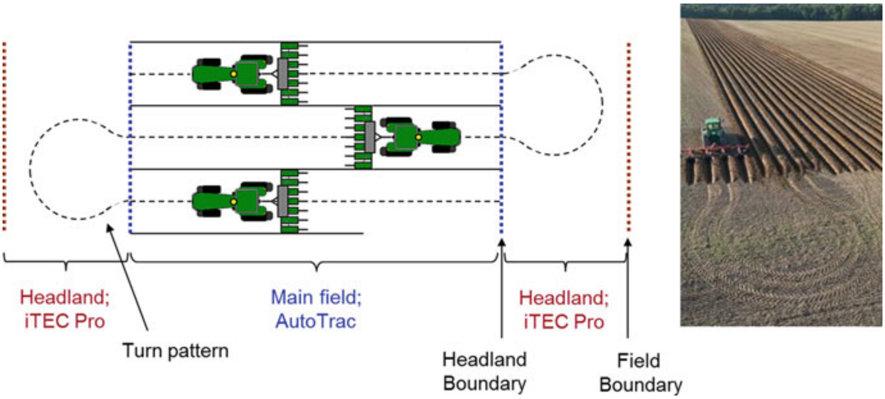


Fig. 15 Automatic headland management John Deere iTec ProTM

A special challenge for tuning of implement steering control systems is caused by the significant change of weight with that of center of gravity during operation. Pulled sprayers, seed drills, and potato harvesters are often more than twice as heavy when filled compared to when empty. Furthermore there are mainly hydraulic steering actuators used for implements. Thus the responsiveness of the system also depends on length/shape of hoses and fittings, oil pressure and temperature. All those factors require continuous adjustment for tuning parameters of implement steering systems.

Today there are active implement steering systems available on the market like John Deere iSteer$^{\mathrm{TM}}$ and Trimble True Tracker (Trimble Navigation Limited 2012).

4.3 Automatic Headland Turns and Automatic Tool Adjustments

The previous sections described solutions to accurately steer vehicles and implements on pre-defined tracks. Now the next logical step toward automation of in-field work process is the integration of headland turns and automatic adjustment of tool-positions. Today all functions shown in Fig. 15 are available within the headland management system John Deere iTec Pro$^{\mathrm{TM}}$.

Therefore at first a field boundary is required. This can be created through boundary recording while driving once around the field or it can be obtained from the land registry office. Based on this information the so called headland is defined. This is the area where vehicles and implements can turn around. This internal headland boundary is created as a parallel line from the field boundary using multiple instances of the working width of the implement. The next step defines the function and timing therefore to be executed at the headland boundary: e.g., the sequence for a drill is activate PTO (power take-off), lower implement, and open seedgate. In doing so the drop point of the seeds defines the position of engagement.

Fig. 16 AGCO Fendt GuideConnect (AGCO Fendt 2011)

In case there is a front-mounted implement too this one needs to be lowered much earlier. Once the vehicle system reaches the other headland boundary all functions need to be deactivated in reverse order. After that a pre-defined turn pattern is executed. Once those sequences are defined for a specific field, work can be completed without additional manual interaction of the operator within the selected boundary.

This system is one first step toward highly automated vehicle fleets or even autonomous vehicles at some point.

4.4 Collaborating Vehicles

On large farms it is very common, that multiple vehicles are doing the same work in the same field to increase ground coverage during operation. This work scenario was addressed by Karlsruhe Institute of Technology (KIT) and AGCO Fendt within a research project developing an electronic drawbar (Zhang et al. 2010). Therefore identical vehicles with the same implements are used. This constellation enables one operator to run multiple vehicles. An unmanned electronically guided vehicle gets connected to a manned vehicle. The leading vehicle transfers its position/track to the following vehicles. Those calculate their track based on implement width and inline offset. Additional information, like engine load and process parameters, is used to navigate the following vehicle dynamically. This solution was shown by AGCO Fendt called GuideConnect System (Fig. 16)

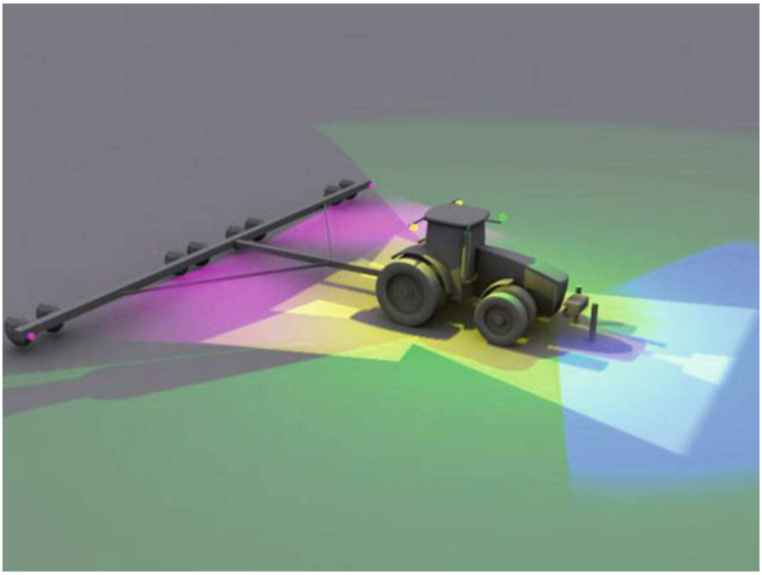


Fig. 17 John Deere concept of perception sensors for an unmanned tractor with implement

(AGCO Fendt 2011). Thereby the operator of the lead vehicle takes responsibility of both vehicles that communicate with each other through an encrypted wireless protocol. In case of errors or loss of communication an emergency stop gets activated on the follower. This means the vehicle is stopped and the engine turns off.

Combinations of tractors and implements as well as autonomous vehicles need to follow high safety standards during operation. The more autonomy that will be given to those vehicles the more important a complete monitoring of the environment will be. Especially during engine startup and motion initiation it is irremissible to ensure that there are no obstacles and people around vehicles. Figure 17 shows an overlay of potential fields of view for a tractor with implement equipped with multiple sensors. It becomes clear that the area in front of the vehicle but also the area between tractor and implement need to be covered. Coverage by LIDAR-sensors is displayed in blue and purple while green and yellow areas are covered by monocular or stereo-vision cameras. Sensor systems like RADAR, LIDAR, and stereo-vision cameras are described in ▶ Chaps. 17, “Automotive RADAR,” ▶ 18, “Automotive LIDAR”, and ▶ 21, “Stereovision for ADAS”.

For wide implements – up to 60 m (197 ft) – it will be challenging to reliably detect the corners of the implement. Figure 17 shows those corner points as purple dots. Sooner or later similar solutions will be used as in automotive industry comprising a combination of short range and long range RADAR sensors.

5 Outlook to Highly Automated Vehicles in Agriculture

For more than 10 years one trend in agricultural engineering is autonomous/highly automated vehicles. John Deere has shown autonomous tractors working in orchards in Florida as described by Moorehead et al. (2010). Those vehicles are used for mowing and spraying operations (Fig. 18): In this case laser scanners and cameras are used to monitor the environment of the vehicle. All vehicles are operated from a central mission control center where the operations manager is able to monitor the environment of the vehicles if needed.

Another solution in that area is offered by KINZE Manufacturing Inc. (Fig. 19): In this setup a tractor is retrofit with sensors enabling the overall system to autonomously run a tractor with grain cart between a combine in the field and a truck at the field boundary (McMahon 2012).

All development in this area is supported by various standards: ISO/CD 18497 contains safety requirements applicable to highly automated vehicles (ISO 2013). This standard contains guidelines for system components, communication protocols, perception systems including test and operation procedures. For different countries those vehicles also need to be homologated by employers' mutual insurance association and technical inspection authorities.

Summarizing the last sections it can be stated that agricultural engineering is stepwise progressing toward highly automated or autonomous vehicles. Further development of the systems described in this chapter will be an important milestone in this direction. Within the next years fully automated systems will first be used in fenced-in areas before releasing them to broader areas. The main challenge

Fig. 18 Autonomous orchard tractor (Moorehead et al. 2010)

Fig. 19 Kinze autonomous grain cart (McMahon 2012)

therefore will be not only to automate the driving but also to monitor the process parameters and ensure high quality execution. It is key that seed-placement, harvesting, overloading etc. are executed reliably.

References

AGCO Fendt (2007) Pressemitteilungen: Die erfolgreichen Vario-Baureihen. Verfügbar unter: http://www.fendt.at/pressebereich_pressemitteilungen_1033.asp. Accessed 24 Apr 2014

AGCO Fendt (2011) Pressemitteilungen: Fendt GuideConnect. Verfügbar unter: http://www.fendt.com/de/pressebereich_pressemitteilungen_7099.asp. Accessed 9 Mar 2014

Anonymous (2008) Produktinformation Pöttinger Ladewagen Jumbo/Torro, Grieskirchen, Österreich Verfügbar unter: http://www.poettinger.at/de_de/Produkte/Kategorie/14/silierwagen-mit-laderotor. Accessed 13 Aug 2015

Balke S (2006) Kostensenkung durch Automatisierungssysteme im Mähdrusch, Diplomarbeit FH Weihenstephan

Bauer W (2008) Hydropneumatische Federungssysteme. Springer, Berlin

Bowman K (2009) Economic and environmental analysis of converting to controlled traffic farming. In: 7th Australian controlled traffic conference, Canberra, pp 61–68

Braess HH, Seifert U (2000) Vieweg Handbuch Kraftfahrzeugtechnik. Vieweg, Braunschweig/Wiesbaden

Dudzinski P (2005) Lenksysteme für Nutzfahrzeuge. Springer, Berlin

Hesse H (2008) Elektronisch-hydraulische Systeme. Expert-Verlag, Renningen

ISO (International Organization for Standardization). ISO 11783 (2012) Tractors and machinery for agriculture and forestry – serial control and communications data network. Geneva

ISO (International Organization for Standardization). CD ISO 18497 (2013) Agricultural machinery and tractors – safety of highly automated machines. Frankfurt

McMahon K (2012) Kinze's autonomous tractor system tested in field by farmers. Verfügbar unter: http://farmindustrynews.com/precision-guidance/kinze-s-autonomous-tractor-system-tested-field-farmers. Accessed 12 Mar 2014

Moitzi G (1997) Vermeidung von Bodenverdichtung beim Einsatz von schweren Landmaschinen; Ländlicher Raum; Online-Fachzeitung des Bundesministeriums für Land- und Forstwirtschaft, Umwelt und Wasserwirtschaft; Jahrgang 2007; Wien Verfügbar unter: www.bmlfuw.gv.at/dms/.../Moitzi/Moitzi_Boxberger_pdf_END.pdf. Accessed 13 Aug 2015

Moorehead S, Stephens S, Kise M, Reid J (2010) Autonomous tractors for citrus grove operations. In: 2nd international conference on machine control and guidance, Bonn, pp 309–313

Schick T, Kearney J (2010) "Steer-by-Wire" for large row crop tractors. In: Tagung Landtechnik 2010, VDI-Max-Eyth-Gesellschaft, VDI-Berichte Nr. 2111. VDI Verlag, Düsseldorf

Thielicke R (2005) Automatisierung und Optimierung traktorgebundener landwirtschaftlicher Arbeiten mit zapfwellengetriebenen Geräten an ausgewählten Beispielen. Dissertation, Halle/Saale

Trimble Navigation Limited (2012) Datasheet true tracker system. Verfügbar unter: http://trl.trimble.com/docushare/dsweb/Get/Document-343005/022503-282A_TrueTracker_FS_0707_lr.pdf. Accessed 9 Mar 2014

Trimble Navigation Limited (2014) Datasheet EZ-Steer system. Verfügbar unter: http://trl.trimble.com/docushare/dsweb/Get/Document-468909/. Accessed 9 Mar 2014

Werner R, Kormann G, Mueller S (2012) Dynamic modeling and path tracking control for a farm tractor towing an implement with steerable wheels and steerable drawbar. In: 2nd commercial vehicle technology symposium, Kaiserslautern, pp 241–250

Wiedermann A (2012) Auslegung von Lenksystemen in modernen Traktoren, Tagung Landtechnik 2012, VDI-Max-Eyth-Gesellschaft, VDI-Berichte Nr. 2173. VDI Verlag, Düsseldorf

Zhang X, Geimer M, Noack P, Ehrl M (2010) Elektronische Deichsel für landwirtschaftliche Arbeitsmaschine. 68. Internationale Tagung Landtechnik, Braunschweig, pp 407–412

Navigation and Transport Telematics

54

Thomas Kleine-Besten, Ulrich Kersken, Werner Pöchmüller, Heiner Schepers, Thorsten Mlasko, Ralph Behrens, and Andreas Engelsberg

Contents

T. Kleine-Besten (✉) • W. Pöchmüller • H. Schepers • R. Behrens
Robert Bosch Car Multimedia GmbH, Hildesheim, Germany
e-mail: thomas.kleine-besten@de.bosch.com; werner.poechmueller@de.bosch.com; heiner.schepers@de.bosch.com; ralph.behrens@de.bosch.com

U. Kersken
Robert Bosch GmbH, Hildesheim, Germany
e-mail: ulrich.kersken@de.bosch.com

T. Mlasko • A. Engelsberg
Bosch SoftTec GmbH, Hildesheim, Germany
e-mail: torsten.mlasko@bosch-softtec.com; andreas.engelsberg@de.bosch.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_55

Abstract

This chapter provides an overview of vehicle navigation and traffic telematics. A brief history of navigation is followed by an introduction to the various technologies involved. Several aspects are considered, like position fixing and route calculation algorithms. On-board, off-board and hybrid navigation systems are compared with one another, and the interaction between navigation functions and traffic telematics is explained. The authors explore in particular the potential of these technologies in the development of future advanced driver assistance functions. In this context, they provide an overview of the radio broadcasting and mobile communications technologies available for supplying the vehicle with telematics data.

1 History

The development of modern radio/navigation head units and telematics devices began with the introduction of car radios in the early 1930s. The first car radios were based on vacuum tube technology and had a space requirement of more than 10 l. The invention of semiconductors and the associated component miniaturization made it possible to design more compact radio/navigation and telematics units, thus paving the way for their widespread adoption in automobiles (see Sect. 9.2).

Increasing motorization in the 1960s drove advancements in navigation and telematics technologies for automobiles. The number of motor vehicles on the road in West Germany had increased since the 1950s by more than ten times to 21.3 million in 1976. Annual mileage in 1974 totaled around 168 billion miles (270 billion kilometers). Freight traffic had increasingly shifted from rail to road, so that in 1975, 43 % of goods transportation was handled by road. This led to traffic congestion problems, particularly on the autobahns (freeways). Even the continued development of the road network could no longer keep pace with the increase in demand for more space for road traffic. Furthermore, influenced by the first energy crisis and other factors, there was an increasing general awareness for the environmental impact of raw material and energy consumption. Connected with the increasing road traffic, the number of traffic accidents and thus the number of injuries and deaths rose significantly.

Within the scope of research projects, scientists developed the first traffic control systems on autobahns for collecting traffic data and for influencing traffic flows by means of speed limits, restrictions on passing, and signage indicating alternative routes. These included, for instance, a warning system at Aichelberg, a traffic regulating system on the A3 autobahn near the Dernbach and Heumar interchanges, and the Rhine/Main alternative route control system. In the early 1960s, radio stations began broadcasting the first traffic reports. Initially these were weekly reports and forecasts and then later daily news about the traffic situation. With the introduction on June 1, 1974 of the Driver Radio Information System (Autofahrer-Rundfunk-Informationssystem – ARI) for broadcasting traffic reports on the radio, a first step was taken in traffic news automation. In connection with this system, sector identifiers for traffic news were defined within the states of the Federal Republic of Germany. Such sector identifiers were displayed on signs on the autobahns and could be selected using a switch on a suitably equipped radio unit. Traffic reports were then broadcast every half hour, and the radio program could be interrupted for urgent news, such as warnings about wrong-way drivers. The further development of this technology to handle a large number of traffic reports and long traffic announcements led to the RDS-TMC telematics service (see Sect. 7).

The first ideas about electronic route guidance systems for automobiles were published in the USA in 1968 by G. Salas (1968). In 1969, this was taken up by Dr. W. Kumm at the Institute of Communication Systems and Data Processing at Aachen Technical University, and induction was proposed as the transmission path for route guidance data. It was suggested that the induction loops that were already in common use for traffic data acquisition should be used for data exchange and that the data acquisition and route guidance functions should be combined in a single system. Thus the basic idea for the Motorist Guidance and Information System (Autofahrer-Leit- und Informationssystem – ALI) was born.

ALI was a personalized, infrastructure-supported route guidance system for car drivers on German freeways and long-distance highways and served two purposes: traffic data acquisition and the transmission of personalized driving recommendations. If there was a traffic disruption, it would guide the driver to the destination by a less busy route. It would therefore have the effect of cutting automobile operating costs, reducing the costs due to time spent driving, and lowering the accident risk. An ALI field trial was carried out in the eastern Ruhr in 1980/1981. The cost-benefit analysis when looking at the economy as a whole, however, came to an unfavorable conclusion: it was estimated that the public sector could expect annual costs of 8.3 million German marks. As a result, attention turned to pursuing an idea from 1978 within the scope of the EVA project (Elektronischer Verkehrslotse für Autofahrer – Electronic Navigator for Motorists): namely, an on-board route guidance system that offered navigation through traffic. Essential elements of the system were installed in the vehicle, thus resulting in a favorable cost-benefit ratio, which was verified in a field trial in 1983. The further development of this system lead in 1989 to the first European series-production navigation device in motor vehicles.

2 In-vehicle Navigation

The primary function of a navigation system is to guide the user to a geographic destination. The system's input variables are provided by the position-fixing sensors and the digitized road data that is held on data storage devices. The road data is a digital representation of the real-world road network. After receiving the relevant user input, the system uses the above initial data to provide the driver with visual and acoustic information for driving the vehicle to the destination (see Fig. 1).

The navigation unit's processor assembly comprises a central processing unit, connected memory, and graphics hardware. The core navigation functionality is realized by means of software modules that are executed on the processor assembly (see Fig. 2). The navigation system uses voice output to inform the driver about the route to take and provides additional visual information on its displays (map display and/or symbol display) or on the instrument cluster (usually a symbol display).

The following software modules exist on the processor assembly:

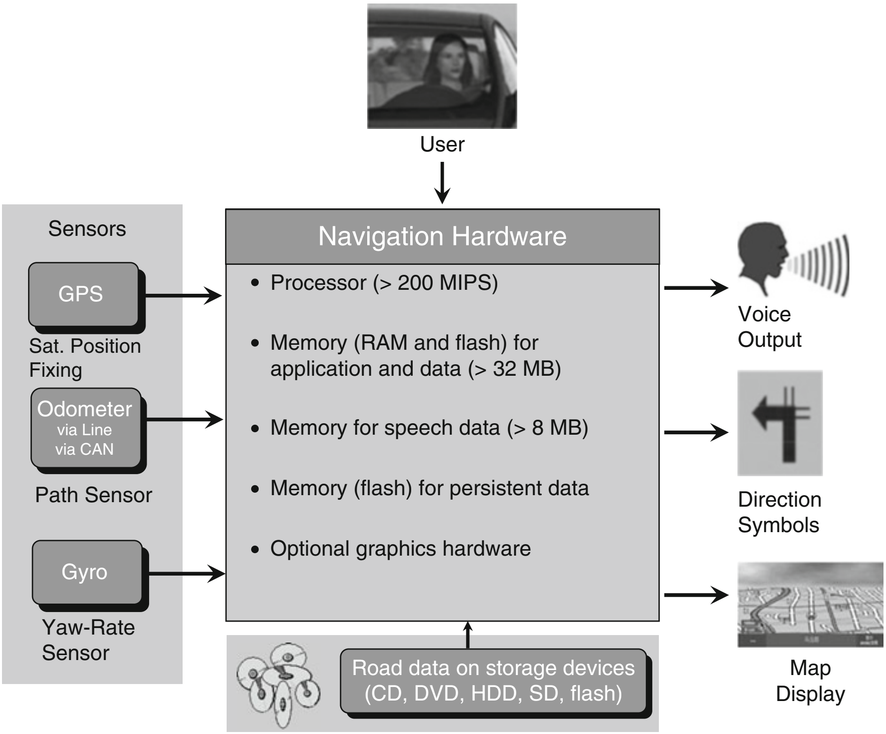


Fig. 1 The navigation environment

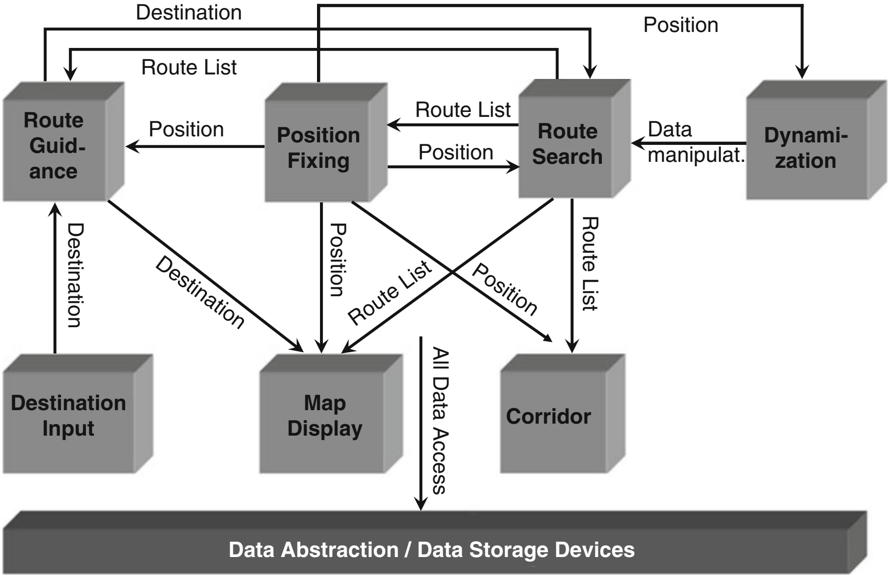


Fig. 2 Software modules of the navigation system

- **Position-fixing module** for determining the position using the data provided by the sensor systems
- **Destination input module** to enable the user to specify the destination
- **Route calculation module** for computing the route from the current position to the entered destination
- **Route guidance module** to guide the driver along the route by means of visual and acoustic information
- **Map display module** for displaying the geographic map with the current location, route and additional information
- **Dynamic navigation module** to incorporate and take into account environmental events (such as fog and ice on the roads) as well as traffic information (like traffic congestion and blocked roads) in the current route
- **Corridor module** for the advance storing of data taken from the navigation data storage device and placed in the main memory of the processor assembly. The corridor is used in particular with CD/DVD-based navigation systems to provide navigation functionality without the data disc having to remain inserted in the disc drive, thus freeing up the drive for playing audio CDs.

2.1 Position Fixing

The task of the position-fixing module involves using the currently available sensor data and its history to accurately determine the current position. In doing so, a distinction must be made between two types of position specification:

(a) The absolute position of the vehicle in space, e.g., specified using WGS 84 coordinates plus a motion vector
(b) The relative position of the vehicle in relation to the road network represented by the digital map (position determined according to so-called "map matching")

The vast majority of functions currently use the position in relation to the road network. The demands placed on the position-fixing module are manifold. New functions, particularly the utilization of navigation systems for driver assistance functions, increase the demands (e.g., position fixing precision, integrity information, error estimation, roadway recognition, and ascertainment of the vehicle position within the boundaries of the roadway).

Implementing a navigation system that will guide a driver to a chosen destination results in the following cross-functional requirements:

The absolute position ascertained by the position-fixing module must be assigned to a position on the map. The position-fixing module must decide whether the vehicle is located on the road (on-road) or off the road (off-road). The position-fixing algorithm ascertains a precise absolute position and at the same time is tolerant of inaccuracies in the digitized map. For the purpose of route guidance, the position error on the road network must at all times be so small that the system can output driving recommendations in good time before every intersection, even in the case of driving maneuvers that need to be performed in quick succession. None of the driving or U-turn maneuvers must cause a loss in position fix. The system must also be able to recognize when the vehicle leaves the road (driving from on-road to off-road), such as when entering a parking lot. When exiting the parking lot at any point and driving onto the nearest road (driving from off-road to on-road), the position-fixing module must automatically switch to the correct position on the road network. After a journey of any distance outside the road network, the system must find the correct position on the network immediately after the vehicle re-enters the digitized area. Particularly with regard to the Asian and North American road networks, it is important that the system be capable of recognizing a change in road altitude in multi-level bridge and roadway systems. In future, users will expect position-fixing capabilities inside buildings, for instance, in parking garages and for pedestrian navigation.

If the navigation unit was moved while switched off, it must be able to quickly and reliably recognize this when switched on again and then determine its current position on the map. In the case of mobile navigation systems (e.g., a personal navigation device – PND) and mobile phones with position-fixing capabilities, this is a normal occurrence; in the case of systems integrated in vehicles, this situation would occur when the ignition is off and the vehicle is moved on a ferry or auto train.

A requirement that improves precision and convenience is the system's automatic calibration. It takes into account tire wear and in particular recognizes when the tires have been changed, thus invoking recalibration. Position fixing is a two-stage process: a position is first determined using dead reckoning and is

Table 1 Navigation system sensor technology

Sensor	Normal installation type	Obtained signal	Error source
Gyro (gyrocompass)	Direct installation of micromechanical gyros or transmission of gyro data from an existing gyro (e.g., ABS gyro) via the CAN	Angle difference	Noise Temperature change Installation angle
Odometer connected by wire	Square-wave pulse	Path difference	Tire expansion
Odometer via CAN		Path difference	Tire expansion
GPS receiver	GPS receiver installed directly	Absolute position Path difference Angle difference	Reception interference like gaps in reception and multipath reception
Accelerometer (multiaxial, if necessary)	Direct installation	Depending on the number of axes: the change in acceleration	Temperature change Installation angle
Wheel pulses	Via CAN from the ABS sensors	Path and angle difference	Tire expansion
Steering-angle transducer	Via CAN	Angle difference	Tire expansion

subsequently depicted on the digitized road geometry by means of map matching. Dead reckoning is a technique that takes a recent position as a starting point and uses the path difference, change in angle, and amount of time elapsed to ascertain a new absolute position. Taking a known initial position and a known initial angle as the starting point, it is possible to use distance and angle measurements and the addition of the path vectors to ascertain the new position that has been reached. The position-fixing module records and synchronizes the signals that are supplied with varying frequency (see Table 1) and attributes them to a common time base in order to then ascertain the dead reckoning position.

In mobile navigation systems (PND), the current position is normally determined using the Global Positioning System (GPS). GPS utilizes satellite positioning to obtain a position fix. The satellites orbit the Earth twice a day at an altitude of around 20,000 km. The United States Department of Defense has placed up to 31 satellites in Earth orbit so as to offset interference and outages. For reasons of cost, however, a reduction in the number of satellites to 25 is being pursued. The satellites broadcast radio signals on 1,575.42 and 1,227.60 MHz while at the same time supplying information about their position and the time of day. If at least four satellites are receivable, it is possible for a receiver to determine its current position based on the position information from the satellites and the transit time of the signals obtained from the clock time difference between the transmitter and receiver. The procedure is equivalent to an equation involving four unknowns (time and three positions) with a point of intersection of three spheres whose radii are determined by the signal transit time. In 2000 the imposed artificial inaccuracy of the signals was deactivated to improve civilian use of GPS, thus enabling position fixing by means of GPS with an accuracy of 10–20 m in situations in which reception is good and the satellite constellation is favorable.

Automobile systems are generally equipped with additional sensor technologies or are connected to sensor systems already fitted to the vehicle, thus enabling them to provide a reliable position even without GPS reception. Since all signals can be inaccurate in certain situations, the position-fixing module must compare the signals against each other and calibrate them. An established method for doing this is Kalman filtering (Kalman 1960) in which the output value is first estimated on the basis of a system model and is then compared with the value measured by the sensor systems. The difference between the estimate and the measurement is then used to improve the current system state. As a result, it is possible to incorporate inaccurate data with a correspondingly lower weighting in the dead reckoning process (see ▶ Chap. 25, “Data Fusion for Precise Localization”).

Once the position has been determined by means of dead reckoning, a comparison with the digital map data on the storage device must be performed, that is to say, “map matching”. The matching process is necessary since not only does the dead-reckoning position contain inaccuracies but the digital road data does too. This is firstly due to imprecision and errors in data acquisition when the road network is digitized; and, secondly, it is due to information having been removed from the digital data during data preparation – a process known as “generalization,” which is used to reduce the volume of data so that it fits onto the data storage device. The employed procedure generally involves using the sensor data to establish a trajectory and then matching this trajectory to the map data. This results in the most probable position on the map being determined. Despite the requirement to ascertain the position as accurately as possible, a margin of error is important here too so that if there is a slight deviation from the digitized road network this will not lead to erroneous behavior by the navigation system. Thus, for instance, if there is a construction site on a freeway and the traffic is diverted onto the oncoming lane, this will not lead to a situation in which the map-matching function matches the vehicle position to the oncoming lane and the navigation system then recommends the driver to make a U-turn. For this purpose, the map-matching function analyzes relevant attributes, like driving directions, that are stored in the map data.

Experience shows that single-path map matching lacks reliability despite the use of sophisticated algorithms. This is because, when only one path is considered, achieving the required high map-matching quality is not possible due to sensor tolerances, minor digitization inaccuracies, and environmental effects. Matters can be improved if several paths are regarded simultaneously (multipath map matching). By tracking several paths, the position-fixing module is provided with the ability to reverse a wrong decision about the “most probable road segment” – the system is thus endowed with a memory, so to speak. By rating the paths that are being tracked in parallel, a principal path will emerge as the path with the highest rating. The position on the principal path is used to control the navigation system’s driving recommendations and to display the vehicle position on the map. If the ratings for the principal path and for one of the parallel paths are approximately equal, the system will then consult the calculated route as an additional criterion in order to finally identify the principal path. In this context, the parallel paths act as a kind of backup feature that the system can fall back on in case a plausibility assessment yields for one of the parallel paths a greater probability for the vehicle position than the hitherto principal path. Alongside the process of parallel path generation, a process for reducing the number of parallel paths is also necessary – since without such a reduction, a highly dense road network would very soon result in an unmanageable quantity of paths having to be considered. A parallel path is deleted when its rating falls below a set threshold value (Neukirchner 1991; Pilsak 1999). The position of the principal path computed by the map-matching function and displayed on the road is made available to the other navigation modules in a suitable form. In order to provide a precise and smooth map display, a high-frequency position signal (>15 Hz) is required, which means that the position-fixing module must, if necessary, extrapolate or interpolate the position.

2.2 Destination Input

The destination input module – also called the "index" – provides the user with various options for entering destinations. The road names and data that describe locations are generally stored in a tree structure in a highly compressed form on a storage device (see Fig. 3).

The hierarchical address input in Europe based on political boundaries is generally performed by entering details in the following sequence: country, place or zip code, street, house number. In North America, one enters: federal state, street, place. The data presented to the user is thinned out to such an extent that, for instance, in Europe only the roads that exist in the selected place or zip code area are shown, and in North America only the places in which the already entered road exists are shown. Further input assistance functions are: the automatic spelling function (ASF), which grays out letters in the user menu for letter combinations that are no longer possible; and the similarity search function that presents the user with similarly written places/roads for selection. Furthermore, in the case of ambiguity (e.g., there are several occurrences of the place name "Frankfurt" in Germany), the user must be offered a selection to resolve the ambiguity.

Besides the method for directly inputting an address, it is often also possible to select an address from a list of particularly interesting places – such "points of interest" (POI) include filling stations, automobile repair shops, tourist attractions, and restaurants, etc. For these kinds of destinations, it is often possible for the user to access a travel guide that provides – alongside the location details – additional information about the destination, like the type of restaurant and opening times.

Particularly with regard to POIs, it is important that there is a function that allows users to perform a vicinity search so that, for instance, they can search for a particular parking lot close to their destination, or for a filling station or restaurant at their present position or on their current route.

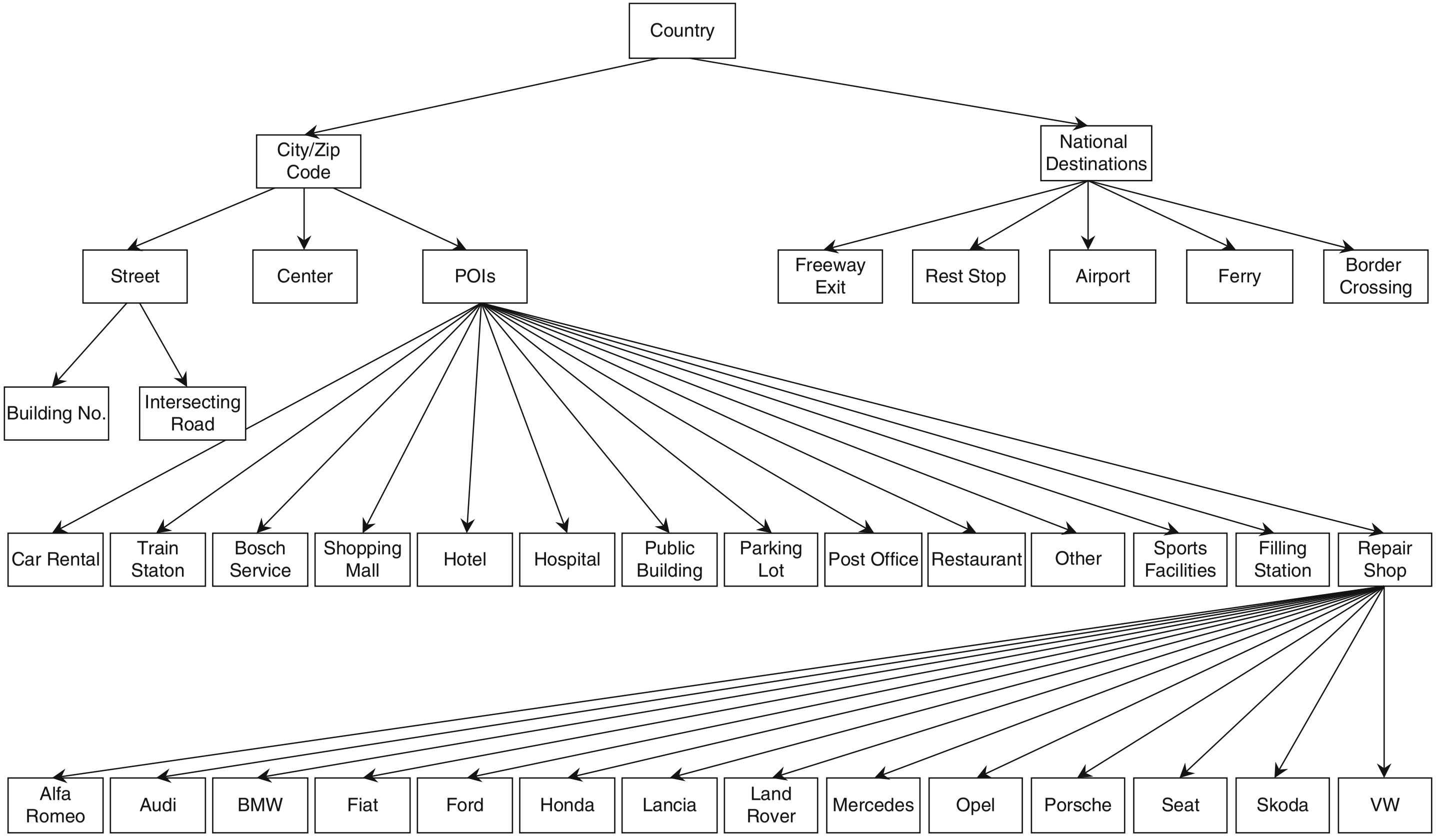


Fig. 3 Destination input tree structure

The response time of the user interface during destination input must be kept short so the user is able to enter the destination quickly. Good performance can be achieved when accessing data stored on fast storage media (hard disk drives, SD cards, flash memory); however, in the case of slower media (CDs and DVDs) that have longer data access times (seek times), an access-optimized algorithm must be used. Despite the provision of all the above-mentioned convenience functions (like the thinning out of data), this algorithm must be able to make do with a minimal number of access operations.

2.3 Route Search

The route search module is responsible for finding the best possible route on the road network from the current position to the destination while taking into account the options and criteria that have been set. The data of the digital road maps is stored as nodes and edges, which correspond to the intersections and to the roads interconnecting the intersections. These elements are assigned appropriate attributes that represent the "resistance," that is to say, the transit speed for the route search algorithm. Important attributes are the road class, length, driving direction, and driving constraints (e.g., tolls). The road class identifies the roads as being a freeway, expressway, highway or residential road. Using algorithms from the field of graph theory (e.g., A Star, Dijkstra 1959) the path of lowest resistance through this resistance network is then searched for in accordance with the set options and criteria. Common route options and criteria are:

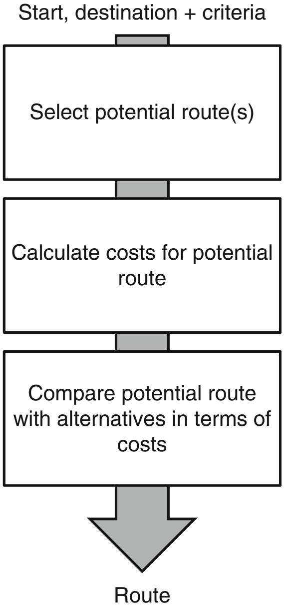
- Fast route: a route optimized for the shortest driving time possible
- Short route: a route optimized for the shortest distance possible
- Optimal route: a route that is a compromise between short distance and driving time
- Dynamic route: a route that takes into consideration traffic reports and the accordingly computed detours
- Avoidance criteria: for avoiding certain route sections, such as freeways, toll roads, ferries, and tunnels

2.4 Route Search Algorithms

The route search function is an important application of mathematical graph theory: within a graph – comprising V nodes (or vertices) connected by E edges (or lines) in which every edge represents a node pair – the aim is to identify an optimal route. Each node represents the place where several road segments meet, i.e., a point of intersection. An edge, on the other hand, represents a road segment of finite length and its route-relevant attributes – e.g., the road class, average speed on the segment, and length of the road segment. The road network is thus represented by a very large graph comprising nodes and edges with route-relevant edge attributes. The following are needed to perform a route search: a starting node and a destination node on the graph; a cost function $f(V_{ij})$ for evaluating the costs of traveling from node V_i to node V_j along an edge E_{ij}; and an optimization method for searching for the route on the graph that represents the optimum as regards the cost function. For the cost function, functions are generally used that are based on the attributes of the edges E_{ij}, such as length (find the shortest route), journey time (fastest route), and fuel consumption (eco route). Cost functions applied in practice often take into consideration several criteria: thus, for instance, when evaluating the cost of a fuel-efficient route, an additional optimization of the journey time is performed or a penalty is applied to the use of very low-order classes of roads. The route search now consists of finding a sequence of edges that are connected to one another, that connect the starting node to the destination node, and that constitute a global optimum with regard to the cost function over the entire sequence of edges ("optimum" in this case means "minimum" with regard to the cost function). A large number of algorithms is available for this.

Since navigation systems are, however, real-time systems that support a great many functions and are subject to certain constraints concerning their response times, it is not possible to utilize every algorithm. Only those algorithms that are particularly efficient in terms of computing time and memory use can be used – even if this results in only a suboptimal (with regard to the cost function) route being found, which nonetheless must be close to the global optimum. Figure 4 shows the basic principle of all algorithms: starting from an initial point, potential routes are selected on the graph, evaluated with regard to their costs, and compared with alternative routes.

Fig. 4 Basic principle of the route search algorithms



This evaluation continues until an optimal or nearly optimal route is found through the graph of the road network. The essential difference between the various algorithms lies in the way in which the search through the graph is performed. In the past, navigation systems employed methods that performed a route search from the starting point of route guidance toward the destination as well as in the opposite direction. Both search directions offer specific advantages and disadvantages. In the case of a "forward search," it is, for instance, possible to supply early driving recommendations, even though the route has not yet been entirely "found." This allows the system to provide the driver with a quick initial driving direction after the start of route guidance – though, with this there is the risk that the route may change again even in the area around the starting point, since the route search process is still running. In the case of a "backward search" from the destination area to the starting area, it is possible to store route alternatives that are found along the final route during the "alternatives search." They therefore constitute quickly accessible "alternatives back to the original route" that can be used if the driver has to deviate from the route (no time-consuming new search is then necessary). The "backward search," however, has the disadvantage that the driver has to wait until the entire route has been found before being supplied with the first driving direction. In future, only searches in a forward direction will be used. This is because, for parts of the road network, navigation map providers have meanwhile begun supplying details on average speeds as they vary with the time of day, known as "time variation curves." Since, in the case of a "fastest" route, the driving speed is relevant for calculation of the cost function, the system must search "forward" – this is because only a forward search will establish at what time of day one will reach an edge in the graph. This time of day information is required in order to select the correct "time variation curve" (average speed on the relevant road segment) for the cost estimation.

The constraints imposed by a navigation system with regard to response time, computing power and available memory mean it is not possible to utilize sophisticated optimization techniques, such as "simulated annealing" or "genetic algorithms," due to the associated computing time and memory requirements. A potentially suitable algorithm is the A* algorithm (A Star). It is an extension of the Dijkstra algorithm. Taking a node as a starting point, neighboring nodes on the graph are "visited" and the alternatives with regard to the cost function are examined. In order not to have to "visit" every node in the entire graph during the search, a heuristic process is used to examine the probable "residual costs" in various directions and thus influence the direction the search takes. Another possible alternative is the Ford-Moore algorithm. This involves performing a full search of the road network between the starting point and destination – though, in order to limit the necessary computing time, practical restrictions are placed on the area to be searched. Before the route search commences, the costs for each edge element are calculated in advance, since the edge elements are "visited" several times during the full search. To limit the computing time, the costs for an edge of the graph must only be calculated once, even if the edge is "visited" several times during the search.

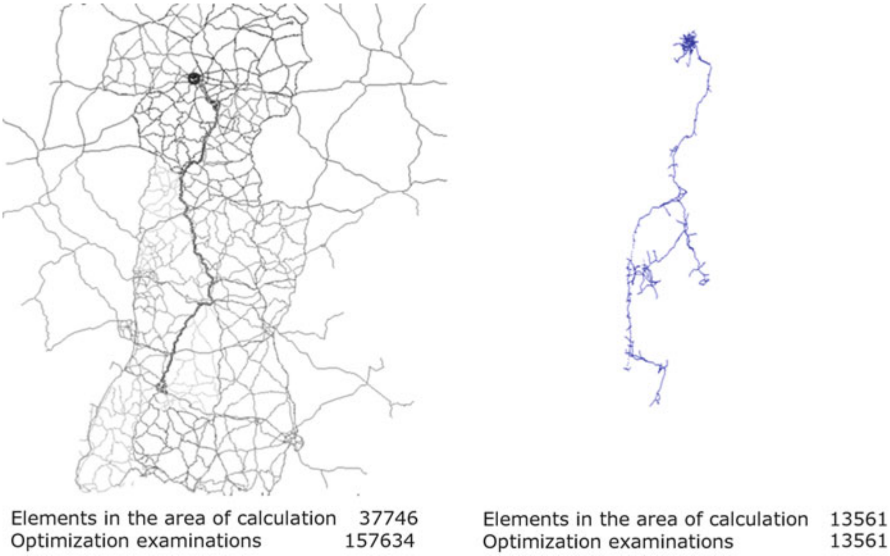


Fig. 5 Search strategies with the Ford-Moore algorithm (*left*) and A* algorithm (*right*)

Figure 5 provides an idea of the elements of the road network examined by the algorithms during an example route search. In the case of the Ford-Moore algorithm (left): during optimization, a search is performed on the entire road network that has been confined to a practical area between the starting point and destination. The A* algorithm (right) limits the extent of its examination by focusing its search on an area around the anticipated optimal route.

In practice, however, there are further constraints, such as the data structure of the digital map, that affect the choice of a suitable route search algorithm.

The route search in moving vehicles is in most cases not a one-time operation: if the vehicle deviates from the calculated route or if traffic reports are received, a new route has to be calculated. A recalculation of the route should take place as quickly as possible, and the following techniques are applied in order to ensure this is the case:

- Caching of data in faster memory if the data storage media are slow (CDs, DVDs)
- Utilization of data hierarchies (see Fig. 6): long roads of a high road class (e.g., freeways, highways) are held in a separate data network. The route search only searches on the lower hierarchy level at the journey's starting point and destination (residential, urban, and country roads) and calculates longer distances on higher levels (freeways and highways) so as to reduce the number of computational steps. The data hierarchies are connected to one another at nodes and/or edges via appropriate cross-references in the digital map data.

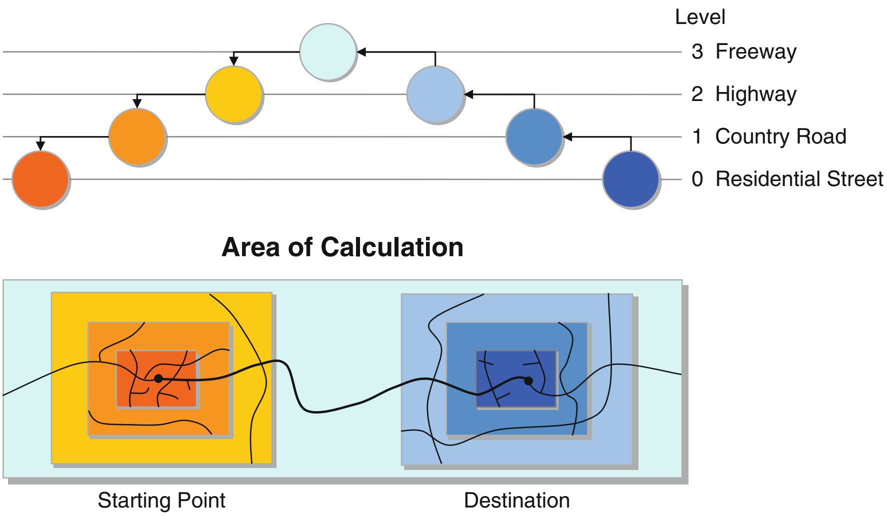


Fig. 6 Route calculation with the aid of data hierarchies

– Use of pre-calculated routes or interpolated points: sections calculated in advance are employed that do not have to be recalculated when needed.

The route search module makes the calculated route available to other navigation modules: the route guidance module for generating instructions for the driver; the map display module for displaying the route; and the user interface for creating the route list, i.e., the sequence for the route to be traveled.

2.5 Route Guidance

The task of the route guidance module is to provide the driver with information about imminent driving maneuvers in good time, but as clearly and concisely as possible. The route guidance module receives the current position from the position-fixing module and the route to be traveled from the route search module. From this information it generates driving recommendations like "turn right" or "follow the road." The system outputs the driving recommendations based on the road situation and current speed. Thus, on a freeway at high speed, it is important that the system provides an instruction to turn off right in good time (advance driving recommendation, e.g., "get ready to turn off right") and that it then repeats the instruction at a distance from the turning-off point or decision point so the driver has sufficient time – appropriate to the current vehicle speed – in which to react (e.g., "in 300 m, turn off right," "now turn off right," see Fig. 7). On residential roads, which are driven on at lower speeds, these intervals are significantly shorter. In addition to this, follow-up or linked driving recommendations (e.g., "now turn

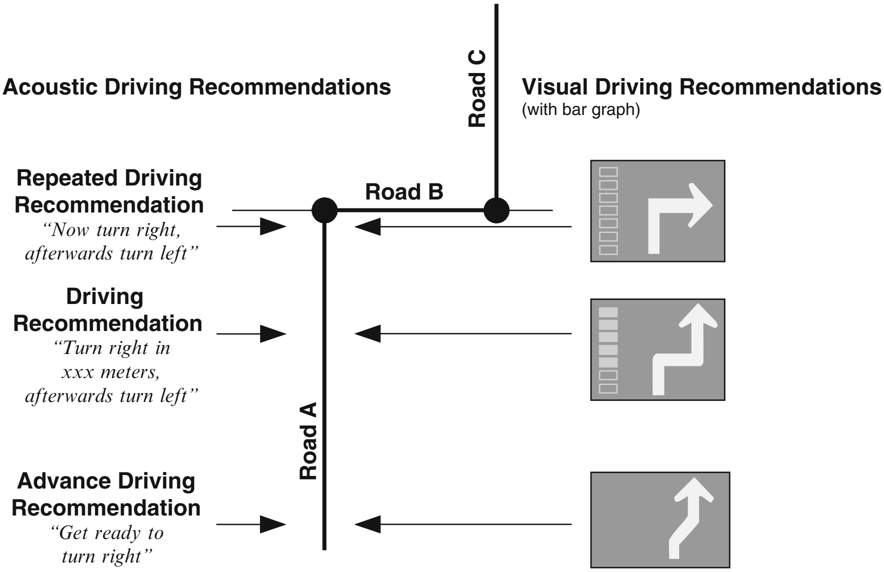


Fig. 7 Example of a route guidance situation

right, afterwards turn left") are generated that enable the driver to get into the correct lane. The announcements must be as clear as possible so that navigation even only with acoustic output is possible without distracting the driver's attention from the road. For this reason, a driving recommendation should not be given or should be delayed if the road situation is such that it would not be clear to which road the recommendation applies. The driving recommendations and grammar as well as the set of rules and parameters that specify the time sequence of the driving recommendations are specific to each manufacturer (i.e., application expertise gained through many years of experience). The driving recommendations are provided acoustically through voice output and visually by means of driving symbols, arrows and, if applicable, a map display (see Sect. 2.6). A visual distance indicator (bar graph) enhances the driving symbol display. Driving recommendations can be supplemented with a wide variety of additional information, such as a recommendation or indicator concerning which lane to use, the name of the road to turn into (turn-to info), and visual indicators for exits and entrances (highway entry/exit recommendations).

2.6 Map Display

Like route guidance, the purpose of the map display is to help drivers find their bearings. To generate the display, the map module receives the following data from the other modules:

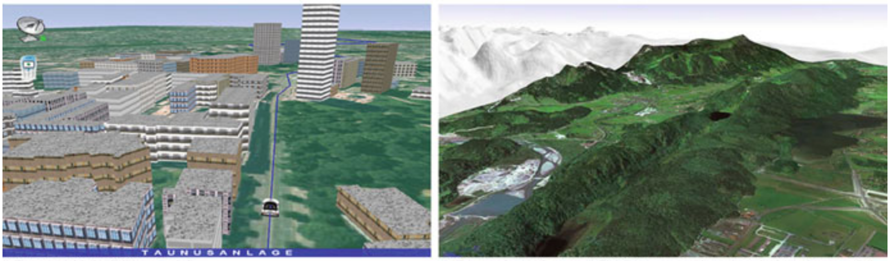
- The vector and bitmap data to be displayed is received from the data storage device or from the corridor that is stored in the cache. Vector data describes geometric shapes like roads, built-up areas and bodies of water. On the storage device, the vector data is enriched with such information as road classes (e.g., different color/width to differentiate between freeways and country roads) and other attributes (e.g., speed limits). In the case of the 3D map view, the vector data is supplemented with 3D building models for individual significant buildings (POI as 3D landmarks) or 3D models of entire regions (currently certain individual cities). Bitmap data is used in the form of satellite maps or textures for models of buildings. Often the map component converts this data into an internal representation that is suitable for displaying as quickly as possible (rendering optimized).
- The information about the position and movement of the vehicle is received from the position-fixing module so a position marker can be displayed at the current position to track the movement of the vehicle as smoothly as possible.
- The information about the current route from the route search module so that it can be highlighted on the map.
- The information about the next driving maneuver from the route guidance module so that the maneuver can be displayed on the map or in special views that provide the driver with a clear representation of the current situation: e.g., intersection zoom (the next intersection and the corresponding driving maneuver are shown enlarged), or highway entry/exit guidance (the entrance/exit is displayed).
- Information about traffic reports from the dynamic navigation module so the information can be displayed on the map.
- Information from the destination input module about POIs that are displayed on the map (e.g., filling stations, parking lots, repair shops, restaurants).
- Information from the user interface about the view that is set, i.e., information about the map area to be displayed (current position map, destination map, overview map), the map scale (usually from a 25 m details view to a 500 km overview) and the type of map (2D, tilted map including the tilt angle, 3D).

An overview of the types of map display is shown in Fig. 8.

A high performance map display is essentially dependent on the capabilities of the hardware used, that is to say, the power of the central processing unit and graphics accelerator. Displaying 2D and tilted perspective-view 2D maps smoothly at frame rates higher than 10 frames per second (fps) requires processors capable of more than 200 MIPS (million instructions per second) and 2D graphics accelerators that can independently display polygons. High-performance 3D map graphics require 3D graphics accelerators that are independently capable of displaying textured surfaces and that possess a z-buffer (depth information) for hidden surface determination in perspective view. To speed up the software, computation is also performed in the map display module using various data hierarchies. This allows the displayed road network to be thinned out when zooming out so the quantity of polygons to be displayed is reduced. A perspective-view map is displayed using

2D overview map displaying the route and traffic congestion

Tilted 2D position map (also referred to as 2.5D or pseudo 3D)



3D map with building models

3D map based on satellite images and a stored elevation model

Fig. 8 Examples of map displays

various levels of detail (LOD): that is to say, objects that are far away are not displayed with as much detail as closer objects. Furthermore, to ensure the map display looks good, a high resolution and the use of anti-aliasing for avoiding aliasing artifacts when displaying lines are important. These, however, place high demands on the system's computing power.

2.7 Dynamic Navigation

The dynamic navigation module has the task of taking external influences into consideration during navigation. Such influencing factors are typically traffic reports in TMC format that are received via radio broadcasting systems (FM-RDS, DAB) or via mobile communications systems (GSM, GPRS). Besides the freely receivable TMC messages that are generally made available by public broadcasters, there are also pay services (pay TMC), such as TMCpro. With its 37-bit coding, TMC is optimized for narrow-band transmission paths (originally FM-RDS had data rates of 60 bit/s). A TMC message contains the event (e.g., traffic congestion, accident, road closure, wrong-way driver) and location information (approx. 65,500 locations defined for each country using tables conforming to an international standard). Thanks to the information (i.e., distance and speed details) conveyed in the event, the dynamic navigation module is able to influence the calculation of a route: after the location contained in the message is represented on the digital road network, the dynamic navigation module can, for instance, increase the resistances of individual nodes/edges according to the disruption. Thus the calculation results in an appropriate detour being generated. A recalculation of the route in this way either takes place automatically ("dynamic route" option has been set) or the user is informed about the new traffic situation and can initiate a recalculation of the route (user-confirmed dynamic navigation).

Besides factoring the event information into the route calculation, the dynamic navigation module is also responsible for processing the encoded traffic reports for displaying in readable form and for making the events portrayed on the road network available to the map display module. This allows the driver or front passenger to get a visual impression of the traffic situation. Besides this visual representation, it is also possible to instruct the route guidance system to provide the driver with an acoustic alert when approaching the traffic disruption ("caution, 2 km long traffic jam"). For higher bandwidth transmission paths (e.g., DAB, WLAN), new standards (TPEG) are in planning that will allow both the event and location information to be specified even more precisely.

2.8 Corridor and Data Abstraction (Data Storage Devices)

The digitized road data is made available to the navigation system as mass data on storage devices. For a country the size of Germany, several hundred megabytes are necessary. Besides the typical data storage media like the CD, which generally covers a single country, and the DVD, with continental coverage (e.g., Europe), electronic mass storage devices are nowadays increasingly being used, like flash memory, SD cards, and hard disk drives. The properties of the data storage device have an influence on the navigation system's range of functions and performance. Important factors in this respect are:

- The volume of data, which not only puts a limit on the regional coverage (individual countries vs. Europe/North America) but also on the range of functions; for certain functions like speed limit information, data attributes must be added that take up storage space
- The access time, which has a significant effect on performance, particularly when data are needed that are spread over the storage medium – such as when calculating long-distance routes over wide areas, when inputting a destination (traversal of the index tree), or during system start-up (when a wide range of data must be read)
- The data transfer rate, when larger volumes of data must be read, such as for the map display
- Additional factors, like wear or dirt that affect rotating optical media like CDs and DVDs

In order to wholly or partially avoid accessing data from these sometimes slow data storage devices, many navigation systems use a corridor function (see Sect. 2.1). The corridor function strategically places the required data into electronic memory in advance so it can subsequently relay this data from cache. The data cached in this way relates in particular to the area around the current position, along the route, or around the destination.

3 Off-board Navigation

When all of the navigation system's subtasks, such as position fixing and route calculation, are performed in the vehicle, this is referred to as autonomous or on-board navigation. In the case of off-board navigation (OBN), certain subtasks, such as route calculation, are performed on an external stationary server. The computed data and information are then transmitted from the server to the device in the vehicle via an air interface (see Sect. 7). With the appropriate air interface design (bandwidth) and sufficient server computing power, there are no limits to such a relocation of subtasks. In the extreme case, only the position-fixing sensors and components necessary for input and output would remain in the vehicle. An established configuration, though, is that the server handles the destination input, route search and dynamic navigation tasks, while the processes that are more time critical – such as those relating to position fixing, route guidance and the map display – remain in the vehicle (see Fig. 9).

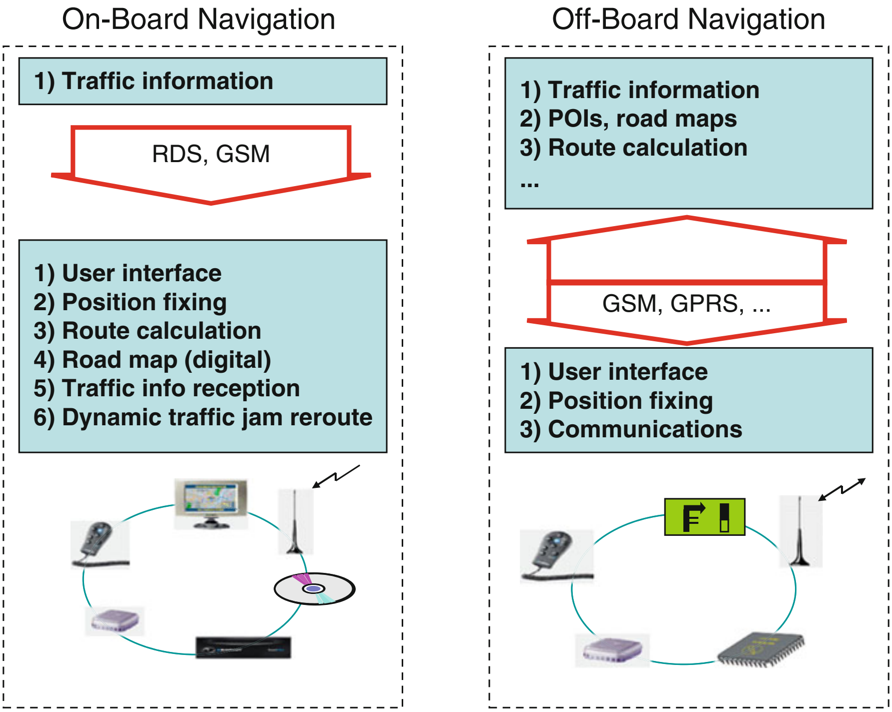


Fig. 9 System comparison of on-board versus off-board navigation

The advantage of off-board navigation compared with on-board navigation lies in the up-to-dateness of the data, which can be better managed on a server. No benefits exist, however, in terms of material costs. In the case of OBN, the data storage device and the drive for importing the data are replaced by the communications unit (e.g., a GSM module); other than that, the same components are required as for on-board navigation. OBN has become established as the standard application in the area of mobile communications for mobile phones with GPS support.

4 Hybrid Navigation

Future requirements on navigation systems are: dynamic updating of points of interest (POIs); access to data or routes against payment of a fee; inner-city dynamic navigation; server-based (off-board) navigation; and map displays using virtual reality and satellite imagery (see Fig. 10). The majority of these objectives require so-called "hybrid navigation," which is characterized by the fact that in order to perform the navigation functions, the system must access a multitude of data sources. Such data sources can be distributed between the vehicle and infrastructure in almost any way possible. For reasons of cost, the efficient transmission of data by means of mobile communications and radio broadcasting is of major importance here. It is furthermore important that methods be developed that facilitate the portrayal (georeferencing) of data from different sources and in

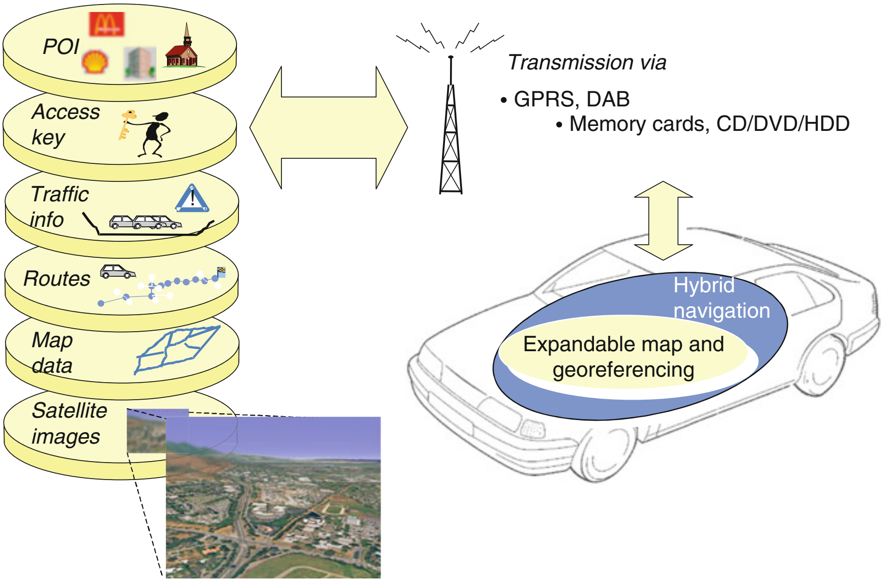


Fig. 10 Hybrid navigation through utilization of data sources external to the vehicle

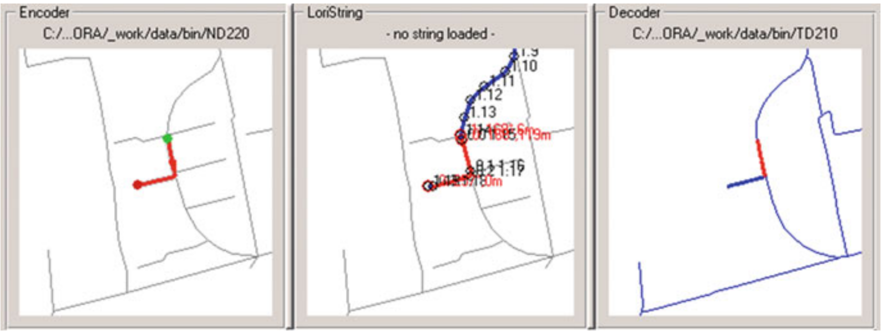


Fig. 11 The AGORA georeferencing method illustrated by means of an example

particular on the digital map available in the vehicle. Simply transmitting the coordinates is not enough, since the correct data item must be identified from the vast number of possibilities.

Referencing methods tailored to specific applications already exist, though they currently do not meet the requirements for flexibility and independence from the utilized maps (in terms of precision and manufacturer independence). A new method, known as AGORA, has been developed within the scope of an EU project under the same name (Hendriks et al. 2005; AGORA Project 2015) and standardized by the ISO.

Using a correlation method, elements from a detailed map can be inserted into a simpler map. The standardization process for this method has been completed (see Fig. 11).

4.1 Map Data: Up-to-Date and Customized

The currentness of the data of any purchased digital map will dwindle as time goes by. The age of a map becomes apparent in the form of deviations from the real-world situation, for instance, when new or altered roads are missing or signs have been changed. If one considers the map to be a sensor, then this sensor supplies data that is incorrect in an unsystematic but reproducible way. Up to now, it was the user who decided when the errors were no longer tolerable and when an update should be performed through the purchase of new map data. The more functions in a vehicle depend on the map data and, in particular, the more these functions are of relevance to safety, the more the decision to make necessary updates to the map data must be made independently of the user. For this reason, various methods are currently under development that will enable map data to be updated automatically (see also ▶ Chaps. 26, "Digital Maps for ADAS," and ▶ 28, "Backend Systems for ADAS").

A controlled, standardized and therefore efficient replacement of parts of the data is being prepared in the "Physical Storage Initiative," which – alongside

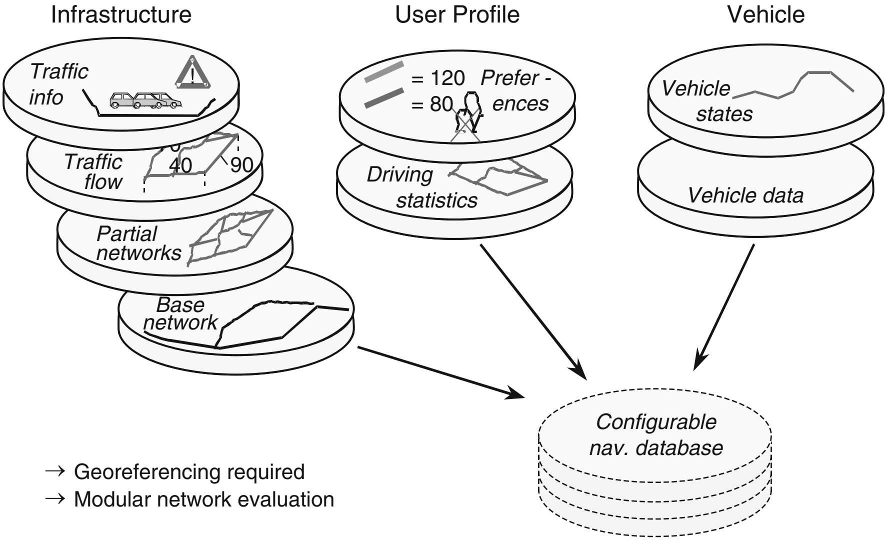


Fig. 12 Configurable adaptive digital map

standardizing the data storage format across OEMs – is defining the mechanisms for the incremental replacement of data. Other methods intend to enable the navigation system to recognize and correct errors in the database itself. Within the scope of the EU-funded projects ActMAP (2015), Otto (2005) and FEEDMAP (2015), researchers investigated options and methods for keeping the map up to date. The range of potential applications already being contemplated today is so diverse that it would entail an enormous effort to keep all the information available in a data pool at all times. Rather, it is probable that a base map will be enriched with additional information from other sources on an application-specific basis. The vehicle too will itself collect information and store it as location-based data on the digital map so it can be used on subsequent journeys (see Fig. 12). This is particularly useful for frequently traveled routes, like the route to work, on which significant fuel savings can be achieved through economic fuel-optimized driving. Even though the map as a whole would only be marginally supplemented with data locally, up to 80 % of one's personal route would be covered.

Optimal assistance can only be provided through adaptation to the driver and his or her preferences. The ability to modify attributes on the map and to add new custom ones is therefore necessary. On-board technologies and processes that provide vehicles with a degree of autonomy can be utilized to add driver- and vehicle-specific attributes to the standard map, which is initially supplied without any customizations. Another application scenario for an adaptive map involves personalization of the map for particular groups of drivers, such as senior citizens. Although older people can manage routine tasks well, the risk of making mistakes when performing unfamiliar tasks, on the other hand, increases markedly with age (Förster 2001; Krüger et al. 2001). The conclusion therefore is that unfamiliar and complex situations should be avoided. Thus, routes and intersections that the user has already used can be marked and rated for that particular user on the digital map, depending on whether they are judged to be "problematic intersections," "high stress routes" or "easy to use," etc. The system must infer such assessments from the user's driving behavior while taking into consideration the road and traffic situation. Since every driver perceives difficulties, demands and stressful situations differently, a personal map inevitably emerges from this. The system will then select a route that preferably uses familiar and "stress-reduced" stretches and of course avoids driving situations that have a high accident risk, such as those identified by (Krüger et al. 2001).

5 Assistance Functions

The term "assistance functions" is very broad and is not used consistently. It ranges from functions of an informative nature (to some extent navigation is even regarded as an assistance function) to safety-related functions that can intervene in the handling of the vehicle.

A distinction can be made between four classes of assistance functions: functions that enhance driving safety (e.g., ESC, brake assist); functions that increase driving convenience (e.g., automatic parking, traffic messages); functions for reducing the fuel/energy requirement (e.g., gear selection recommendation, coasting assistance); and functions that boost vehicle performance (e.g., deactivation of the air conditioning during acceleration). The navigation system can play a decisive role in one of these assistance functions. In the following paragraphs, a distinction is made between navigation-based and navigation-aided assistance functions.

Navigation-based assistance functions are induced and made available by the navigation system itself. Examples of such assistance functions are:

- The "congestion ahead" warning, which enables the driver to take the next exit and drive around the congestion or at least approach the tail end of the congestion at an adjusted speed
- The curve warning assistant, which warns the driver if the speed is too high for the bend ahead
- The danger spot warning assistant, which warns the driver about hazardous spots (e.g., high collision zone, kindergartens/schools)

In the case of navigation-aided assistance functions, the navigation system acts as a sensor for other assistance functions that are typically implemented on other control units. Via a standardized interface (ADASIS 2015), the navigation system supplies the situation at the current position, the route during activated route guidance, and the road network ahead (also called the ADAS horizon or electronic horizon). Using the so-called AHP (ADAS Horizon Provider), the data is broken down on the transmitting side into messages and made available on the vehicle bus. On the receiving side, the AHR (ADAS Horizon Reconstructor) reassembles the

ADAS horizon from the messages. Thanks to standardization, the assistance functions are independent of the navigation system used.

Examples of navigation-aided assistance functions are:

- Adaptive beam assist – for better illumination of bends and intersections
- Predictive gear selection appropriate to the route profile – performed by the automatic transmission to save fuel
- For further examples, see (Nöcker et al. 2005).

The existence of new assistance functions is leading to new sensor technologies becoming established in the vehicle environment, such as video cameras and RADAR. Navigation systems also benefit from the presence of such technologies. Thus, for instance, by analyzing the camera image, it is possible to identify the precise lane in which the vehicle is situated. The navigation system will in future be able to utilize current information (such as detected road sign information or resolved exceptional situations, like a construction site) to improve the driving recommendations and alerts that it provides the driver.

6 Electronic Horizon

A navigation system can also serve as a sensor for other assistance functions. In this case, location-based information is typically transmitted from the navigation system to another control unit in the vehicle; the transmitted location-based data is referred to as the electronic horizon (see Fig. 13).

The electronic horizon is transmitted using the standardized interface protocol ADASIS (Advanced Driver Assistance Systems Interface Specification) (ADASIS 2015). Thanks to this standardization, the assistance functions are independent of the navigation system used.

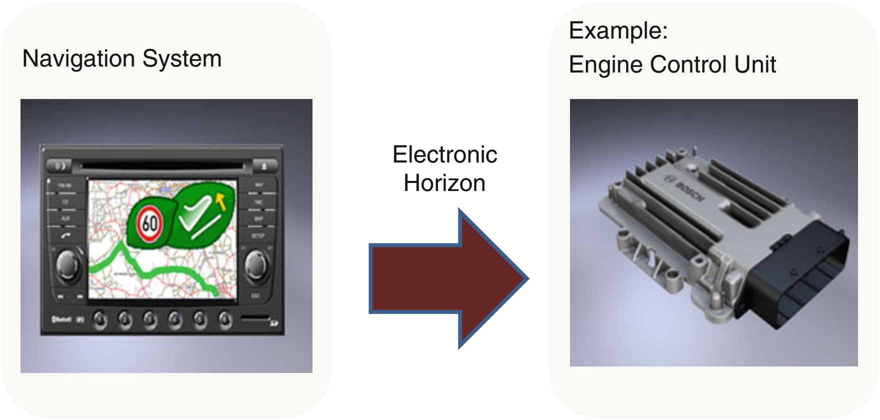


Fig. 13 The navigation system as a sensor (Source: Robert Bosch GmbH)

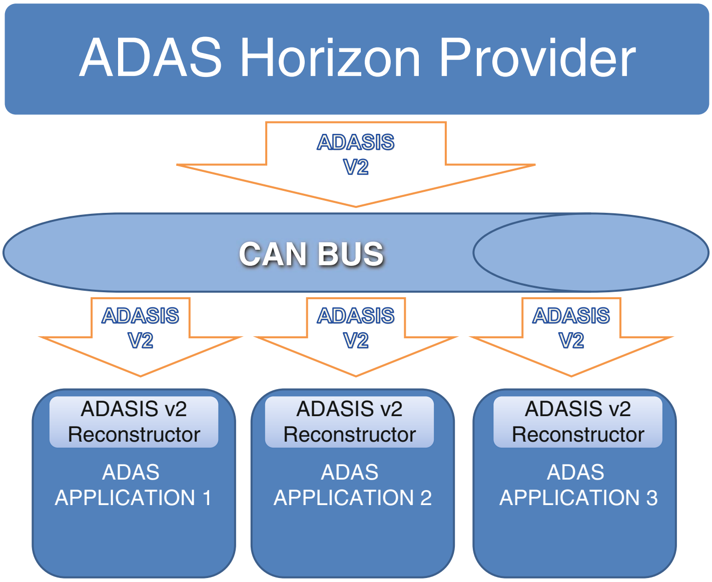


Fig. 14 Basic principle of the ADASIS interface protocol (Source: (ADASIS 2015) v2 Protocol)

The basic principle of the ADASIS interface protocol is that the so-called AHP (ADAS Horizon Provider) breaks down the data on the transmitting side into messages and makes them available on the vehicle bus. On the receiving side, the AHR (ADAS Horizon Reconstructor) reassembles the electronic horizon from the messages (see Fig. 14). In this context, the CAN bus is typically used as a communication channel.

The electronic horizon contains information about the road network ahead, though only those roads are relevant that can be reached from the vehicle's current position and that will also actually be driven on with a certain degree of probability. The route that will be driven along with the greatest probability is referred to as the MPP (Most Probable Path). All the information and the current position of the vehicle are specified with relative offsets with respect to the possible paths (see Fig. 15).

The electronic horizon's prediction distance is typically around 500 m to 6 km, depending on the requirements of the assistance function that is using it. Due to the way in which the information is encoded, the prediction distance for the content of the electronic horizon is limited to around 8 km when the usual length increment of 1 m is employed.

Besides typical information from a digital map used for navigation systems (such as road classes and speed limits), specialized data, in particular information about road gradients and bends, is also transmitted in the electronic horizon. Special demands are also placed on the quality of the location accuracy of information in the electronic horizon. For this reason, this data is generally offered by map data providers as additional "ADAS data" and marked accordingly on the digital map.

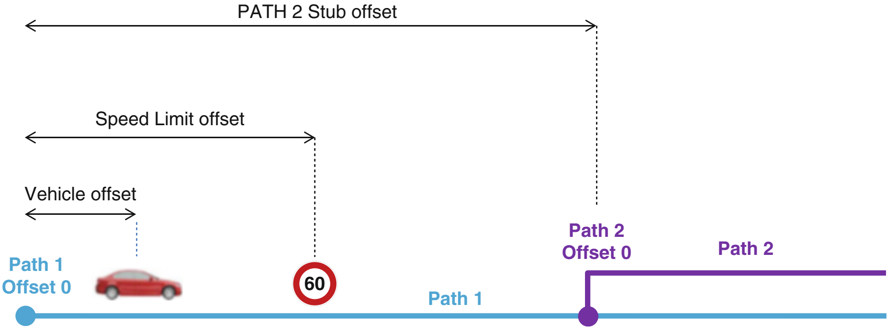


Fig. 15 ADASIS position coding (Source: (ADASIS 2015) v2 Protocol)

Utilization of the electronic horizon for assistance functions also requires that the navigation system can supply very accurate position data. Only very good position fixing will provide an accurate and robust electronic horizon. For this reason, only navigation systems are used that are permanently installed in the vehicle and that make use of the vehicle's sensor systems for optimizing the position computation.

7 Traffic Telematics

The word "telematics" is a blend of the words "telecommunications" and "informatics."

The term "telematics" is used in various specialist fields (e.g., medical telematics and building telematics) and is therefore not clearly defined. Telematics systems in traffic telematics generally comprise the following elements:

- A stationary services server with a telecommunication or broadcasting facility for processing and transmitting data
- A (mobile) terminal device with a telecommunication facility for receiving data
- A local computer in the terminal device that provides the user with functions based on the data from the stationary services server or that transmits data to the stationary server so that the server can provide services

The stationary services server can be omitted when data are exchanged and processed directly between the mobile terminal devices (see Sect. 7.4).

The transmission of the data via the air interface – that is to say, the transmission path by means of electromagnetic waves – can be approximately divided into two categories:

(a) Radio broadcasting-based technologies: These technologies allow information to be transmitted to a large number of recipients; the communication is unidirectional, non-personal, and takes place over a wide area of coverage.

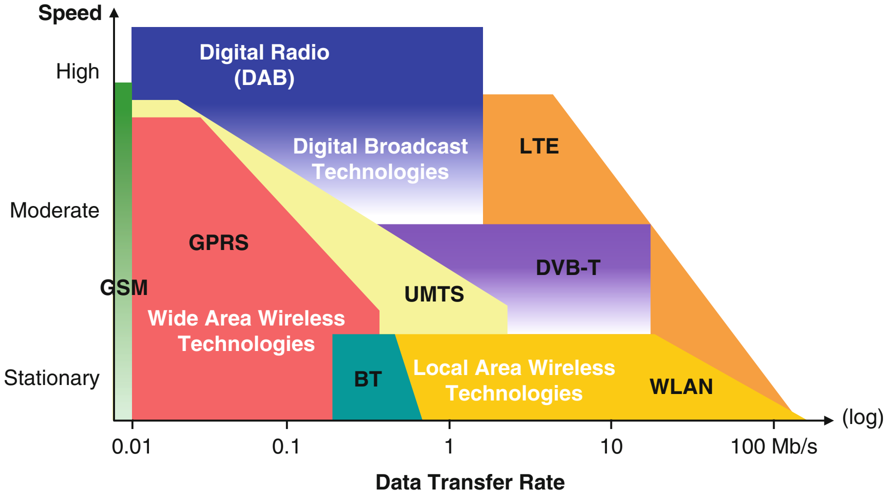


Fig. 16 Data transmission technologies for telematics services

(b) Mobile communication-based technologies: These technologies allow information to be transmitted specifically to an individual or just a few recipients; the communication can be bidirectional, is personal, and generally takes place over a limited area of coverage.

Besides this classification into categories and the associated suitability for particular telematics services, the technologies furthermore differ in terms of the data transfer rate with dependence on vehicle speed. Telematics services for motor vehicles require transmission technologies that are insusceptible to traveling at speed (>150 km/h) (see Fig. 16).

7.1 Radio Broadcasting-Based Technologies

Radio broadcasting-based technologies can be divided into analog and digital transmission modes.

Alongside the transmission of the radio program and thus the transmission of spoken (traffic) information, analog transmission modes (e.g., FM) also allow data to be transmitted on a very narrow-band channel. This data contains, for instance, radio text for displaying on the radio/navigation system's HMI or RDS-TMC information for dynamic navigation (see Sect. 2.7 Dynamic navigation). Although the analog receiver technology can also be used at high speeds, it is susceptible to interference: multipath propagation due to reflection from buildings, hills and mountains may occur; shadowing in the direction of propagation can disrupt reception; and the Doppler effect diminishes the signal quality. To improve reception, multi-tuner designs or digital tuner designs (not to be confused with digital

Table 2 Analog and digital radio broadcast transmission technologies

Analog	Digital		
	Europe	USA	Republic of Korea/China
VHF (87–108 MHz)	DAB	HD Radio	DMB
MW (530–1,710 kHz)	DRM (Digital Radio Mondial)	HD Radio	DRM
LW (148–284 kHz)	DRM		DRM
SW (3–30 MHz)	DRM		DRM
TV terrestrial	DVB-T		
Satellite		SDARS	

transmission modes) can be used. Digital tuner designs employ a strategy of digitizing the incoming analog signal as early on as possible (e.g., digitizing the signal after mixing down to the intermediate frequency). Using modern mathematical methods in digital signal processing, it is thus possible to form a directional antenna characteristic via two antennas.

Digital transmission modes (e.g., DAB – Digital Audio Broadcasting) have been developed to minimize reception interference due to environmental effects (e.g., multipath propagation, etc.) while at the same time increasing the data rate. The transmission can take place via stationary terrestrial transmitters (e.g., DAB) or via satellites (e.g., SDARS – Satellite Digital Audio Radio System).

The significant increase in data rate up to 1.5 Mbit/s in the case of DAB enables the transmission and displaying of complex data and images in a telematics system. An overview of analog radio broadcast transmission modes and the advanced digital transmission modes is shown in Table 2.

Compared with mobile communications, the distribution of information using broadcasting technologies is a low-cost way of disseminating information that is of interest to many people. New services made available by means of digital transmission modes will therefore continue to increase the importance of broadcasting technologies in the vehicle environment alongside the increasing importance of mobile communications. These services are already today transmitting data that is of interest to many customers, such as weather and traffic situation services via satellite broadcasting (SDARS) in North America. Digital broadcasting cannot, however, be utilized on a customer-specific or vehicle-specific basis, which is why mobile communication-based services are required.

7.2 Mobile Communication-Based Technologies

Mobile communications according to the GSM standard (Global System for Mobile Communications) is widespread in Europe. GSM has a cellular network structure with terrestrial base stations that have a varying cell radius depending on local conditions. The maximum cell radius can be up to 35 km.

An FDMA/TDMA combination is used on the physical layer for GSM. Two frequency bands with a 45 MHz band gap are reserved for GSM operation: 890–915 MHz for the uplink and 935–960 MHz for the downlink. Each of these bands of 25 MHz width is divided into 124 individual channels spaced at 200 kHz. The frequency channels are uniquely numbered, and a pair of the same numbers from the uplink and downlink forms a duplex channel with a duplex spacing of 45 MHz. Each of these channels (200 kHz) is divided into 8 TDMA channels (8 time slots). The uplink channels are transmitted with a delay of three time slots compared with the downlink. In the uplink and downlink, a mobile station will use the same numbered time slot in each case, so transmission and reception do not have to take place simultaneously. Low-cost terminal devices can be offered because no duplex units are necessary for this.

Several technologies for mobile communications-based data transmission are explained below.

Circuit Switched Data (CSD) is the name for the transmission mode in which a data link is established from the mobile terminal device to a remote terminal (e.g., server). The connection is technically comparable to a voice connection, and the payload can be transmitted at 9.6 kbit/s.

High Speed Circuit Switched Data (HSCSD) is an enhancement to CSD. To make a higher bandwidth available, several time slots are combined leading to a payload rate of 38.6 kbit/s.

In the case of both CSD and HSCSD, the utilized time slots are permanently unusable for voice data; alternatively the data are transmitted via SMS (Short Message Service). The GSM standard defines several logical channels for the transmission of (voice) data and signaling data. SMS are transmitted over the signaling channel, thus enabling the simultaneous transmission of voice data and SMS. This is of significance to some telematics services, since a call can be made and data transmitted at the same time, for example, when an emergency call is made while position data is sent via SMS. An SMS comprises a header and the content of the message, which is limited to 1,120 bits (160 characters in the case of text messages). It is possible to link together up to 255 short messages. An SMS is not sent directly from one terminal device to the other terminal device; instead it is transmitted via an SMS service center, thus also allowing the message to be stored temporarily.

SMS makes it possible for telematics applications to offer push services: if the phone number is known, information can be actively sent from outside exclusively to a terminal device without the information having been requested (e.g., advertising).

General Packet Radio Service (GPRS) makes packet-oriented data transmission possible: the data are grouped by the sender into individual packets, transmitted and then reassembled at the receiving end. GPRS technology allows a practical data rate of up to 53.6 kbit/s. One of the advantages is that a virtual connection is established that is only occupied during data transmission. Billing of the data is thus dependent on volume and not time, as is the case, for example, with CSD.

The **Wireless Application Protocol** (WAP) is a protocol for making internet content accessible in the case of slow transfer rates and long response times in mobile communications; the HTTP internet protocol is used for communication. The off-board navigation system described in Sect. 3 uses the WAP protocol for transmitting the navigation data. Furthermore, push services are possible with WAP.

Universal Mobile Telecommunications System (UMTS) is the name of the third-generation mobile communications standard. UMTS provides data rates of up to 384 kbit/s (out of town at 500 km/h) to up to 2 Mbit/s (in town at 10 km/h). UMTS makes it possible to provide telematics services that require a high data rate (e.g., the transmission of video data).

The performance of mobile communications is being enhanced still further through the current introduction of the fourth-generation mobile communications standard known as **Long Term Evolution** (LTE). This will result in increased data rates (in the range of >100 Mbit/s for the downlink) and reduced latencies for transmitted information (in the range of 5–100 ms). The reduction in latencies makes the LTE standard highly appealing for telematics functions that depend on short transmission times, e.g., certain warning functions based on the transmission of data from one vehicle to other vehicles.

Bluetooth (BT) is an industry standard for the wireless transmission of data and voice. Transmission takes place in the 2.4 GHz ISM band and has a range of 10 m to max. 100 m. Bluetooth is suitable for quasi-stationary use; the data rate is 723 kbit/s. BT modules are very compact and, compared with other mobile communications modules, are inexpensive and have a low power consumption. BT offers a range of profiles, several of which can be used well in the vehicle environment. An example is the Hands-Free Profile (HFP) for using a hands-free facility integrated in the radio/navigation unit with an external mobile phone. This permits utilization of a low-cost BT module in the navigation device to make use of the mobile phone's more cost-intensive GSM module.

A further example is the Phone Book Access Profile (PBAP) for exchanging phone books between the radio/navigation unit and the mobile phone.

BT gives users the opportunity to introduce consumer electronics (CE) devices into the car environment (see also Sect. 9).

A further option for exchanging data in quasi-stationary mode is via the **Wireless Local Area Network** (WLAN). Similar to BT, the transmission takes place at 2.4 GHz or 5.4 GHz, depending on the standard used. With the established standards, the transfer rate is 54 Mbit/s, and, as is the case with BT, the range is limited: 100–300 m. WLAN can thus be used for data exchange at dedicated access points; services of this kind are used in the vehicle-to-infrastructure (V2I) context (see Sect. 7.4).

7.3 Telematics: Basic Services

Besides being able to categorize telematics services according to the type of communication technology employed, it is also possible to classify them according

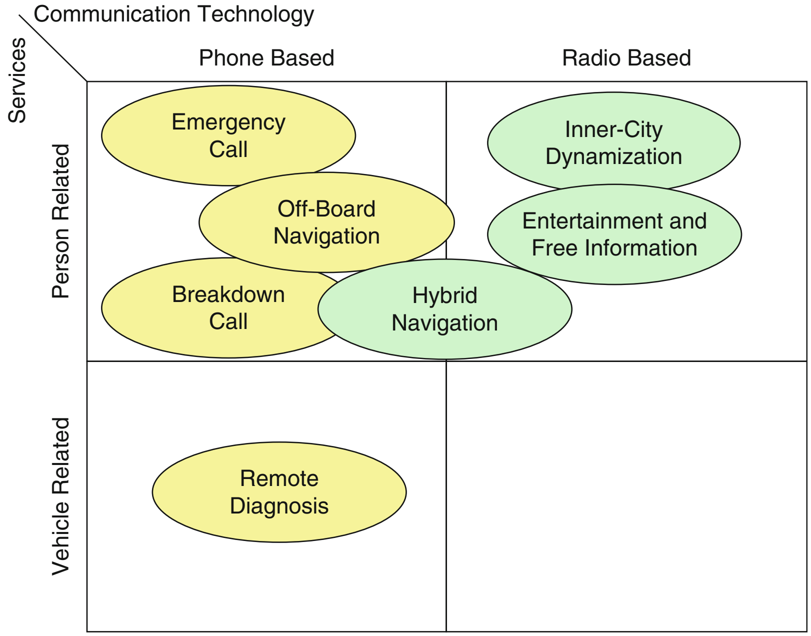


Fig. 17 Categorization of telematics services

to their type of utilization. A distinction between person-related services and vehicle-related services (see Fig. 17) is useful in this context. Below, some examples of basic telematics services are provided.

Communications: Information exchange with a call center or server takes place in an automated way via a voice call, SMS or email. This information can be personalized for each customer.

Safety: An emergency call (eCall) can be made automatically or manually – for instance, after airbag deployment. The emergency call can go out to a service center or directly to a rescue services control center. At the same time as the outgoing emergency call, the location data – determined via GPS – is typically sent too (e.g., by SMS) in order to initiate the emergency response. Within the scope of a European initiative to reduce the number of road traffic-related injuries and deaths, the development of a Europe-wide eCall system is in planning. To achieve this, appropriate communications equipment must be kept in all vehicles, standardization implemented (e.g., a standardized emergency phone number in Europe), and a services infrastructure developed (emergency response system).

A breakdown call can be made if the driver manually triggers it. In addition to establishing a voice connection, it is also possible to transmit the vehicle's location and diagnostics data. The diagnostics data can be retrieved over the vehicle network either automatically or after initiation by the user. The service center personnel can then use this data to decide whether a repair of the vehicle is possible or whether the vehicle must be towed. Furthermore, based on this data, it is also possible to order replacement parts and schedule the repair work.

In contrast to the breakdown call, an off-board diagnosis can be initiated by the user or service center independently of a fault. In predictive systems, it is possible to ascertain how much mileage particular components have seen and whether it is necessary to make repairs.

Route guidance: The relevant "off-board navigation" and "hybrid navigation" services are described in Sects. 3 and 4.

Convenience functions: These services facilitate the remote control of components in the vehicle, such as opening doors and activating the auxiliary heating system. They are to some extent of relevance to safety and are therefore only offered with restrictions by OEM manufacturers. A further potential convenience function is the ability to manage a driver's logbook via an external server: this involves the user logging the beginning and end of a journey on the server. In addition, further data like the fuel consumption is also documented.

Traffic information is generally broadcast by radio and is thus made available to many users. This information can be used to make the navigation and route guidance dynamic (see Sect. 2.7 TMC). Traffic information can also be transmitted on an individual basis, i.e., by request, which can be realized in an off-board navigation solution. In this case, the external navigation server is queried at regular intervals to find out whether there are any traffic problems that could affect the current route guidance. In order to keep the amount of transmitted data as small as possible, a fingerprint of the route is sent to the server for comparison.

General information can be accessed via the internet, and this can include the option to download entertainment content, such as music files. Downloading control unit software is – despite the availability of the technology – nowadays generally not offered in the OEM business, since the risks outweigh the possible benefits, particularly when it comes to safety-related components.

7.4 Vehicle-to-Vehicle Communication and Vehicle-to-Infrastructure Communication

In addition to communication by the vehicle using radio broadcasting-based and mobile communication-based technologies, there is increasing demand for bidirectional communication with other vehicles. A distinction is made between communication among vehicles – vehicle-to-vehicle (V2V) – and communication of the vehicle with infrastructure components – vehicle-to-infrastructure (V2I). These communication scenarios make further telematics services possible (Eberhardt 2003) (see also ▶ Chap. 27, "Vehicle-2-X").

In all the scenarios, a mixture or supplementation of V2V and V2I is possible. At certain danger spots, local data can be transmitted to individual vehicles (V2I) that is then passed on along a chain of vehicles (V2V). To make services for V2V and V2I available to the automotive mass market, solutions must first be found to address various challenges.

A reliable communications system is needed that should, if possible, be available free of charge. To achieve this, the requirements must be broken down into the individual layers of the OSI layer model:

- For communication from the vehicle toward the back and front, the range should be approx. 1,000 m; the range to each side has been estimated at approx. 250 m.
- The number of participating members of the network is subject to considerable variation (e.g., the traffic on a country road compared with that in the inner city). To prevent collisions on the individual transmission channels at maximum transmitting power, the transmitting power must be scaled depending on the situation.
- Vehicles may be moving at high speeds and thus also at high relative speeds with respect to one another; the associated Doppler effect must be compensated.
- Particularly in inner-city traffic, effects like shadowing and reflections or multipath propagation (caused by buildings) must be considered.
- In the case of safety-related applications, it must be ensured that the connection is uninterruptible and immune to interference. Furthermore, safety-related data must be transmitted with minimal delay time (prioritization) so it is always given priority over data from entertainment applications.
- Suitable routing and forwarding strategies must be implemented so that information can be forwarded with precision to the specific recipients. An item of information about traffic congestion may, for instance, only be relevant to vehicles on a freeway, while it may be unimportant for vehicles on a parallel highway.
- A standard must be established that allows vehicles from different manufacturers and of various product and vehicle generations to communicate with one another.

A potential standard for vehicle communication that is currently under development is IEEE 802.11p/IEEE 1609, which is an advanced flavor of the Wi-Fi standard IEEE 802.11. To meet the challenges of V2V and V2I, it operates in a dedicated frequency range and has been designed for data communication at speeds of up to 200 km/h.

7.5 Toll Systems

With simple toll systems, the toll is either collected at points of payment in the toll zone or by means of a toll sticker that permits road utilization in a restricted region – generally limited to one or more countries – and for a limited time period. These implementations do not permit custom charging with time- and location-related precision. Further disadvantages of these implementations are, depending on the specific solution, the interruption to the traffic flow and the complex monitoring of a wide area.

A complex toll system was introduced in Germany in the form of the ETC system (Electronic Toll Collection): with this system, the toll is levied for trucks on freeways according to the user pays principle, without interruption to the traffic flow.

The main components of this system are the dual toll collection system, the monitoring system, and the operational center for controlling the processes. The dual toll collection system gives the freeway user the option to participate in the automatic collection system or use the manual log-on system.

In the case of automatic collection, an autonomous on-board unit (OBU) with a DSRC module and GSM/GPS combination antenna ascertains the toll amount and transmits this information to the billing center using mobile communications. The driver enters vehicle-specific information into the OBU (e.g., axles, weight, etc.) to activate the vehicle for the automatic system. The OBU uses GPS and the stored freeway network to autonomously detect which toll route sections on a freeway are being used. Based on this route data as well as the preprogrammed tariff data and vehicle data, the amount of toll to be paid is then calculated, stored and sent via the integrated GSM module to the toll collection center. The manual collection system is intended for occasional users: they have the option to log on over the internet or at stationary payment terminals at rest areas and filling stations. To detect who has not paid the toll or who has paid incorrectly, payment is monitored: this can be performed automatically using enforcement gantries, by means of stationary checks (e.g., at freeway parking lots), or through mobile enforcement vehicles in passing. The automated enforcement gantries bridge the entire road and are equipped with detection equipment for each lane. Approaching vehicles are registered by the gantries using laser distance sensors so the vehicles can be assigned to individual lanes. Measurement sensors with 3D laser distance scanners then scan the vehicle to classify it and determine whether a toll is liable for it. CCD cameras with LED flashes make an overview image of the vehicle and record the vehicle registration, which is automatically detected and analyzed. The ascertained data is communicated to the central database via ISDN/GSM technology where it is compared with the data that was transmitted earlier via the OBU or the stationary toll terminals (Daimler Chrysler 2003).

7.6 Modern Traffic Control

Modern traffic control systems collect current traffic data over a wide area and then use it to generate information for road users and forecast developments in the traffic situation. Traffic flows can then be selectively influenced using these forecasts.

An example of a successful introduction of a modern traffic control system is the Vehicle Information and Communication System (VICS) in Japan. The public sector began introducing VICS in 1996, starting with the metropolitan areas of Tokyo and Osaka; by 2003 the system had been introduced throughout Japan (VICS 2015). The data is collected by the police and the Road Administration. It is then forwarded via the Japan Road Traffic Information Center to the VICS Center, which informs drivers in real time about the current traffic situation in text and graphical form – provided that an appropriate navigation device is installed in the vehicle. The processing of the data into graphical form involves classifying the traffic flow on individual roads according to color (red, yellow, and green) and then displaying this on the navigation system's overview and detail maps, thus enabling the driver to judge the traffic situation. Information is transmitted concerning traffic disruptions, estimated driving time, accidents, construction sites, speed limits, blocked roads, and parking space availability. Three channels of communication are used to distribute the information:

- FM radio broadcasting with broad coverage up to a range of 50 km to transmit all the previously mentioned information at an interregional level (the traffic situation within a surrounding radius of 100 km) – local radio stations are responsible for the broadcasting
- Infrared beacons on major roads that typically have a range of 3.5 m and transmit information about the traffic situation within a surrounding radius of approx. 30 km
- Microwave beacons are installed on expressways and have a range of approx. 70 m; they distribute information about the traffic situation on divided highways predictively up to 200 km

7.7 Future Development of Telematics Services

Telematics services in cars are generally characterized by a long value chain (content provider, service provider, network provider, terminal device manufacturer) that has to be governed in order to be able to offer high-quality services. Furthermore, the long value chain with its many participants all wanting to make a profit brings with it the risk of end user prices that are too high. The value chain participants come from the automotive and CE worlds and have different interests and business models that all have to be integrated. A major commitment is therefore generally required from a carmaker in order to develop and offer a high-quality "car services portal."

The users, however, only show a limited willingness to spend money on in-vehicle services. An important function on which users are willing to spend money is mobile communications: telephoning using a hands-free system. Though this function is generally not sufficient to cover the costs of a telematics module. There is also only a low willingness to spend money even on additional services that offer a direct customer benefit, like "emergency call" and "off-board navigation." Services like "remote diagnosis" do not create any immediate customer benefit but an indirect benefit, or they open up opportunities for "customer relationship management."

Technical limitations too diminish the speed at which telematics-based devices and functions are introduced to the vehicle environment. Low data transfer speeds and the long connection setup times for GSM place a limit on the device's responsiveness. Cellular signal losses and problems with cellular coverage can lead to an impairment of functions and thus to unsatisfied end users. “Remote control functions” in a connected vehicle environment are critical to safety, and a satisfactory solution has not yet been found in this regard. The technologies in the world of CE will not be optimized to suit the needs of automotive technology. Carmakers will therefore continue to experience difficulties in making CE technologies fully usable in the vehicle environment (see Sect. 9.2).

Standardized interfaces will help with the integration of telematics functions in vehicles, but they at the same time pose other problems (e.g., easy/uncontrolled access for products from competitors). Telematics functions will not be replacing today’s “on-board functions” but will supplement them. An essential element for OEMs – in conjunction with telematics – will be customer relationship management. More recent technologies (LTE, Java, etc.) will make it easier to implement telematics functions. Services offering a high degree of customer benefit, while taking into consideration convenience, safety and costs, are necessary for business success. Telematics will presumably not establish itself as a value-added service paid for by the customer but will become established through the continued development of existing distribution media (e.g., SDARS) and enhanced navigation technology (hybrid navigation). Realistic time frames must be set for the penetration of new technologies and telematics services.

8 Smartphone Integration in Cars

Smartphone integration in automobiles is understood to mean the connection of a smartphone by cable or wirelessly to the car’s infotainment system. This system enables users to interact with their smartphone by, for instance, displaying the smartphone’s screen on the vehicle’s display and letting users operate it with control buttons such as those on the head unit or steering wheel remote control.

8.1 Motivation for Smartphone Integration in Cars

The forecast for global smartphone sales indicates a distinct upward trend: while around 300 million smartphones were sold in 2010, sales are expected to increase to around 1.4 billion smartphones in 2016. This growth is linked to a sharply increasing number of applications that users can run on their smartphones whenever and wherever they want. According to Paragraph 23 of the German road traffic regulations, however, drivers are not permitted to operate smartphones while driving: in Germany, a mobile phone may not be used while driving if it has to be picked up or held in order to do so. This therefore also means that use of a mobile phone as a navigation aid while driving is restricted – though use of a smartphone car mount or operation through voice control provides a possible alternative means of operation.

Solutions therefore need to be sought to integrate smartphones in automobiles so that users can access them while driving without violating applicable legislation or compromising safety during the journey (Visveswaran 2012).

8.2 Smartphone Integration Options

8.2.1 Dock-Based Integration

Dock-based smartphone integration is among the easiest options. A distinction is made between two different approaches (Visveswaran 2012).

8.3 Semi-Integrated Approach

The dock and head unit have been designed to be able to function independently of one another. Basic smartphone integration functionality is utilized that enables access to the smartphone's audio channel so audio can be outputted on the loudspeakers integrated in the vehicle.

8.4 Fully Integrated Approach

In the fully integrated approach, the head unit is dependent on the smartphone for the full range of functions. The main user interface is provided by the smartphone app for controlling various hardware components, such as a radio.

8.4.1 Market Solutions for Dock-Based Integration

The concept of dock-based integration went hand in hand with the continuously growing popularity of the Apple iPhone$^{\mathrm{TM}}$. The majority of the currently available docking solutions have therefore been designed solely for the iPhone. Due to the increasing popularity of smartphones running the Android operating system from Google, the manufacturers are being confronted with new challenges to develop universal docking solutions that are compatible with smartphones from various manufacturers running various operating systems.

8.4.2 Advantages and Disadvantages of Dock-Based Integration

The advantages of dock-based integration are in particular the low cost of the docks and the potential to adapt quickly to the market of new smartphones. Before docking solutions can be deployed, however, the necessary apps must first be developed, since this ensures optimal utilization and compatibility of the smartphones.

A big disadvantage for docking solutions is the legal requirements of some countries that to some extent limit or even completely prohibit their use. In addition to this, the displays that are increasingly being installed in new vehicles as standard are a threat to docking solutions, which could in the long term become redundant.

Fully integrated solutions have in particular the disadvantage that they must maintain compatibility in the long term with every new smartphone generation.

8.4.3 Proxy Solutions

The term “proxy” refers to the part of the head-unit software that is capable of communicating with the smartphone’s installed apps, which make information available in a format compatible with the head unit. Most smartphone integration solutions currently on the market utilize this proxy solution, which allows the smartphone apps to be run directly on the head unit and displayed on the screen of the infotainment system.

8.4.4 Manufacturer-Specific App Approach

The aim of this approach is to create a carmaker-specific smartphone application that permits communication with the car manufacturer’s head unit. The backdrop to this approach is that third-party applications do not have direct access to the infotainment system, since the car manufacturers only provide select contractual partners with full access.

A distinction can be made between three variations of this manufacturer-specific app implementation:

(a) Implementation as a “meta app”: The head unit communicates with a “meta app” installed on the smartphone, which in turn controls sub-apps (like internet radio) that are embedded in the main app.
(b) Implementation as a “gateway app”: The head unit in this case interacts with a “gateway app” that communicates with other compatible apps that are hierarchically equal and thus independent.
(c) A combination of both of the above variants: The head unit in this case exchanges data with a “meta app,” which in turn can control both subordinate and hierarchically equal compatible apps.

An advantage of proxy methods is the seamless implementation that enables data to be exchanged between the car and smartphone – this makes it possible, for instance, to read out vehicle data. Using the APIs offered by some carmakers, third-party manufacturers can develop compatible apps themselves.

Among the disadvantages is that there is so far no generalized solution available for integrating an app in vehicles from different carmakers. A modification of the app is therefore often required to provide the necessary compatibility; furthermore, in many cases the app must also be installed on both the smartphone and the car head unit in order to display the user interface. A further issue is that the user interface is limited to the predetermined functionality and that an expansion of the API necessitates new apps or software updates in the car. The maintenance requirements in such cases are rather high while flexibility is low.

Despite the disadvantages, car manufacturers are currently tending to choose proxy solutions over integrated solutions since they offer greater potential for implementing future developments.

8.4.5 The Future of Smartphone Integration in Cars

The future and further development of smartphone integration depend, among other things, on which technologies are developed to connect cars to the cellular network. Three technologies can be distinguished in this regard (GSMA-mAutomotive 2012):

Embedded solutions: The connection to the cellular network as well as all provided functionalities are realized by systems integrated in the car.

Tethering solutions: To ensure functionalities that are dependent on a cellular network can be utilized, it is necessary to provide a connection to a mobile phone, which is used as a modem. The connection types available are modem and hotspot/access-point solutions via Bluetooth and Wi-Fi.

Integrated solutions: Smartphone functionalities – particularly apps – are integrated in the car.

None of these solutions should be seen as an exclusive solution. Most carmakers are developing strategies in which several of these connectivity options can be utilized for different market segments (e.g., embedded solutions for compact class models and tethering solutions for sub-compact class models). Furthermore, different technologies will be used depending on the application. Embedded solutions are preferred for safety-related functions, while integrated solutions will be used for infotainment functionality (GSMA-mAutomotive 2012).

9 Aspects of Mobile Communications for Navigation and Telematics

Compared with other electronic control units in cars, radio/navigation and telematics systems are subject to very different constraints that have a decisive influence on their development. These constraints are examined below.

The end user has a very clear awareness of the function of a radio/navigation system. In contrast to a brake control unit or engine control unit, whose functionality is similarly complex but which are deployed largely unnoticed by the driver, a radio/navigation or telematics system possesses a complex interface with the driver. The driver perceives the functionality via this interface and experiences many of the devices features directly. Slow start-up behavior or a sluggish HMI, even with only a slightly delayed response to operating actions, will therefore become immediately apparent and make a negative impression.

Furthermore, due to its strong presence in the area of the car’s center console and its control elements, a radio/navigation system constitutes a design-relevant component. It is not uncommon for the design requirements and the need for straightforward and safe operation to conflict with one another. For example: design requirements for chromed smooth and shiny controls, but functional requirements for non-slip, safe-to-operate surfaces.

Last but not least, developments in consumer electronics (CE) exert a considerable influence. On the one hand, functions offered by consumer electronics devices are also expected in vehicles in a radio/navigation system. This leads to components from consumer electronics having to be adopted in vehicles. Consumer electronics devices are often produced in considerably higher numbers than devices installed in vehicles. This leads to the pressure to adopt CE components that, for reasons of cost, have undergone no or only slight modification, even though they do not fully satisfy the requirements of the vehicle environment. On the other hand, consumer electronics devices are in direct competition with devices installed in vehicles: a current example is the portable navigation device. Since the consumer electronics sector has shorter development times and alternative distribution channels, there is considerable pressure to innovate, resulting in significant innovation from generation to generation, and cost pressure is very high. An average price reduction of more than 10 % per annum with a simultaneous increase in the range of functions is a customary requirement.

9.1 Consumer Electronics (CE) Versus Automotive Electronics (AE)

Particularly the utilization of components commonly used in consumer electronics poses a variety of challenges for the development and automobile-compatible certification of navigation systems. The pressure to use such components firstly arises from the fact that consumer electronics devices offer functions that drivers also expect to see in vehicles. An example of this is the ability to play music from audio or data storage devices that are used in the home, such as, music on CD, or MP3 files stored on CD or SD card. Secondly, due to the very large-scale manufacturing employed, the consumer electronics sector can supply components in a price category that would not be achievable with components specially made for automotive use only. Examples are: CD drives for portable devices, home devices and PCs; and hard disk drives for PCs and video recorders.

The requirements and resulting challenges are illustrated below using the example of a DVD drive (see Table 3). DVD drives have been utilized in navigation systems as mass storage devices for digital maps due to their capacity of around 7 GB. Currently they are being superseded by electronic media like SD cards. In high-end navigation systems, the drive is additionally used for playing video DVDs. Using this component as an example, the table illustrates the requirements of the consumer electronics sector (home and PC; CE requirement) alongside the requirements of the automotive sector (AE requirement).

9.2 Design of the Navigation System

The design of a navigation or telematics system depends on the planned installation type. Systems therefore also exist as a functional component without a dedicated

Table 3 Demands placed by the CE and AE environments on a component (using a DVD drive as an example)

Parameters	CE requirement	AE requirement	Practical compromise solution
Ambient temperature	0–60 °C	−40 °C to +95 °C	Operating temperature of −20 °C to +80 °C with functional limitation outside this range (drive deactivation). Challenges: drive lubrication, vibration-absorbing elements, distortion of the plastic lens. These elements are, if necessary, modified for the vehicle
Media temperature (CD, DVD)	55 °C[a]	95 °C	None, since it is not possible to specify to end users which particular media they must use. If need be, a warning notice can be included in the operating instructions explaining that unsuitable data storage media can, in extreme cases, ruin the unit
Installation angle	Around 0°	−30° to 90°	Through modified drive suspension and vibration clearance it is possible to achieve an operative range of −15° to +45°. For a more extensive range, drive variants with different mechanical properties must be installed (resulting in different device variants in vehicles)
CD/DVD loading time	No specifications; generally non-critical	Max. 3–6 s until function available (i.e., audio signal can be heard after inserting a CD)	Cannot currently be resolved; usual times are 7–15 s before function availability after disc insertion
Full stroke seek (time required by the read head to traverse the entire disc)	Non-critical, since a read head is generally not repositioned that often due to large data blocks being stored linearly one after another; generally: 800 ms	As short as possible (if possible <150 ms)	Two-dimensional navigation map data requires that the head be repositioned numerous times in order to load navigation data Great effort is put into storing data as optimally as possible on a linear data spiral on the disc so that head repositioning is kept to a minimum
Streaming rate	High, in order to load large amounts of data quickly	Low, since processor performance is lower than for PCs	For navigation systems, head-repositioning time is more important than streaming rate
Period of product availability	2–3 years	15 years	Product maintenance after start of production so new drive variants can be installed

(*continued*)

Table 3 (continued)

Parameters	CE requirement	AE requirement	Practical compromise solution
Read-head position control	Slow	Very fast	Challenge: deviation control of the read head during severe vibrations in the vehicle
Dirt	No special requirements	Operation under warm, moist conditions and after simulated dust exposure	Challenge: simulation of convection heat currents to predict dust accumulation characteristics; if necessary, enclosure of the drive

[a] Temperature specification of a well-known brand manufacturer for the operative range of its "burnable" data disc blanks

operating interface, that is to say, they are utilized as a component in a larger integrated system. In this case, they are a so-called "silver box" with a connectivity interface (e.g., electrical CAN interface or optical MOST interface). This type of design is common for pure telematics devices without additional functionality. A fundamentally different type of device is one with a dedicated operating interface ("silver box" with a plastic front panel). This type of design is generally used for radio/navigation units and head units that provide several functions (i.e., radio, navigation, and audio playback from audio media and data storage devices). A typical radio/navigation system for vehicle installation in the entry-level segment comprises approx. 1,500 components (mechanical and electronic).

9.3 Development Process

The development process for navigation or telematics devices is characterized by a high level of complexity and a great many requirements. Pure navigation or telematics systems in the form of telematics or navigation modules that are connected up to head units in order to be controlled are rare. The majority of navigation systems on the market are infotainment systems that also include a radio tuner and media functions, such as playback capabilities for audio/MP3 CDs or SD cards and USB flash drives.

Due to the influence of the rapidly changing CE world, even entry-level radio/ navigation systems (RNS) are subject to a considerable degree of innovation from one device generation to the next. Even entry-level models meanwhile feature high-resolution TFT color graphics displays; such display resources require the support of powerful graphics processors so 3D navigation maps and graphic animations for menu changes can be displayed smoothly. Although CD and DVD drives are still present in many devices for playing conventional data storage media, they are already to some extent being replaced by electronic devices, like SD cards and USB flash drives. The extensive data storage options are stimulating demand for broadband interfaces that can supply data quickly

(USB, WLAN). Bluetooth-capable mobile phones encourage the utilization of Bluetooth technology, at least as an option, even at the entry level. Smartphones must be connected using various standards to enable their operation in the vehicle and to integrate data and services from them. New reception methods, such as antenna diversity and the background TMC tuner (for continuous reception of RDS-TMC messages independent of the "foreground tuner" being used for radio listening) have already largely become standard through the use of multi-tuner systems.

The high level of innovation in each new device generation only permits limited reuse of already developed hardware and software components. In the area of software development, the trend is increasingly moving toward use of open source software (OSS) in order to limit the investment in software development and to be able to offer new functions sooner.

Added to this is the wide variety of functional requirements. At the start of development, there are generally up to several hundred documents that define the functional specifications, representing thousands of detailed requirements. A list of functions for a radio/navigation system usually comprises 2,000–4,000 elements, and behind each individual function there are anywhere between several and numerous very specific functions and different detailed requirements. The graphical user interface comprises 500–2,000 different screens whose design is specified by the vehicle manufacturer. The high number of detailed requirements, which make the task of drawing up a consistent requirements specification difficult, as well as the changes to the functional specifications during development lead to extensive modifications being implemented during the development process.

The applied development process must therefore satisfy the following requirements:

1. Management and configuration of an extensive, constantly changing quantity of documents (requirements specifications)
2. Identification and handling of inconsistent requirement specifications
3. Handling of extensive modifications during the development phase
4. High level of innovation in the requirements (only partial availability of past experience)
5. Flexibility in order to take into consideration the development process requirements of the customer (the OEMs pursue distinct development models and define very different development specifications)

These constraints lead to considerable challenges in project planning and cost estimation, particularly at the beginning of the project, since the task of processing and resolving the device specifications in detail can itself result in several months of work.

To handle such projects, database systems are used during the development process to manage the customer's requirements – they support the process from requirement specification evaluation to the development phase to product testing. Only by using such a system is it possible to ensure all aspects are considered throughout the entire development process. Approaches to hardware and software design and standardization in the automotive industry will become increasingly established, which, when implemented consistently, are expected to facilitate the reusability and interchangeability of components (examples of this are AUTOSAR 2015, VICS 2015, Zimmermann and Schmidgall 2007 and the GENIVI Alliance 2015).

References

ACTMAP Project (2015) http://www.transport-research.info/web/projects/project_details.cfm?id=14953. Accessed 18 Mar 2015

ADASIS Forum (2015) http://adasis.ertico.com/. Accessed 18 Mar 2015

AGORA Project (2015) http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=45962. Accessed 18 Mar 2015

Daimler Chrysler (2003) Systembeschreibung ETC Deutschland (ETC Germany system specification)

Dijkstra EW (1959) A note on two problems in connexion with graphs. In: Numerische Mathematik. 1, pp 269–271. Available at http://www-m3.ma.tum.de/foswiki/pub/MN0506/WebHome/dijkstra.pdf

Eberhardt R (2003) Car to car communication consortium. EuCar SGA, 23 Oct

FEEDMAP Project (2015) http://www.mapchannels.com/FeedMaps.aspx. Accessed 18 Mar 2015

Förster HJ (2001) Das Automobil, ein Lebenselixier für alte Menschen (The automobile, an elixir of life for the elderly), VDI Report 1613. VDI-Verlag, Conference Berlin, May 2001

GENIVI Alliance (2015) https://www.genivi.org/. Accessed 18 Mar 2015

GSMA-mAutomotive (2012) Connecting cars: the technology roadmap. GSMA

Hendriks T, Wevers K, Pfeiffer H, Hessling M (2005) AGORA-C specification

Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82(Series D):35–45

Krüger K et al (2001) Optimierung der Kompetenz älterer Fahrerinnen und Fahrer durch frühzeitige Navigationshinweise und Knotenpunktsinformationen (Optimization of the competence of older drivers through navigation instructions and intersection information provided in good time), VDI Report 1613. VDI-Verlag, Conference Berlin, May 2001

Neukirchner EP (1991) Fahrerinformations- und Navigationssystem (Driver information and navigation system). Informatik-Spektrum 14(2):65–68

Nöcker G, Mezger K, Kerner B (2005) Vorausschauende Fahrerassistenzsysteme (Predictive driver assistance systems), FAS 2005, 3. Workshop Fahrerassistenzsysteme (3rd driver assistance systems workshop), Watling, 6–8 Apr 2005, pp 151–163

Otto HU (2005) The ActMAP approach – specifications of incremental map updates for advanced in-vehicle applications. In: Proceedings of “ITS in Europe Congress”, Hanover

Pilsak O (1999) Routensuche, digitale Karte und Zielführung. (Route search, digital map, and route guidance.) In: Talk held at seminar: Kfz-Navigation Überblick über Entwicklung und Funktion (Motor vehicle navigation – overview of development and function)

Salas G (1968) Highway coding for route destination and position coding. Highway Research Board, Publ. No. 1642

VICS (Vehicle Information and Communication System) (2015) http://www.vics.or.jp. Accessed 18 Mar 2015

Visveswaran A (2012) A status update on in-car smartphone integration. SBD

Zimmermann W, Schmidgall R (2007) Bussysteme in der Fahrzeugtechnik – Protokolle und Standards (Bus systems in automotive engineering – protocols and standards), 2nd edn. Vieweg-Verlag, Wiesbaden. ISBN 978-3-8348-0235-4

Additional References

Autosar (2015) http://www.autosar.org. Accessed 18 Mar 2015

IDC (2013) Prognose zum Absatz von Smartphones weltweit bis 2017 (Forecast global smartphone sales up to 2017), weblink: http://de.statista.com/statistik/daten/studie/12865/umfrage/prognose-zum-absatz-von-smartphones-weltweit

Part X

Future of DAS

Future Integration Concepts for ADAS

55

Peter E. Rieth and Thomas Raste

Contents

Abstract

The automotive industry is facing the next major evolutionary step. New functions for highly automated driving are entering the vehicles. This is accompanied by increased E/E and mechatronic contents, leading to increased topological complexity. At the same time the system/component and development costs should remain stable, and product quality should be further improved.

A promising strategy to master the complexity of the E/E architecture is the clustering of already intensively networked functional elements either by physical integration or by functional integration into a handful of functional domains. One of these functional domains is the motion domain, which is needed to execute the driving strategy. In recent times there has been a trend toward

---

P.E. Rieth (✉) • T. Raste
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: peter.rieth@continental-corporation.com; thomas.raste@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_56

automation of selected elements of the driving strategy, like driving with a predefined speed or distance in a specific lane.

The main purpose of motion control is to execute the driving strategy by generating and managing the forces at the wheels. Motion control structures and coordinates the access to the actuators. The command flow is hierarchically organized in a three-layer sequence. The standardization of the interfaces of each layer is an important task, which finally has to lead to an extension of the AUTOSAR application interface catalog. A new and challenging requirement for motion control is to provide a tracking control capability to follow a predefined trajectory autonomously.

For a custom-specific realization of motion control, powerful and flexible integration platforms equipped with multi-core microcontrollers are available. The new architecture is highly scalable and fulfills all requirements from ISO26262 ASIL D. The design is fail operational due to additional on-chip diagnosis (1oo2D), i.e., in case of a permanent failure in one channel, a limp-home mode is entered. The AUTOSAR compliant software can be configured to satisfy different customer needs and requirements, e.g., most flexible hardware resource usage or maximum independency between OEM and supplier software (virtual ECU).

1 Introduction

The automotive industry is once again poised for another evolutionary leap. Cars will soon benefit from functions that will make highly automated driving possible. An increasing amount of mechanical and electrical/electronic components will team up with software that is expanding in geometric fashion. All of this means that the complexity of E/E architecture will skyrocket. At the same time, the costs of the system, its components, and their development must not rise while their quality has to improve steadily.

Managing this complexity requires new solutions to architectural concepts. The variety and thus the changes in requirements need to remain manageable. Platform strategies and modular design are today’s answer to these challenges. The domain approach (Reichart et al. 2007) promises still more improvement. This provides for a regrouping of functions and electronics into just a few (four to five) domains so that, as much as possible, changes will affect only one domain without spilling over into other domains.

The growing power of modern multi-core processors has fed the expectation that it will be possible to integrate more and more control functions within a single domain into domain control devices. This will render it necessary to master many new challenges such as scheduling with abbreviated latency so as not to put the stability of closed loops at risk. It will also be necessary to devise safety mechanisms so as to achieve a sufficient measure of isolation between software components in spite of a highly integrated hardware basis (Hilbrich and Gerlach 2011).

The choice of architecture is a function of the features required and the extent to which a car is to be equipped with them. It is also necessary to keep an eye on additional cost centers such as developing and testing functions, modifying specifications for specific vehicles, and managing functions and components (Weidner et al. 2013). Carmakers must decide carefully whether an additional function will result in a new component, or whether several basic functions can be integrated into a component along with the expanded function. Here, it will be necessary to examine at which price integration of additional components would be more favorable for a given quantity. Of course, this does not only apply to control devices but for mechanical/electrical systems in general. An example of successful integration of a mechanical/electronic braking system will appear in the following chapter. We will then consider the functional integration and major aspects of domain architecture in closer detail.

2 Physical Integration

The many new automotive trends affecting braking systems have led to a situation where fulfilling additional requirements by expanding the architecture of the ordinary basic system has reached its technical limits. This has provided motivation for a consistently new approach that increasingly will be based on commercial considerations (Feigel 2012a). The architecture of today's automotive braking systems is the product of historical evolution (▶ Chaps. 29, "Hydraulic Brake Systems for Passenger Vehicles," and ▶ 39, "Brake-Based Assistance Functions"). Brakes originally had no boosters. The vacuum brake booster was the next evolutionary step. It derived its boosting power from the vacuum that arose in classical, normally aspirated engines. Hydraulic brake-pressure modulation was a later development which made braking more easily controllable (anti-blocking system, ABS). Then came the electronic stability control (ESC) which enhanced stability through electronically controlled braking. In the meantime, a third component has joined these systems. In many cars now, a pump supplies the vacuum needed by the brake booster because modern engines are no longer throttled in the interest of efficiency. Therefore, a brake system today consists of two, and often three, separate components (Feigel 2012b). Up until now, this architecture served a purpose. Now, however, the situation has changed dramatically with the advent of hybrid and electric vehicles on the one hand and increasing demands from emergency-brake systems plus increasing NVH demands from comfort-assistance systems on the other hand. The following will serve to illustrate this point.

An example of two emerging trends is Euro NCAP active pedestrian protection, which will be mandated starting in 2016, and the growing demand for the traffic-jam assistant function. Fulfilling both functions will lead to a conflict, based on current ESC systems. On the one hand, as much output as possible from the recirculation pump is necessary in order to achieve the emergency-braking characteristics necessary for protecting pedestrians. On the other hand though, as little pump pulsation as possible is needed to satisfy the requirements of follow-to-stop comfort. Achieving this with high-output pumps is only possible at great additional cost.

The efficiency of gasoline engines has increased, but it has entailed throttle losses along the intake tract which are no longer acceptable. As a result, a special vacuum pump has become necessary for gasoline engines as is already the case for diesel engines. If, moreover, the engine sailing mode is engaged in the interest of greater efficiency, thus disengaging the mechanical vacuum pump that the combustion engine drives, a vacuum pump driven by an electric motor often will take its place. This again serves to increase costs, weight, space requirements, complexity, and the risk that a pump may fail.

Since the energy density of the vacuum (limited by the ambient pressure) is very low anyway, and the packaging density in cars is constantly increasing, it makes sense to search for alternative types of energy. Since the availability of electrical energy in cars is on the upswing, an obvious alternative is to support brake boosting electrically, as is already the case with power steering. Likewise, instead of using vacuum as a source of energy, cars can now make use of the electric motor of the ESC unit.

If it is then possible (as the basic idea of the MK C1 provides (▶ Chap. 29, "Hydraulic Brake Systems for Passenger Vehicles")) to combine the functions of electric brake boosting and stability control into one unit, it would not only greatly reduce the cost of assembly and the total weight but also enhance the quality and robust nature of the functions. It would obviate the need of not only installing a vacuum pump and ESC system but also their brackets, their electric lines, and their fluid lines.

Figure 1 shows the functional components of the MK C1. External characteristics like the pedal stem and reservoir that at first glance signal a typical stability control system are clearly recognizable as an integrated brake-activation system. A driver activates the pedal stem, which is connected to a tandem master cylinder built into the valve block. Under normal conditions, the master cylinder supplies the pedal simulator with pressure. Should the electrical system fail, the master cylinder

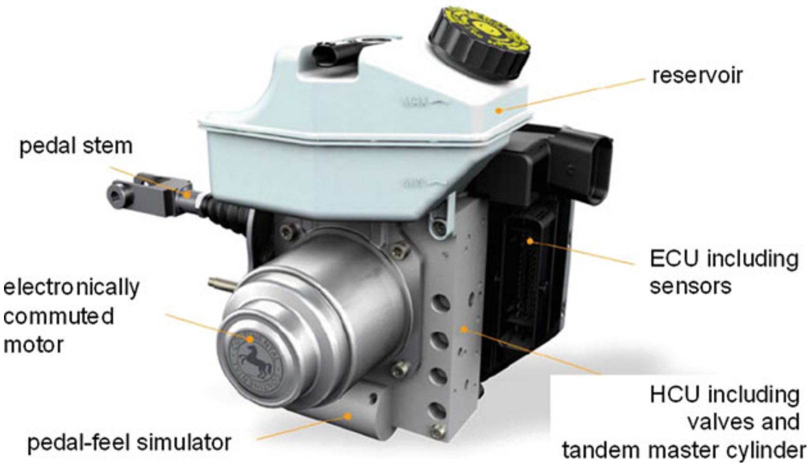


Fig. 1 MK C1 integrated brake system

will also activate all four-wheel brakes directly, i.e., without boost, creating a hydraulic fallback level.

This brake-by-wire system permits normal braking with the aid of a plunger, which is driven by an electrically commuted DC motor. The plunger pressurizes and depressurizes in accordance with a driver's wishes. Sensors monitoring pedal travel and pressure assess a driver's intention. All sensors are integrated into the system and connected to the ECU. The valve block houses both the valves necessary for normal activation and the control valves for antiskid functions.

Another advantage of this very compact arrangement is the very short length of the entire assembly. The outer dimensions of the MK C1 fit almost completely into the hollow of a classical brake booster (an 8″/9″ device). This is important so as to not run afoul of packaging restrictions in modern engine compartments, even where both types of devices are present. The fact that in some cases the MK C1 is shorter than a tandem master cylinder by approximately one length could prove particularly advantageous in the case of a collision. The distance to other assemblies is sufficiently large that it could prevent or mitigate intrusion of the pedal assembly into the passenger compartment. There are likewise no particular features on the side of the MK C1 facing the passenger compartment that might influence the usual interface with the pedal assembly.

3 Functional Integration

The demand for more safety, comfort, and, at the same time, lower fuel consumption requires innovative system solutions in the realm of chassis and drive train. Development up until now has not been continual but has been characterized by several leaps in technology. The first leap that customers could register and appreciate as progress was the introduction of such electronic/mechanical systems as ABS in the late 1970s or ESC in the mid-1990s. Since optimization of electronic/mechanical systems increasingly bumped up against its limits, the next leap occurred in the first decade of the twenty-first century, namely, the functional integration of systems that, as already described, was rooted in the strong growth of driver-assistance systems and systems designed to improve fuel economy. Given the existence of high-performance bus systems connected to brakes, steering, drive, and shock absorbers, it was now possible to get a better grip on the conflicts posed by active safety, driving enjoyment, comfort, and efficiency. Today there are essentially two approaches to functional integration:

1. *The fully integrated approach*: A central control device directs the intended driving dynamics and coordinates the necessary actuators such as in Smakmann et al. (2008). This approach offers the potential for high performance but frequently leads to high integration costs for carmakers.
2. *A cooperative coexistence approach*: A central coordinator activates predefined operating modes or parameters such as in Held (2009). This approach does not

lead to maximum performance; however, the cost of integration is reasonable due to extensive separation of the work of carmakers and automotive suppliers.

A development known as global chassis control (GCC) aims at functional integration for the purpose of regulating horizontal dynamics. Chassis subsystems – divided according to components – were initially treated by development departments of carmakers and suppliers as stand-alone systems for historical reasons and because of purchasing strategy. As a result, it was possible to have a car with active steering, active stabilizers, electronic stability control (ESC), and electric differential with up to four independent horizontal dynamic controls, each with its own sensors, its own method of calculating reference values, and its own dynamic control. In this type of functional architecture, known as the "coexistence approach," the individual controls pursue different objectives (comfort, handling, and safety) independent of the system they are supposed to regulate. Their action area must be coordinated in such a manner that they cannot negatively influence each other. In effect, they need to be isolated from each other. The result is that it is not possible to achieve optimum overall control. ESP II represents the first time that the distribution of functions known as the "integrated approach" has been achieved (Schwarz et al. 2003). With this approach, each of the individual systems – steering, brakes, chassis, and drive train – possesses a basic function. With regard to horizontal dynamics, this basic function is limited to pure feedforward control such as speed-dependent steering ratio or right/left brake-force distribution as a function of lateral acceleration. The functions constantly exchange information with the overall horizontal dynamic controller in the ESP II and report their current reserve status.

The functional integration of sensors represented the next evolutionary leap (Kelling and Raste 2011). Longitudinal-guidance functions such as adaptive cruise control (ACC) (▶ Chap. 45, "Adaptive Cruise Control") and emergency brake assist (▶ Chap. 46, "Fundamentals of Collision Protection Systems") or lateral-guidance functions such as tracking or lane-departure warning or warnings of following traffic are already in use (▶ Chaps. 48, "Lateral Guidance Assistance," and ▶ 49, "Lane Change Assistance"). New functions such as construction-site assistance or emergency evasive maneuvers are in the development stage. All of these functions are based on information obtained from sensors such as RADAR (▶ Chap. 17, "Automotive RADAR"), camera (▶ Chap. 19, "Automotive Camera (Hardware)"), LIDAR (▶ Chap. 18, "Automotive LIDAR"), and ultrasound (▶ Chap. 16, "Ultrasonic Sensors for a K44DAS"), all of whose information melds into a model of the surroundings (▶ Chap. 24, "Representation of Fused Environment Data"). Any safety information that may become available in the future via vehicle-to-vehicle or vehicle-to-infrastructure communication (V2X) (▶ Chap. 27, "Vehicle-2-X") will also register as sensory input. Ultraprecise information as to the car's position will be necessary in order to achieve this, though. Positioning using a digital map such as those used by navigation systems will not be possible here because map data could be faulty or even entirely unavailable. Approaches that couple information from the driving-dynamics sensors to the GPS signal in a car have shown promise.

They are capable of precise positioning and are highly available (▶ Chap. 25, "Data Fusion for Precise Localization").

When it comes to passive safety, meaning any functions that mitigate the consequences of an accident, cars today contain a number of functions that are undergoing constant optimization and refinement. Among them are crash sensors (frontal, rear, side, rollover) and the activation of appropriate restraints (airbags, belt tighteners, headrests). It is also possible to trigger an automatic emergency call (eCall) and to trigger the brakes to diminish the effects of a secondary crash. Of similar importance are ways of protecting pedestrians. For example, in case of a frontal collision with a pedestrian, the hood will draw up to diminish injuries. It is possible to improve the passive safety functions, which are based primarily on crash sensors, by linking information from the model of the surroundings. This represents a type of "prospective" protection against accidents which can predict crashes or recognize pedestrians.

4 Domain Architecture

Once the functional network has been established, it will be possible to work out the architecture of the system. One major challenge will be partitioning the functions onto the control devices, a process that, although it has already been mathematically solved (Lochau et al. 2009), at least approximately, still requires a good deal of experience today. It should be noted that strong interactions among the system's components will also lead to strong interactions in developing and manufacturing the system on the one hand and the structure of suppliers and collaboration on the other. Changes to one component are hardly possible without having an impact on all the other components. This is where the modular concept comes in. The draft system and the steps necessary to achieve it can be broken down into modules which can then be assigned to either of the two following categories with the aid of rules and standards:

1. Externally visible modules such as operating systems and things like middleware or design parameters such as specifications for interfaces or integration and test procedures
2. Externally invisible modules, i.e., components whose design parameters remain hidden in each module (and their own development departments)

4.1 Approaches to Standardizing Architecture

Successful modularization is based on detailed knowledge of the interactions and interdependencies among the design parameters. A system architect will work out the following rules and standards:

- *Architecture*, i.e., determining what modules and components are part of the system, what roles they play, and what modules and components are sources for externally visible standards
- *Interface*, i.e., a detailed description of how the modules are joined, communicate, exchange energy, etc. In other words, the interface as a component of the overall system
- *Integration and test*, i.e., instructions for assembling the system and for determining how well the system works and how well one version of a module works relative to another

Modularization is the first step to enhance value for the system. Investing in modularization will only pay off, however, if the hidden modules/components undergo permanent refinement or are exchanged for better ones after the module has been taken apart. The number of experiments necessary to accomplish this will depend on the size and the technical potential of the modules/components. The basic principle is that large modules/components require fewer development projects than small ones because the costs per experiment are high in comparison to the small modules (Baldwin and Clark 2000). The evolution of visible modules requires high outlays and a good deal of coordination because of their wide-ranging effects on the system, and therefore, these modules are comparatively constant.

Figure 2 clearly points up the advantage of modular design by making use of fictional system architecture with four components (A, B, C, D). In the directional graphs in Fig. 2 above, arrows indicate how the components interact. The design structure matrix (DSM) is the equivalent of the graph c.f. (Eppinger and Browning 2012). The DSM depicts relationships between the elements in a system in a compact and visually advantageous format. A DSM is a square matrix in which the elements under examination appear on the diagonal. If a relationship exists between two elements, an “x” appears in the matrix. One can alternatively use numeric values to document the magnitude of the relationship, for example. The matrix runs in a predetermined direction, thus providing directional information. The applicable convention here is that the input is in the columns, while the feedback reaction appears under the diagonal.

In the example of a highly networked system, the interfaces are proprietary, i.e., there are no open, industry-wide rules for linking the components. The upshot is that any change to any one component has an effect on all the other components. On the other hand, any relationships among components in a modular system are ideally determined by open, industry-wide rules and standards (design rules, DR) plus integration and test procedures (I&T). Even though the components may remain hidden, development work can proceed completely independently as long as open standards are adhered to. The matrix makes this apparent wherever no “x” appears in the marked area. Thus, changes to any hidden, outwardly invisible design parameters have no effect on the system’s other hidden modules. Only changes to the visible design parameters will require changes to the hidden modules.

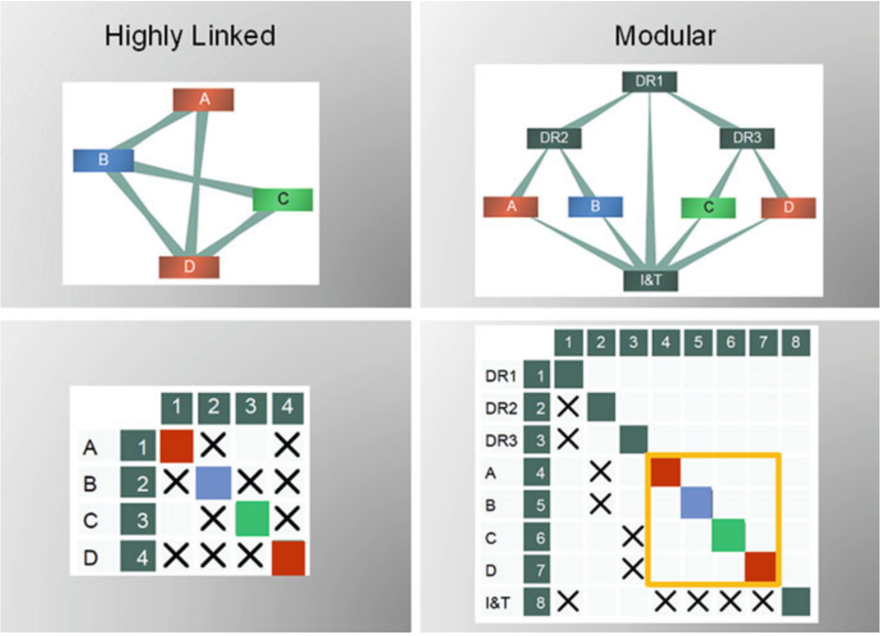


Fig. 2 Depiction of a highly linked (*left*) and modular system (*right*) as a directional graph and as a design structure matrix (DSM). *A to D* system components, *DR* design rules, *I&T* integration and test

4.2 Approaches to Standardizing Interfaces

The next step is to determine interfaces whereby the following two objectives for domain architecture stand in the forefront (Eriksson and Alminger 2013):

- Reduce complexity by exchanging as little information as possible between domains
- Facilitate reusability by employing coordinated interfaces that will thus remain stable over the long term

A major building block of system architecture is what is known as AUTOSAR (basic software, BSW) (▶ Chap. 7, “AUTOSAR and Driver Assistance Systems”). It makes it possible to develop software applications (software components, SWC) independently of hardware (Weidner et al. 2013). One component of basic software is middleware, which links the components of the application software with each other and with platform services such as communication or system services.

According to today’s standard, the middleware and the interfaces to the application software are generated statically in what is known as a runtime environment, RTE. Future generations of control devices could see dynamic integration of the

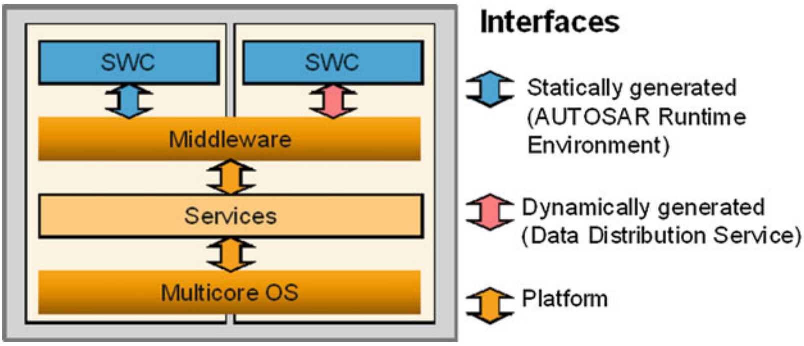


Fig. 3 Control device architecture with application software (software components, SWC) and platform software (basic software, BSW) with today's static interfaces as well as the dynamic interfaces of tomorrow

application software (see Fig. 3). The motivation behind this is that carmakers intend to install software on control devices via Internet connections in the future so that they can constantly offer their customers up-to-date functionalities (Bulwahn et al. 2013). For this reason, the basic software must contain middleware that offers Data Distribution Service (DDS). Open standards for implementing DDS already exist such as the Robot Operating System (ROS) (ROS 2015) and the Open DDS (OpenDDS 2014). These solutions, however, are not yet standard in the automotive industry.

4.3 Approaches to Standardizing Integration

The most important innovations in the future of domain architecture will be domain control devices, which will serve as integration platforms for carmakers' software and thus afford a unique selling proposition. Additional drivers of domain architects will be effective handling of variants such as repartitioning and upward integration, the physical encapsulation of domains and protection of know-how.

The AUTOSTAR standard has introduced a new method for integrating basic and application software. Instead of recoding the assignment of software to hardware for each control device, predefined modules need only be configured. Appropriate tools support configuration and generate XML descriptive files, which carmakers and suppliers exchange among themselves. When development commences, a carmaker produces a description of such things as the topology, communication details, and the partitioning of the application software onto the different control devices. The system description provides the source for extracting information relevant for each control device (ECU extract), and this information is passed on to the control device's manufacturer. The manufacturer then configures the control device (ECU config) and generates executable software with the aid of code generators (see Fig. 4).

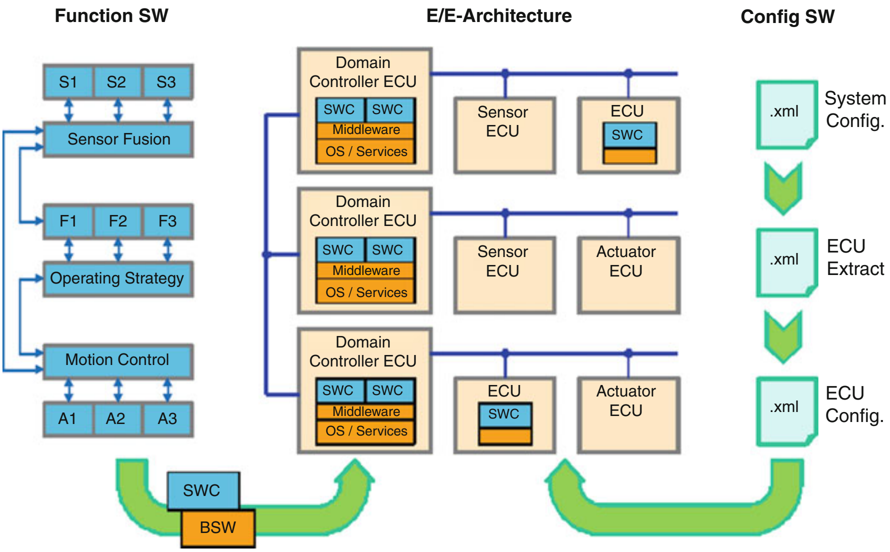


Fig. 4 Integration process for system architecture with domain control devices as integration platforms and sensor/actuator control devices in the sub-domains

The domain control devices are scalable with regard to computing power (single-, dual, and – in the future – even multi-core controllers), memory, and security level (up to ASIL D). They feature AUTOSAR-compatible software architecture and offer the opportunity to integrate software modules from various partners. The modular structure permits variants with reduced functionality.

5 Motion Control

Motion control is a domain for itself and demands careful functional and engineering integration due to the great variety of functions and control elements. The following describes an approach to structuring.

Tire forces acting on the plane formed by the street determine the horizontal motion of the system of mass known as a car. Governing the tire forces, in turn, is a car's current steering angle and wheel torque. Chassis forces determine vertical motion. The basic principle of motion control contemplates the opposite effect: using an intended motion as a starting point, motion control determines the appropriate tire forces and, hence, such variables as steering angle, steering torque, drive, or braking torque for the wheels. Effects such as crosswind or sloping road surface must be compensated for by adjustments in conjunction with feedforward control bits.

Dividing the system into several logically separate levels, with clearly defined interfaces, yields advantages in determining any adjustments (see Fig. 5). It will

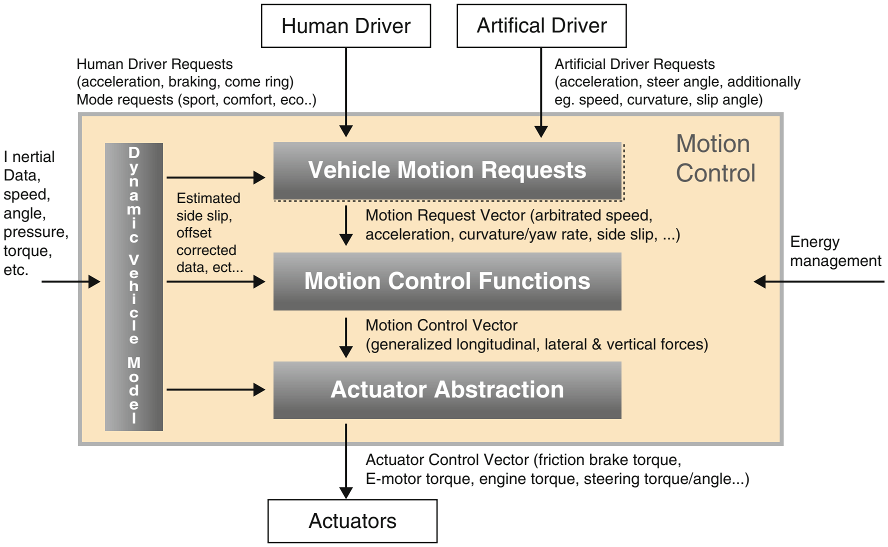


Fig. 5 Motion control-functional architecture

then be easier and more efficient to perform changes or additions to the system because usually only one level is affected instead of the entire system.

The *vehicle-motion-request level* continually registers driver intentions. It can also accept and process separate signals via switches, buttons, or the like. The third class of input includes motion requests from external systems such as driver-assistance systems. All requests are converted into projected data, which is then coordinated and passed on to the next level.

The *motion-control-function level* sees functions that control the motion of the car in longitudinal, lateral, and vertical directions. The goal is conflict-free optimization of safety, comfort, emotion, and efficiency. The output of the motion-control-function level is a vector referring to the entire car.

The adjustments to the different actuators are determined at the *actuator-abstraction level* from the reference vector. The focus here is on coordinating the chain of events. The actuator-abstraction level must have precise data on the current status and limitations of the adjusting actuator systems. The maximum tire forces in the Kamm circle must also be taken into consideration.

References

AUTOSAR (2014) AUTOSAR homepage. www.autosar.org. Accessed 18 July 2014

Baldwin CY, Clark KB (2000) Design rules, vol 1, The power of modularity. MIT Press, Cambridge, MA

Bulwahn L, Ochs T, Wagner D (2013) Research on an open-source software platform for autonomous driving systems. In: Proceedings of 19. Esslinger Forum für Kfz-Mechatronik

Eppinger SD, Browning TR (2012) Design structure matrix methods and applications. MIT Press, Cambridge, MA

Eriksson R, Alminger H (2013) Volvo Cars system architecture that fulfills vehicle needs of the future. In: Proceedings of VDI-Kongress Elektronik im Kraftfahrzeug, Baden-Baden

Feigel HJ (2012) MK C1 – a new generation of integrated brake. In: Proceedings of μ-symposium, Bad Neuenahr

Feigel HJ (2012) Integriertes Bremssystem ohne funktionale Kompromisse. Automobiltechnische Zeitschrift ATZ, 07–08, 2012, pp 612–617

Held V (2009) The chassis control systems of the Opel Insignia: design goals, control strategies and the system integration. In: Proceedings of TÜV-congress chassis tech, Munich

Hilbrich R, Gerlach M (2011) Virtualisierung bei Eingebetteten Multicore Systemen: Integration und Isolation sicherheitskritischer Funktionen. In: Proceedings of 41. Jahrestagung der Gesellschaft für Informatik e.V., Berlin

Kelling E, Raste T (2011) Trends in der Systemvernetzung am Beispiel einer skalierbaren E/E-Architektur für die Domänen Vehicle Motion und Safety. In: Proceedings of VDI-Kongress Elektronik im Kraftfahrzeug, Baden-Baden

Lochau M, Müller TC, Steiner J, Goltz U, Form T (2009) Optimierung von AUTOSAR-Systemen durch automatisierte Architektur-Evaluation. In: Proceedings of VDI-Kongress Elektronik im Kraftfahrzeug, Baden-Baden

OpenDDS (2014) OpenDDS homepage. www.opendds.org. Accessed 18 July 2014

Reichart G, Vondracek P, Bruckmeier R (2007) Systemarchitektur im Kraftfahrzeug – Status und künftige Anforderungen. In: Proceedings of Automobil Elektronik, Ludwigsburg

ROS (2015) ROS homepage. www.ros.org. Accessed 18 July 2014

Schwarz R, Bauer U, Tröster S, Fritz S, Muntu M, Schräbler S, Weinreuter M, Maurischat C (2003) ESP II – Fahrdynamik der nächsten Generation. Automobiltechnische Zeitschrift ATZ, Teil 1: Komponenten und Funktionen, 11, 2003, S. 1062–1069, Teil 2: Funktionsintegration und Elektronik, 12, 2003, S. 1178–1182

Smakmann H, Köhn P, Vieler H, Krenn N, Odenthal D (2008) Integrated Chassis Management – ein Ansatz zur Strukturierung der Fahrdynamikregelsysteme. 17. Aachener Kolloquium Fahrzeug- und Motorentechnik, Aachen

Weidner T, Stüker D, Wender S, Katzwinkel R, Vukotich A (2013) Skalierbare E/E-Architekturen als Enabler innovativer Fahrerassistenzfunktionen bis zum hochautomatisierten Fahren. In: Proceedings of VDI-Kongress Elektronik im Kraftfahrzeug, Baden-Baden

Integrated Trajectory Optimization

56

Moritz Werling

Contents

Abstract

New active driver assistance systems that work at the road and navigation level as well as automated driving face a challenging task. They have to permanently calculate the vehicle input commands (such as those for the steering, brakes, and the engine/powertrain) in order to realize a desired future vehicle movement, a driving trajectory. This trajectory has to be optimal in terms of some optimization criterion (in general a trade-off between comfort, safety, energy effort, and traveling time), needs to take the vehicular dynamics into account, and must incorporate lane boundaries or the predicted free space amid (possibly moving) obstacles. This kind of optimization can be mathematically formulated as a so-called optimal control problem. In order to limit the calculation effort, the

---

M. Werling (✉)
BMW Group, München, Germany
e-mail: moritz.werling@bmw.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_64

optimal control problem is usually solved only on a limited prediction interval (starting with the current time) leading to a receding horizon optimization.

The chapter illustrates this practically proven approach in detail. Furthermore, the three general principles of dynamic optimization known from control theory and robotics are presented, namely, calculus of variations, direct optimization, and dynamic programming. Furthermore, their application to driver assistance systems and automated driving is exemplified and the high practical relevance supported by the given literature. Finally, the respective advantages and limitations of the optimization principles are discussed in detail proposing their combination for more involved system designs.

1 Introduction

Advanced collision avoidance systems, lane keeping support, traffic jam assistance, and remote valet parking all operate on the actuators to relieve the drivers of the lateral and/or longitudinal vehicle control or make it safer for them. Well-defined tasks, such as staying in the middle of the marked lane while following the vehicle ahead, can be handled by a set-point controller (Gayko 2012). And yet standard automated parking maneuvers already require a calculated trajectory (more precisely a path, which does not have any time dependency) that has to be adapted to the available parking space (Katzwinkel et al. 2012). As systems need to cover more and more situations, the number of degrees of freedom increases, which makes a trajectory parameterization very complex, especially when the vehicle has to take numerous obstacles into account. This calls for a systematic approach on the basis of mathematical optimization (as opposed to heuristic approaches such as potential field and elastic band methods, see, e.g., Krogh (1984) and Brandt (2008), with their inherent limitations, cf. Koren and Borenstein (1991)).

In the chapter at hand, we address real-time *trajectory optimization*, a task that an automated vehicle faces when it travels through its environment, also referred to as *motion planning* in robotics (Latombe 1990; LaValle 2006). (Notice that in control theory the term *trajectory planning* usually implies that there is no feedback of the actual system states on the trajectory. The dynamical system is then only stabilized by a downstream trajectory tracking controller, which is not always advisable. We therefore use the term *trajectory optimization* instead to be independent of the utilized stabilization concept.) The focus will be on methods that engage with the longitudinal and lateral vehicle movement. However, the results can be transferred to novel warning systems that can also benefit from an optimal trajectory prediction, see, e.g., Eichhorn et al. (2013).

Speaking most generally, a trajectory optimization method is sought that can handle both structured (e.g., streets) and unstructured environments (parking lots), one that works among cluttered static obstacles and in moving traffic as well as exhibits a natural, human-like, anticipatory driving behavior.

Using more technical terms, the method should be easy to implement, parameterize, adapt, scale well with the number of vehicle states and the length of the optimization horizon, incorporate nonlinear, high-fidelity vehicle models, combine the lateral and longitudinal motion, be complete, allow for both grid maps and object list representations (with predicted future poses) of the obstacles, be numerically stable, and be transparent in its convergence behavior (if applicable). (A complete algorithm always finds the solution if it exists.) Also, the calculation effort has to be low to allow for short optimization cycles on (low-performance) electronic control units so that the vehicle can quickly react to sudden changes in the environment.

Unfortunately, there is no such single method that has all these properties. And, most likely, there will never be one. However, different optimization methods can be combined in order to get as close as possible to the above requirements. The next section therefore gives a closer look into the basic principles of trajectory optimization and their application.

2 Dynamic Optimization

When engineers speak about optimization, they usually refer to *static optimization,* in which the optimization variables $\boldsymbol{p}$ are finite, also called *parameters* (e.g., finding the most efficient operating point of an engine). Then *optimal* refers to some well-defined optimization criterion, usually the *minimization* of a *cost function* $J(\boldsymbol{p})$ (e.g., fuel consumption per hour).

Trajectory optimization is different in that the optimization variables are *functions* $\boldsymbol{x}(t)$ of an independent variable t, usually time. It is also called *dynamic* or *infinite-dimensional optimization*. Evaluating $\boldsymbol{x}(t)$ therefore requires a cost *functional* (a "function of a function"), which quantifies the "quality" of the trajectory $\boldsymbol{x}(t)$ by a scalar value.

Due to the vehicular focus, a special case will be considered, one that requires the trajectory $\boldsymbol{x}(t)$ to be consistent with some dynamical system model which has an input $\boldsymbol{u}$. Without such a model, the optimization cannot incorporate the inherent properties and physical limitations of the vehicle. This special case of dynamic optimization is called an *optimal control problem* (e.g., Lewis and Syrmos 1995).

2.1 Optimal Control Problem

A fairly general formulation of the optimal control problem (OCP) reads,

Minimize the cost functional

$$J(\boldsymbol{u}(t)) = \int_0^{t_f} l(\boldsymbol{x}(t), \boldsymbol{u}(t), t)\mathrm{d}t + V\left(\boldsymbol{x}\left(t_f\right), t_f\right) \tag{1}$$

subject to the system dynamics

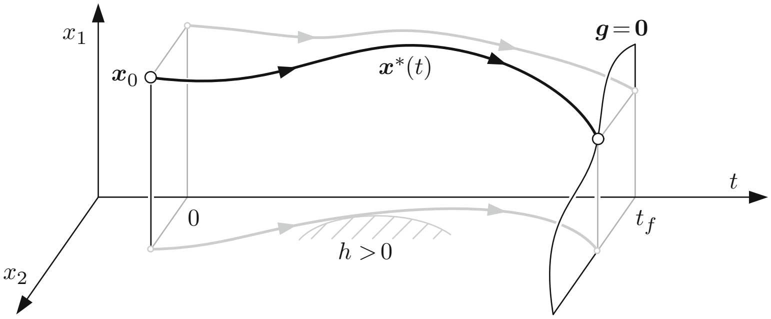


Fig. 1 Example of an optimal trajectory $x^*(t)$ in $\mathbb{R}^2$ with inequality constraint $h \leq 0$, a fixed final end time t_f, and end constraints $g = 0$

$$\dot{x}(t) = f(x(t), u(t), t), x(0) = x_0 \tag{2}$$

as well as the equality and inequality constraints

$$g\left(x\left(t_f\right), t_f\right) = 0 \quad \text{and} \tag{3}$$

$$h(x(t), u(t), t) \leq 0 \text{ for all times on the interval } \left[0, t_f\right]. \tag{4}$$

In other words, for our (possibly nonlinear and time-variant) system with state $x \in \mathbb{R}^n$ and input $u \in \mathbb{R}^m$ we seek on the interval $t \in \left[0, t_f\right]$ the input trajectory $u(t)$ that minimizes the cost functional J while steering (in the truest sense of the word) the system from its initial state x_0 to an end state $x(t_f)$, so that the equality and inequality constraints are fulfilled at all times. The optimal input trajectory is usually denoted by $u^*(t)$ and the resulting state trajectory by $x^*(t)$, see Fig. 1. Notice that J comprises integral costs l and endpoint costs V and is not only a functional of the input $u(t)$ but also of the states $x(t)$. Furthermore, if the final time t_f is not given, it then becomes part of the optimization, so that the length of the trajectory will also be optimized.

2.2 Problem Formulation for DAS and Automated Driving

Coming back to the automotive application, Eq. 2 of the previous section describes the dynamics of the vehicle. This state space model also includes the planar motion, either relative to some reference such as the lane center, see Sect. 3.1.2, or relative to a stationary origin, see Sect. 3.3.4. Undesired vehicle motion such as deviations from the lane center, detours, dangerous vehicle states (e.g., large slip angles), or uncomfortable jerks, e.g., caused by hectic steering actions, will be penalized in the cost functional Eq. 1. The free space prediction between the generally moving obstacles can be described by the inequality constraints Eq. 4. And the end constraints Eq. 3 can be utilized to require that the optimized vehicle trajectory will be aligned to the road at the end of the optimization interval. As will be explained in Sect. 5, this final state plays an important role for the stability of the "replanning" algorithm, therefore special costs $V(\boldsymbol{x}(t_f), t_f)$ can be introduced in Eq. 1.

3 Solving the Optimal Control Problem

All known approaches to the OCP can be assigned to one of the following three principles, see, e.g., Diehl et al. (2006).

3.1 Approach I: Calculus of Variations

The classical approach to the OCP is *calculus of variations*, which delivers valuable insight in the solution.

3.1.1 Theoretical Background: Hamilton Equations

Static optimization problems can be tackled by *differential calculus*. It is well known that the first derivative of a function $J(\boldsymbol{p})$ is equal to zero at a minimum (or any other stationary point). For multivariate problems we can write $\nabla J(\boldsymbol{p}) = \mathbf{0}$, which leads to a set of (algebraic) equations that the optimum $\boldsymbol{p}^*$ has to satisfy. The extension to problems with equality constraints requires the method of so-called Lagrange multipliers, yielding *first-order necessary conditions* for optimality.

Analogously for dynamic optimization, variational calculus requires that the *first variation* of the functional $J(\boldsymbol{u}(t))$ vanishes for the optimal control function $\boldsymbol{u}^*(t)$, which is often written as $\delta J(\boldsymbol{u}(t)) = 0$. In order to incorporate the system dynamics in the OCP, which constitute (differential) equality constraints, we can also apply the Lagrange multiplier method. This yields a *set of differential equations*, the so-called Hamilton equations:

$$\dot{\boldsymbol{x}} = f(\boldsymbol{x}, \boldsymbol{u}, t) \tag{5}$$

$$\dot{\boldsymbol{\lambda}} = -\frac{\partial l}{\partial \boldsymbol{x}} - \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right]^{\mathrm{T}} \boldsymbol{\lambda} \tag{6}$$

$$\mathbf{0} = -\frac{\partial l}{\partial \boldsymbol{u}} - \left[\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{u}}\right]^{\mathrm{T}} \boldsymbol{\lambda} \tag{7}$$

They are *first-order necessary conditions* for our OCP, when there are no inequality constraints Eq. 4 involved. The function $\boldsymbol{\lambda}(t)$ constitutes the Lagrange multipliers, here called *costates*. Besides the initial condition

$$\boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{8}$$

the optimal trajectory has to fulfill the (algebraic) *transversality conditions*, depending on whether the end state $\boldsymbol{x}(t_f)$ is constraint by Eq. 3 and/or the final time t_f is given (see, e.g., Lewis and Syrmos 1995). The most simple condition requires a fixed end state $\boldsymbol{x}_f$ at a given end time t_f so that

$$\boldsymbol{x}\left(t_f\right) = \boldsymbol{x}_f. \tag{9}$$

Either way, this results in a *boundary value problem*, which in general needs to be solved numerically. This so-called indirect approach is very accurate but not as flexible as the direct approach that we will introduce in Sect. 3.2. However, for simple OCPs the resultant boundary value problem can be solved analytically, leading to fast computable optimal trajectory primitives with broad applications (see Sect. 4).

3.1.2 Example Application: Automated Lane Change

We will now apply the described method to the generation of optimal lane change primitives. The lateral motion across the road can be modeled as a triple integrator system with states $\boldsymbol{x} = [x_1, x_2, x_3]^{\mathrm{T}}$, namely, the position, the lateral velocity, and the lateral acceleration, respectively, all within the reference frame of some curve, see Fig. 2.

Then, the system dynamics are described by

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u) = [x_2, x_3, u]^{\mathrm{T}}, \tag{10}$$

where u represents the lateral jerk, which is the third derivative of the lateral position. We will now seek for the optimal system input $u^*(t)$ that transfers the integrator system from its initial state

$$\boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{11}$$

to a given end state

$$\boldsymbol{x}\left(t_f\right) = \boldsymbol{x}_f \tag{12}$$

at a given end time t_f. Among all trajectories we seek for the one that minimizes the integral of the jerk-square, that is,

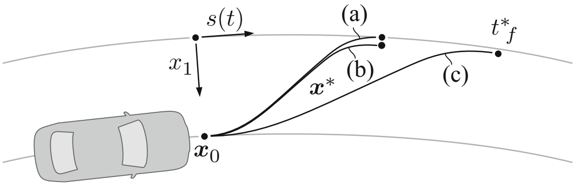


Fig. 2 Optimal lane changes for (*a*) a given end time and end state; (*b*) a given end time and a free end state; (*c*) a free end time and a free end state

$$l = \frac{1}{2} u(t)^2, \tag{13}$$

so that the movement feels most pleasant to the passengers. Notice that the final cost here has no influence on the solution due to the fixed end state, so that we can set $V = 0$ in Eq. 1.

With $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^{\mathrm{T}}$, evaluating the so-called control equation (7) we obtain

$$0 = u + \lambda_3 \Rightarrow u = -\lambda_3. \tag{14}$$

The costate equation (6) yields

$$\dot{\boldsymbol{\lambda}} = \left[\dot{\lambda}_1, \dot{\lambda}_2, \dot{\lambda}_3\right]^{\mathrm{T}} = [0,\ -\lambda_1,\ -\lambda_2]^{\mathrm{T}} \tag{15}$$

and therefore we get

$$\Rightarrow \lambda_1 := -c_1, \quad \lambda_2 = c_1 t + c_2, \quad \text{and} \quad \lambda_3 = -\frac{1}{2} c_1 t^2 - c_2 t - c_3. \tag{16}$$

With λ_3 from Eq. 16 it can be seen from Eq. 14 that the optimal control input function is a third-order polynomial with yet unknown integration constants c_1, c_2, and c_3. Substituting $u(t)$ in the state equation (10) we find the optimal trajectory to be

$$\dot{x}_3 = -\lambda_3 \Rightarrow x_3(t) = \frac{1}{6} c_1 t^3 + \frac{1}{2} c_2 t^2 + c_3 t + c_4 \tag{17}$$

$$\dot{x}_2 = x_3 \Rightarrow x_2(t) = \frac{1}{24} c_1 t^4 + \frac{1}{6} c_2 t^3 + \frac{1}{2} c_3 t^2 + c_4 t + c_5 \tag{18}$$

$$\dot{x}_1 = x_2 \Rightarrow x_1(t) = \frac{1}{120} c_1 t^5 + \frac{1}{24} c_2 t^4 + \frac{1}{6} c_3 t^3 + \frac{1}{2} c_4 t^2 + c_5 t + c_6 \tag{19}$$

with additional integration constants c_4, c_5, and c_6. To comply with the initial state Eq. 11 and end constraint Eq. 12 the integration constants need to be chosen accordingly. As Eqs. 16, 17, 18, and 19 are linear with respect to the constants this can be done by simple linear algebra, leading to the trajectory (a) in Fig. 2.

Similarly, we can find the solution to the OCP with a free end state (and a free end time). In this case we require the end state to be as close to (and as soon at) the reference as possible by setting

$$V(x) = k_1 x_1 \left(t_f\right)^2 + k_2 x_2 \left(t_f\right)^2 + k_3 x_3 \left(t_f\right)^2, k_i > 0$$
$$\left(V(x) = k_t t_f + k_1 x_1 \left(t_f\right)^2 + k_2 x_2 \left(t_f\right)^2 + k_3 x_3 \left(t_f\right)^2, k_i, k_t > 0\right)$$

The new transversality conditions will only lead to a different end state and end time, see (b) and (c) in Fig. 2. However, the optimal function class, that is, the fifth-order polynomial, will stay the same.

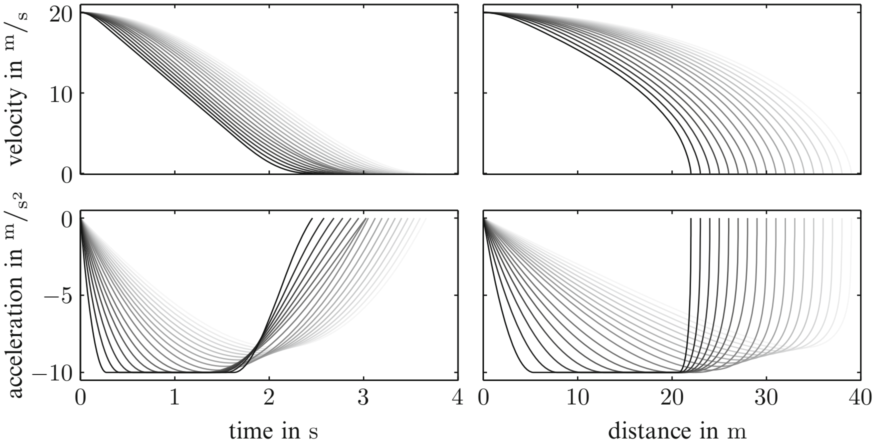


Fig. 3 Minimal jerk-square minimal time stopping trajectories with acceleration constraint $a \geq -10\ \mathrm{m/s^2}$

3.1.3 Further Readings

The previous calculations can be analogously carried out for the longitudinal movement, which leads to a very comfortable braking characteristic (Gutjahr and Werling 2014). The longitudinal and lateral movements can also be combined by a regular local and temporal sampling of target states across and along the street. This leads to a reactive algorithm which prevents collisions with static and moving obstacles (Werling et al. 2012).

The variational approach was generalized to problems with input constraints, known as *Pontryagin's minimum principle*. As for the vehicular application, it yields the shortest path connecting two poses of a vehicle with a limited turn radius (Dubins 1957; Reeds and Shepp 1990; Boissonnat et al. 1994). Even more involved is finding variational solutions to problems with state constraints such as the optimal braking application in Fig. 3 (e.g., Bryson and Ho 1975).

Numerical solutions to the first-order necessary conditions can be found by so-called indirect methods (see e.g., Graichen 2012) which provide very accurate results in general. Contrary to direct methods, as described in the sequel, they need to determine initial conditions for the costates, which has made the applications unsuitable for automotive online applications yet.

3.2 Approach II: Direct Optimization Techniques

Direct optimization is probably the most widely explored approach in model-predictive control. It approximates the dynamic optimization problem of the OCP to a static one, as the latter can be efficiently solved by well-established numerical

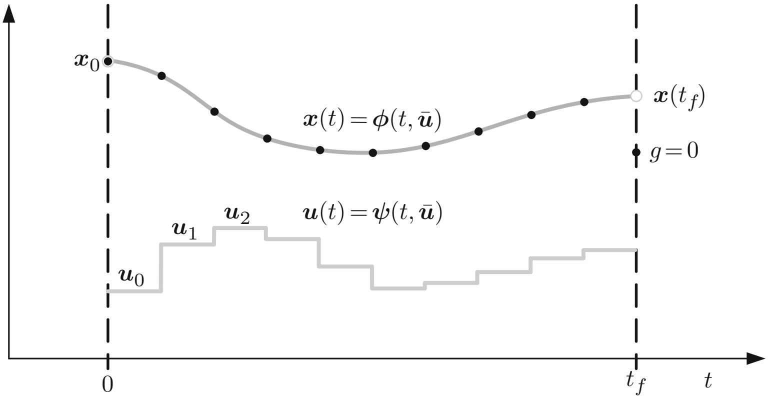


Fig. 4 Finite parameterization of the input and sampling of the constraints

solvers. The essence of this approach is *a finite-dimensional parameterization* of the input, state, or output trajectory.

3.2.1 Theoretical Background: Finite Parameterization Approximation

Here, we will introduce a very common method called single shooting. In a first step, we chose a finite-dimensional parameterization of the input

$$\boldsymbol{u}(t) = \boldsymbol{\psi}(t, \bar{\boldsymbol{u}}), \tag{20}$$

such as a piecewise constant interpolation (see Fig. 4), a polynomial, or a spline. The input trajectory is therefore fully described by the finite parameter vector $\bar{\boldsymbol{u}}$. The system dynamics Eq. 2 now read

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{\psi}(t, \bar{\boldsymbol{u}}), t), \quad \boldsymbol{x}(t_0) = \boldsymbol{x}_0. \tag{21}$$

This constitutes an initial value problem, which can be solved by an ordinary differential equation solver. We denote the resultant trajectory by

$$\boldsymbol{x}(t) = \boldsymbol{\phi}(t, \bar{\boldsymbol{u}}). \tag{22}$$

Furthermore, it is standard practice that the inequality constraints are only required to hold at N discrete equidistant points in time $t_i, i = 1, \ldots, N$, so that the number of inequality constraints is also finite.

Thus, the OCP was transformed to the following static optimization problem:

Minimize the cost function

$$J(\bar{\boldsymbol{u}}) = \int_0^{t_f} l(\phi(t, \bar{\boldsymbol{u}}), \psi(t, \bar{\boldsymbol{u}}), t) \ \mathrm{d}t + V\big(\boldsymbol{\phi}\big(t_f, \bar{\boldsymbol{u}}\big)\big) \tag{23}$$

subject to the equality and inequality constraints

$$\boldsymbol{g}\left(\boldsymbol{\phi}\left(t_f, \bar{\boldsymbol{u}}\right), t_f\right) = \boldsymbol{0} \tag{24}$$

$$\boldsymbol{h}(\boldsymbol{\phi}(t_i, \bar{\boldsymbol{u}}), \boldsymbol{\psi}(t_i, \bar{\boldsymbol{u}})) \leq \boldsymbol{0}, \quad i = 1, \ldots, N. \tag{25}$$

Loosely speaking, for a first guess $\bar{\boldsymbol{u}}_0$ the system will be simulated in a “single shoot” for $t \in \left[0, t_f\right]$ starting from $\boldsymbol{x}_0$. Then, the total costs J are evaluated as well as the constraints $\boldsymbol{g}$ and $\boldsymbol{h}$ in Eqs. 23, 24, and 25, respectively. These values are then fed back to a numerical solver, which repeats this procedure by a variation of $\bar{\boldsymbol{u}}$ to conclude how to modify the parameter so that J gets smaller without violating $\boldsymbol{g} = 0$ and $\boldsymbol{h} \leq 0$. When the solution does not significantly change any more or a certain number of iterations have been reached, the optimization will be terminated.

3.2.2 Example Application: Emergency Obstacle Avoidance

The direct method will now be applied to a combined steering and braking maneuver as part of a possible future active safety system (Werling and Liccardo 2012). For simplicity, we will approximate the vehicle dynamics by the static one-track model with only two parameters, the wheel base l and the characteristic speed v_{ch}. We then get

$$\dot{\theta} = v \frac{1}{l\left[1 + \left(\frac{v}{v_{ch}}\right)^2\right]} \delta, \tag{26}$$

where δ denotes the steering angle at the wheels, v is the velocity, and θ describes the course angle of the vehicle in some stationary reference frame. To avoid discontinuities in the steering angle trajectory, we extend the model by

$$\dot{\delta} = u, \tag{27}$$

so that the steering rate serves as the system input. We parameterize it by a *piecewise constant* input

$$u(t) = \psi(t, \bar{\boldsymbol{u}}) = u_i, t \in [t_i, t_{i+1}) \quad \text{and} \quad \bar{\boldsymbol{u}} = [u_0,\ u_1,\ \ldots, u_{N-1}]^{\mathrm{T}}$$

with N equidistant intervals.

Throughout the maneuver the brake should be engaged as much as physically reasonable, and therefore it is required that the vehicle accelerations stay right on the friction ellipse

$$\left[\frac{a_n}{c_n}\right]^2 + \left[\frac{a_t}{c_t}\right]^2 - 1 = 0,$$

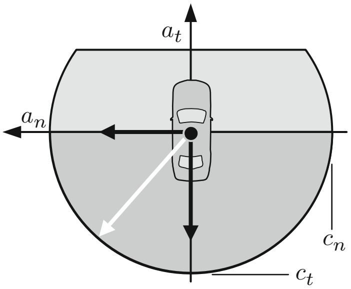


Fig. 5 Physical acceleration limits

see Fig. 5, where the lateral and longitudinal acceleration is given by $a_n = v\dot{\theta}$ and $a_t = \dot{v}$.

Solving the equation for a_n leads to the longitudinal dynamics

$$\dot{v} = -c_t\sqrt{1 - \left[\frac{v\dot{\theta}}{c_n}\right]^2} \tag{28}$$

of the full braking vehicle, where $\dot{\theta}$ has to be replaced by the expression in Eq. 26. In other words, the steering angle δ directly influences how much the vehicle is allowed to slow down. Hence, there is no need to optimize the braking power by a separate optimization variable.

Next, the planar vehicle motion will be modeled. Choosing street-relative coordinates bears some calculative advantages over Cartesian coordinates $[x_1, x_2]$ in the cost functional. We therefore describe the vehicle motion (approximately) by

$$\dot{d}_r = v \sin(\theta - \theta_r) \tag{29}$$

$$\dot{\theta}_r = v \cos(\theta - \theta_r)\kappa_r(s_r) \tag{30}$$

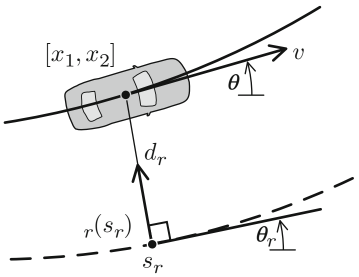
$$\dot{s}_r = v, \tag{31}$$

with the curve offset d_r with respect to the center of the lane, the curve's orientation θ_r and curvature κ_r, as well as the curve length parameter s_r. All variables refer to the vehicle projection point on the curve, see Fig. 6. The set of ordinary differential equations (26, 27, 28, 29, 30, and 31) is particularly nonstiff so that it can be integrated by the simple first-order Runge–Kutta scheme.

For the costs we choose

$$J = \int_0^{t_f} \left[u^2 + k_1(\theta - \theta_r) + k_2(\kappa - \kappa_r)\right] \,\mathrm{d}t \quad \text{with } \kappa = \frac{1}{l\left[1 + \left(\frac{v}{v_{ch}}\right)^2\right]}\delta \tag{32}$$

Fig. 6 System states relative to the road center



to avoid excessive steering rates as well as large deviations from the reference curve heading and curvature. The costs can be simply calculated by integrating the running costs l in parallel to the system dynamics, that is, $\dot{J} = l(\boldsymbol{x}, u)$.

In order to formulate inequality constraints that prevent the vehicle from colliding with the obstacle, for a given trajectory we define the point in time $t_{d,\min}$ when the (possibly moving) obstacle is the closest to the vehicle. The projection of the distance vector $\boldsymbol{\xi}^{\mathrm{T}}(d_r(t_{d,\min}), s_r(t_{d,\min}), t_{d,\min})$ to the obstacle onto the normal vector $\boldsymbol{n}\big(\theta\big(t_{d,\min}\big)$ of the vehicle course has to yield some safety clearance d_{safety}. Therefore

$$\boldsymbol{\xi}^{\mathrm{T}}\big(d_r\big(t_{d,\min}\big), s_r\big(t_{d,\min}\big)\big) \cdot \boldsymbol{n}\big(\theta\big(t_{d,\min}\big)\big) + d_{\text{safety}} \leq 0 \tag{33}$$

has to hold, when we want to force the vehicle to pass on the left side of the obstacle.

Also, we require the optimal trajectory to stay within the left lane boundaries, which can be expressed as

$$d_r(t_i) - d_{\max} \leq 0, i = 1, \ldots, N. \tag{34}$$

The resultant so-called nonlinear program executes with $N = 20$, $t_f = 2.0$s, and an initial guess of $\overline{\boldsymbol{u}}_0 = \boldsymbol{0}$ in approximately 100ms on an i5-520 M (2.4 GHz). A pedestrian example is shown in Fig. 7. As can be seen, the vehicle uses the full friction potential (bottom left) while braking and evading the pedestrian (black dot within gray safety margin) without leaving the given lane boundary (dotted line).

3.2.3 Further Readings

Single shooting has been intensively studied for many vehicular applications such as Kelly and Nagy (2003), Falcone et al. (2007a), Howard and Kelly (2007), Yoon et al. (2009), Gerdts et al. (2009), and Park et al. (2009). Nevertheless, many other numerical methods exist such as *multishooting* (Bock and Plitt 1984) and *collocation* (Hargraves and Paris 1987) that might gain in importance in the future when dealing with unstable vehicle models in challenging driving conditions. As for all direct optimization methods they lead to a static optimization problem, which can be most successfully solved by *sequential quadratic programming* (SQP) techniques or *interior point methods* (IP), see Nocedal and Wright (2006).

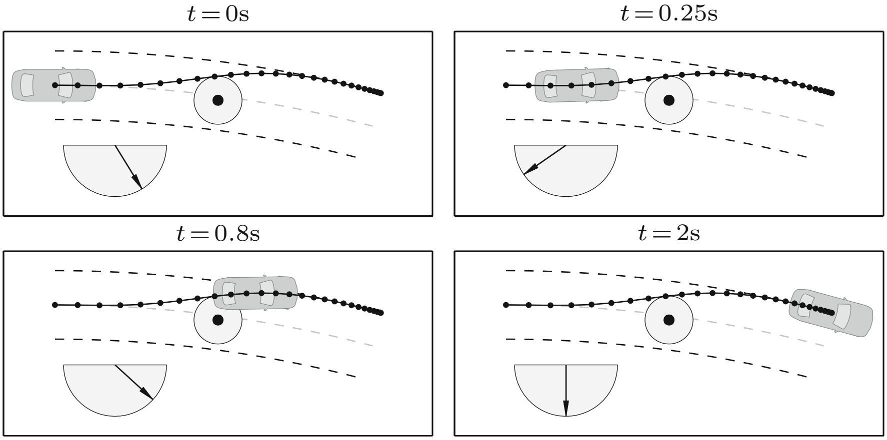


Fig. 7 Automatic obstacle avoidance with combined braking and steering at the friction limit; the safety margin d_{safety} is shown as a *gray circle* around the stationary pedestrian depicted as a *black dot*; the lane boundaries given by $d_{\max}$ are drawn as *gray dotted lines*

As an alternative to a finite parameterization of the input, for flat systems (Rouchon et al. 1993) the flat output can be parameterized instead (Kang 2012), which has been successfully demonstrated in complex inner-city scenarios by Ziegler et al. (2014).

In order to reduce the computational effort to a few milliseconds, Falcone et al. (2007b) and Carvalho et al. (2013), for instance, approximated the formulation of trajectory optimization as a *quadratic problem* (QP). This way no iterations are required as the solution can be found in a single step.

3.3 Approach III: Dynamic Programming

Certain tasks such as parking in several moves or finding the way through multiple moving obstacles are combinatorial (*nonconvex*) problems. They cannot be tackled by (local) direct optimization methods as the latter rely on an initial solution. However, dynamic programming is a principle from Bellman (1954) that significantly reduces the computational burden of these combinatorial problems, so that the *global optimum* can be efficiently obtained. It is therefore no surprise that it can be found in numerous (discrete) optimization algorithms.

3.3.1 Theroetical Backgorund: Bellman's Principle of Optimality

Richard Bellman's *principle of optimality* states,

"An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

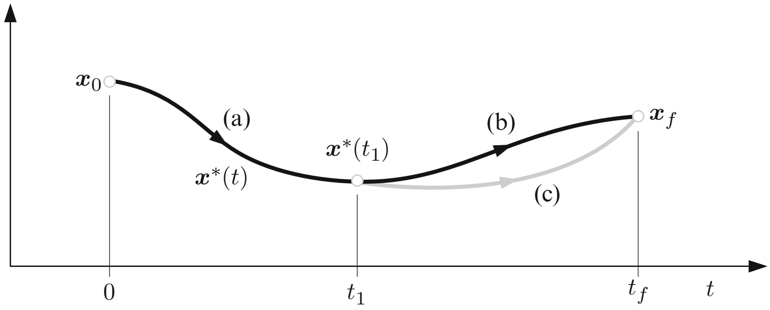


Fig. 8 Illustration of Bellman's principle of optimality

In other words, an optimal trajectory is composed of optimal subtrajectories. This can be understood by looking at Fig. 8, which shows $x^*(t)$ in black connecting x_0 with x_f. The subtrajectory (b), bringing $x^*(t_1)$ to x_f, has to be optimal, too. Otherwise, there would be another subtrajectory (c) (gray) that would lead to a better total trajectory (a) + (c), which is contrary to the optimal trajectory $x^*(t)$ comprising (a) + (b).

The largest practical benefit from this principle is received when time-discretizing the OCP. We therefore consider the time-discrete process

$$\boldsymbol{x}(k+1) = \boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{u}(k)), k = 0, \ldots k_f - 1, \boldsymbol{x}(0) = \boldsymbol{x}_0 \tag{35}$$

and seek for the optimal steering sequence $\boldsymbol{u}^*(k)$ that minimizes the cost function

$$J = \sum_{k=0}^{k_f-1} l(\boldsymbol{x}(k), \boldsymbol{u}(k), k). \tag{36}$$

Notice that constraints can be easily incorporated in Eq. 36 by setting the costs $l = \infty$ when the equality or inequality equations are violated. A basic element of dynamic programming is memorizing the costs of subtrajectories. Therefore, we define the so-called cost-to-go

$$G = \sum_{\kappa=k}^{k_f-1} l(\boldsymbol{x}(\kappa), \boldsymbol{u}(\kappa), \kappa) \tag{37}$$

(notice the difference between κ and k), which incurs when going from an intermediate state $\boldsymbol{x}(k)$ all the way to k_f.

At this point, it should be noticed that the minimal cost-to-go, denoted by G^*, is nearly as good as the solution $\boldsymbol{u}^*$ itself, as will be seen later. The minimal cost-to-go can be found in Bellman's *Recursion Formula*

$$G^*(\boldsymbol{x}(k), k) = \min_{\boldsymbol{u}(k)} \left\{ l(\boldsymbol{x}(k), \boldsymbol{u}(k), k) + G^*\left(\boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{u}(k), k), k+1\right\} \right. \tag{38}$$

which can be derived from the principle of optimality in a few steps. In words, it relates the minimal cost-to-go of the state $\boldsymbol{x}(k)$ at the k^{th} step, namely, $G^*(\boldsymbol{x}(k), k))$,

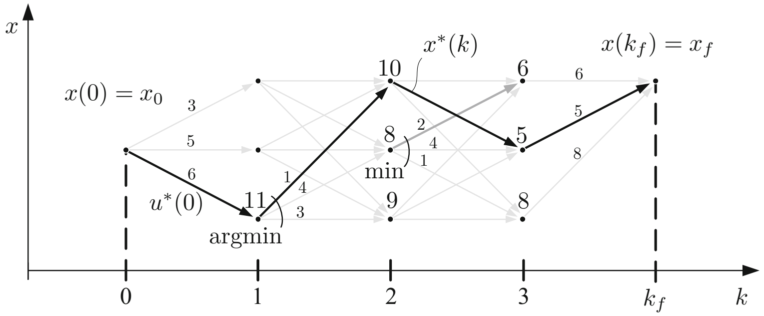


Fig. 9 Discrete decision process with the optimal trajectory

to the subsequent minimal cost-to-go $G^*(\boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{u}(k), k), k+1)$ in consideration of the best choice of the possible inputs $\boldsymbol{u}(k)$ with its associated stage cost $l(\boldsymbol{x}(k), \boldsymbol{u}(k), k)$. The formula is used to break down the multistaged OCP into simpler single-staged optimizations, which numerous algorithms exploit it to their advantage.

3.3.2 Value Iteration

A very suitable and likewise simple algorithm to solving OCPs is *value iteration* that we will present in its backward form. We assume that the system input can only take discrete values from a given set, which depends on the current state and the current time, that is, $\boldsymbol{u}(k) \in \mathcal{U}(\boldsymbol{x}(k), k)$. The set ensures that the input will transfer the system from one discrete state $\boldsymbol{x}(k) \in \mathcal{X}(k)$ to the next. We therefore get a multistage decision process, such as the simple example in Fig. 9 with three discrete states in each time step (except for the goal state $\boldsymbol{x}_f$).

Even though the naive approach of evaluating all $3^3 = 27$ possibilities would here be feasible, this procedure is clearly prohibitive for realistically sized problems due to an exponential runtime of $O\left(m^{k_f}\right)$, where m is the number of state transitions and k_f the optimization horizon. However, we can apply the recursion formula, start with the last stage $k = k_f - 1$, and work our way back to the first step. In doing so we calculate the optimal cost-to-go of every state of each stage by contemplating all possible transients and memorize it. This way it can be accessed when evaluating the optimal cost-to-go of the previous stage. Every stage can be done in $O(m \cdot n)$, where m and n equal the number of its inputs and states. This adds up to a total runtime of $O(N \cdot m \cdot n)$, which is only linear in the length of the optimization horizon. Figure 10 summarizes the algorithm.

The optimal input and state sequences $\boldsymbol{u}^*(k)$ and $\boldsymbol{x}^*(k)$ can now be found by an efficient forward search starting at $\boldsymbol{x}_0$ and alternating

$$\boldsymbol{u}^*(\boldsymbol{x}(k), k) = \text{argmin}_{\boldsymbol{u}(k)}\left\{l(\boldsymbol{x}(k), \boldsymbol{u}(k), k) + G^*\left(\boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{u}(k), k), k+1\right\}\right. \tag{39}$$

and Eq. 35 for every stage to the end, see black arrows in Fig. 9.

```
$G^*(\boldsymbol{x}_f,\ k_f) \leftarrow 0$
for $k = k_f - 1$ to $0$ do
    for all $\boldsymbol{x} \in \mathcal{X}$ do
        $G^*(\boldsymbol{x}(k),\ k) = \min_{\boldsymbol{u}(k)} \{\, l(\boldsymbol{x}(k), \boldsymbol{u}(k), k) + G^*(\boldsymbol{f}(\boldsymbol{x}(k), \boldsymbol{u}(k), k), k+1) \,\}$
    end for
end for
```

Fig. 10 Value iteration algorithm

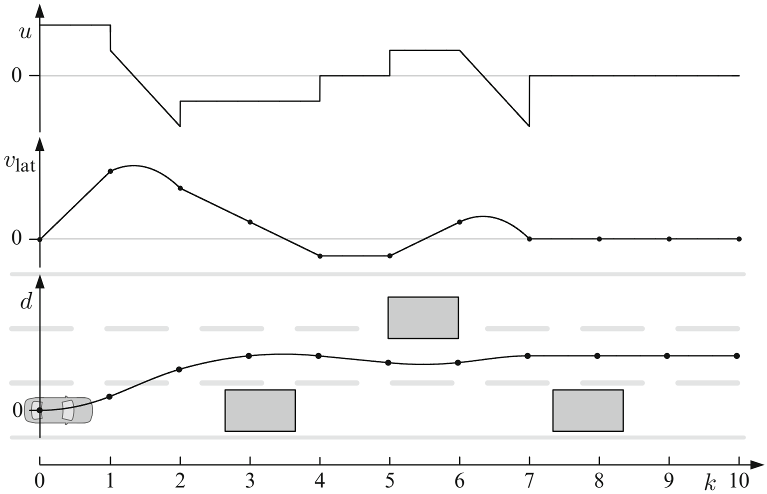


Fig. 11 Dynamic programming example

3.3.3 Example Application: Optimal Overtaking

A simplified problem setup and its solution can be found in Fig. 11. Here, the lateral motion is optimized, so that the car can proceed at a constant speed without getting too close to the other vehicles. The system state comprises of only the lateral position and velocity, hence $\boldsymbol{x}(k) = [d(k), v_{\text{lat}}(k)]^{\text{T}}$, which is discretized with $\Delta d = 0.5\text{m}$ and $\Delta v_{\text{lat}} = 0.5\text{m/s}$ at $t_k = k \cdot \Delta t$ with $k = 1, .., 10$ and $\Delta t = 1.0\text{s}$. Furthermore, the input set $\mathcal{U}$, i.e., all lateral acceleration profiles, is chosen as a first-order polynomial with $-10.0\text{m/s}^2 < u(t) < 10.0\text{m/s}^2$. And lastly, the stage costs are defined as

$$l(\boldsymbol{x}(k),\ u(k), k) = \int_{t_k}^{t_k+\Delta t} u(t)^2 \mathrm{d}t + k[d(k) - d_{\text{nearest}}(k)]^2 + C_{\text{collision}}(\boldsymbol{x}(\boldsymbol{k}),\ u(k), k)$$

with $k = 1.0$, where $d_{\text{nearest}}(k)$ denotes the lateral coordinate of the lane center closest to $d(k)$. The cost term $C_{\text{collision}}(\boldsymbol{x}(k),\ u(k),\ k)$ equals zero if the vehicle does

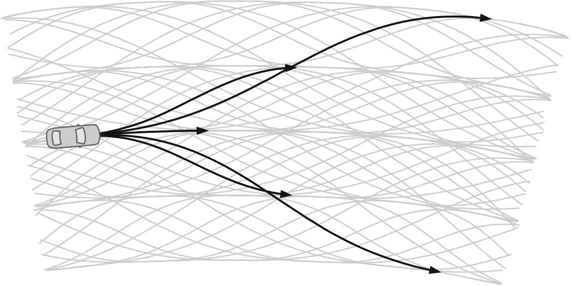


Fig. 12 State lattice for structured environments (illustration based on McNaughton (2011))

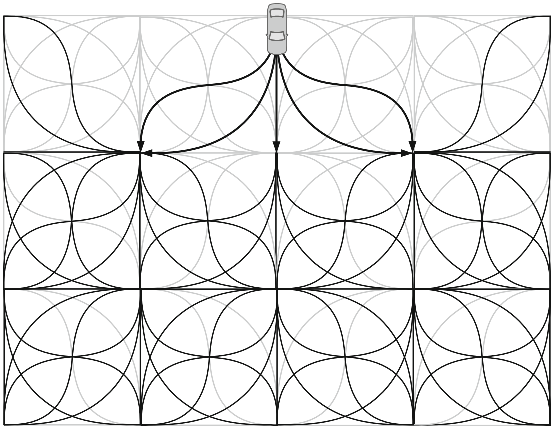


Fig. 13 State lattice for unstructured environments (illustration based on McNaughton (2011))

not collide within the interval $[t_k, t_k + \Delta t]$ and else equals infinity. Due to the few system states with their coarse discretization, the (global) solution is found within a few milliseconds.

3.3.4 Further Readings

The presented value iteration scheme is especially suited for *structured dynamic* environments as shown by Ziegler and Stiller (2009) and McNaughton (2011), as well as Gu and Dolan (2012). As can be seen in Fig. 13, for natural, human-like trajectories it is most advantageous to align the sampling along and across the road course. The costs usually penalize not only jerky, accelerant movements, and the approximation to obstacles, but also deviations from the desired road center.

As opposed to that, in unstructured environments, such as parking lots, there is no preferred direction, so that the sampling uniformly covers the $[x, y]$ plane as shown in Fig. 12. As static obstacles dominate the problem, temporal aspects are usually neglected meaning that it is not important at what time a location is visited by the vehicle (see Montemerlo et al. 2008; Ziegler et al. 2008; Pivtoraiko et al. 2009; Likhachev and Ferguson 2009; Dolgov et al. 2010). This drastically simplifies the problem; however, value iteration cannot be applied in a single sweep anymore due to the loss of the processing sequence given by k (the decision graph becomes cyclic). Therefore a substitute order needs to be chosen, one that still leads to the optimum: Investigating the state (expanding the node) with the lowest *cost-to-come* leads to *Dijkstra's* algorithm, unlike considering the state with the lowest (under)estimate on the total cost, which results in A* algorithm, an informed search. The latter is especially efficient in unstructured environments, as the costs are dominated by the covered distance, which can be well estimated by the Euclidean distance (the so-called heuristic), see Bertsekas (1995) (Fig. 13).

4 Comparison of the Approaches

When we compare dynamic programming with direct optimization we realize that the two approaches possess orthogonal capabilities, see Table 1. Dynamic programming suffers the so-called curse of dimension (Bellman 1954) meaning that the approach does not scale well with the number of states. Its application is therefore limited to models with few system states (<3–4), and also requires their coarse discretization. Direct optimization, however, can incorporate system models with numerous and continuous system states (>4) and is therefore able to directly feed the system input $\boldsymbol{u}(t)$. In turn, direct optimization cannot deal with arbitrary cost functionals due to numerical limitations (local convergence) of the underlying solver. Also, its runtime usually grows exponentially with the number of optimization variables so that the length of the optimization horizon is restricted to a few seconds. As opposed to that, dynamic programming can handle arbitrary costs and will always lead to the global optimum. Furthermore, dynamic programming scales comparably well with the length of the optimization horizon (cf. the linear runtime of the value iteration scheme in Sect. 3.3.3).

For complex, farsighted trajectory optimization tasks, the two approaches need to be combined. Dynamic programming will then yield only a "rough long-term plan," which serves as the reference trajectory (see, e.g., Ferguson et al. 2008a, b; Gu et al. 2013) and/or provides an initial guess (Kang 2012) for the locally working direct optimization method. The latter takes a detailed model of the vehicle into account and improves the dynamic programming solution on a reduced optimization horizon to a feasible trajectory (also see Sect. 5). The optimal state trajectory is

Table 1 Comparison and combination of the approaches

Approach	Many states	Continuous states	Global optimum	Long horizon
DP	⊖	⊖	⊕	⊕
DO	⊕	⊕	⊖	⊖
DP + DO	⊕	⊕	⊕	⊕
DP + DO + CV	⊕	⊕	⊕	⊕⊕

DP dynamic programming, *DO* direct optimization, *CV* calculus of variations either forwarded to a low-level feedback steering/acceleration controller or the optimal input trajectory is directly fed to the vehicle actuators.

Closed-form solutions from the calculus of variations are thereby often used to speed up dynamic programming. This can be in the form of a heuristic for an informed search (see, e.g., Ziegler et al. 2008) or so-called analytical expansions (Dolgov et al. 2010), both of which, roughly speaking, approximate the remaining trajectory and therefore extend the computable optimization horizon, see Table 1.

5 Receding Horizon Optimization

The receding horizon approach is the gist of *model-predictive control* (MPC, see, e.g., Rawlings 2000), which makes a numerical optimization practical for closed-loop control. Therein, in each step t_k, the OCP is solved on a finite horizon T, which calculates the optimal open-loop trajectories $\overline{\boldsymbol{x}}^*(\tau)$ over $\tau \in [t_k, t_k + T]$, see Fig. 14. Only the first part of the optimal control $\overline{\boldsymbol{u}}^*(\tau)$ is implemented on Δt. Right in time the new solution is available of the OCP that has been shifted by Δt. In classical MPC, at each t_k the current plant state is fed back as the new initial state of the OCP. Altogether, this leads to a closed control loop that anticipates future events, such as input and state saturation, and takes control actions accordingly.

This procedure is completely compatible with trajectory optimization for vehicles. Even more, its replanning mechanism can innately take the limited sensor range and predictability of the other traffic participants into account, which can fundamentally change the OCP from one optimization step to the other.

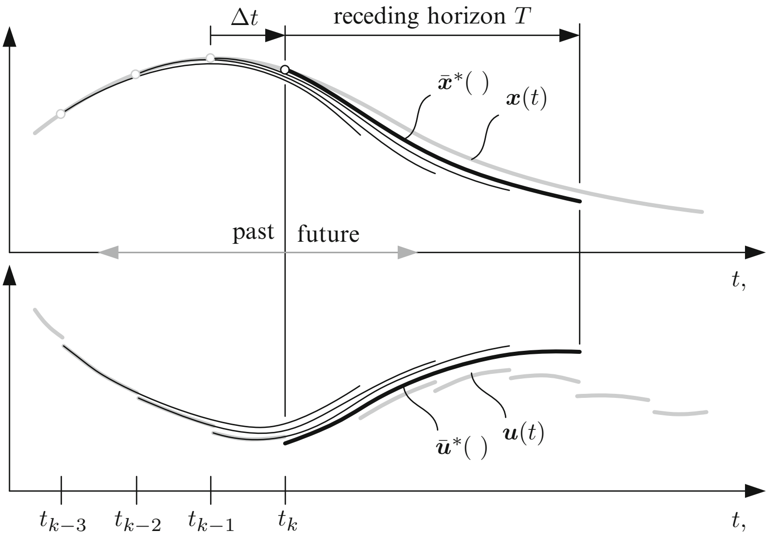


Fig. 14 Receding horizon optimization with optimization length T, cycle time Δt, current time step t_k, predicted and actual trajectory $\overline{\boldsymbol{x}}^*, \boldsymbol{x}$, predicted and actual system input $\overline{\boldsymbol{u}}^*$, $\boldsymbol{u}$

Furthermore, the approach leaves additional degrees of freedom, which can be used to increase the overall robustness of the closed-loop system. Firstly, the prediction model may not only include the plant but also the underlying fast low-level feedback or feedforward controllers, which are intended to simplify the resultant optimization model Eq. 2, see also Sect. 6. And secondly, the initial state of the OCP does not necessarily have to be the actual vehicle state but can also be the optimal trajectory of the last step sampled at the current time – or a combination of both. This works as long as the fed-back desired states of the optimal trajectory are tracked by an abovementioned low-level feedback controller. However, as soon as components of the current vehicle state are used that show up in the optimization constraints, so-called slack variables need to be introduced in the optimization. They transform the otherwise hard inequalities into *soft constraints* as they are referred to in the MPC lingo. This is required as otherwise the slightest disturbance or model uncertainty would ultimately lead to an initial state that the OCP with hard constraints does not have a solution for.

MPC theory offers even more. It also deals with stability issues of the closed-loop system such as in Grüne and Pannek (2011). More precisely, even for time-invariant system dynamics, constraints, and costs (typical of conventional control problems), the OCP solution changes over time due to the receding horizon, as indicated by Fig. 14. In the worst case the differences of the consecutive solutions build up and the system destabilizes. Well-established MPC schemes therefore propose an augmentation of the cost functional and constraints, e.g., by a special terminal constraint (*invariant set*, (Blanchini 1999)) and cost (*control Lyapunov function*, (Grüne and Pannek 2011)) in order to guarantee stability. These schemes can also be transferred to the automotive application. Even more interesting for collision avoidance are the *permanent feasibility* guarantees, which can prevent the suboptimal shortsighted solutions from “dead ends” also known as inevitable collision states (ICS, Fraichard (2007)) in robotics, but this is ongoing research (e.g., Althoff et al. 2012; Lawitzky et al. 2014).

6 Conclusion

Owing to the increase in processing power, computationally intensive optimization algorithms can be executed in real time in many industrial areas. As automotive electronic control units follow this trend, it is only a matter of time until receding horizon control techniques will emerge in production cars. As has been shown in this chapter, these techniques are most suitable for solving complex trajectory optimization tasks for novel driver assistance systems and automated driving. Based on well-known, elaborated principles we can combine different optimization techniques and implement fast, powerful algorithms with no need to reinvent the wheel.

In order to surmount the complex integration of numerous safety and comfort functions to a complete driver-friendly unit, an integrated trajectory optimization module is required. Ideally, such a module covers the superset of all emerging use cases. Conventional functions such as emergency braking and lane keeping will then only be special cases within the algorithm.

References

Althoff D, Werling M, Kaempchen N, Wollherr D, Buss M (2012) Lane-based safety assessment of road scenes using inevitable collision states. In: Intelligent vehicles symposium, Madrid (Spain), pp 31–36

Bellman R (1954) The theory of dynamic programming. RAND-P-550. Rand Corporation, Santa Monica

Bertsekas DP (1995) Dynamic programming and optimal control. Athena Scientific, Belmont

Blanchini F (1999) Set invariance in control. Automatica 35:1747–1767

Bock H, Plitt K (1984) A multiple shooting algorithm for direct solution of optimal control problems. In: 9th IFAC world congress, Budapest (Hungary), pp 242–247

Boissonnat JD, Cerezo C, Leblond J (1994) A note on shortest paths in the plane subject to a constraint on the derivative of the curvature. Report 2160, INRIA

Brandt T (2008) A predictive potential field concept for shared vehicle guidance. PhD thesis, University of Paderborn

Bryson AE, Ho YC (1975) Applied optimal control. Wiley, New York

Carvalho A, Gao Y, Gray A, Tseng H, Borrelli F (2013) Predictive control of an autonomous ground vehicle using an iterative linearization approach. In: Conference on intelligent transportation system, IEEE, The Hague (Netherlands), pp 2335–2340

Diehl M, Bock HG, Diedam H, Wieber PB (2006) Fast motions in biomechanics and robotics. Lecture notes in control and information sciences. Springer, Berlin, Heidelberg. pp 65–94

Dolgov D, Thrun S, Monteberlo M, Diebel J (2010) Path planning for autonomous vehicles in unknown semi-structured environments. Int J Robot Res 29(5):485–501

Dubins LE (1957) On curves of minimal length with a constraint on average curvature and with prescribed initial and terminal positions and tangents. Am J Math 79(3):497–516

Eichhorn A, Werling M, Zahn P, Schramm D (2013) Maneuver prediction at intersections using cost-to-go gradients. In: International conference on intelligent transportation systems, IEEE, The Hague (Netherlands), pp 112–117

Falcone P, Borelli F, Asgari J, Tseng E, Hrovat D (2007a) Predictive active steering control for autonomous vehicle systems. IEEE Trans Control Syst Technol 15(3):566–580

Falcone P, Tufo M, Borrelli F, Asgari J, Tsengz HE (2007b) A linear time varying model predictive control approach to the integrated vehicle dynamics control problem in autonomous systems. In: Conference on decision and control, IEEE, New Orleans (USA), pp 2980–2985

Ferguson D, Howard TM, Likhachev M (2008a) Motion planning in urban environments: part I. In: International conference on intelligent robots and systems, IEEE/RSJ, Nice (France), pp 1063–1069

Ferguson D, Howard TM, Likhachev M (2008b) Motion planning in urban environments: part II. In: International conference on intelligent robots and systems, IEEE/RSJ, Nice (France), pp 1063–1069

Fraichard T (2007) A short paper about motion safety. In: International conference on robotics and automation, IEEE, Singapore, pp 1140–1145

Gayko J (2012) Lane keeping support. In: Handbuch Fahrerassistenzsysteme. Vieweg + Teubner Verlag, Wiesbaden, pp 554–561

Gerdts M, Karrenberg S, Müller-Bessler B, Stock G (2009) Generating locally optimal trajectories for an automatically driven car. Optim Eng 10(4):439–463

Graichen K (2012) A fixed-point iteration scheme for real-time model predictive control. Automatica 48(7):1300–1305

Grüne L, Pannek J (2011) Nonlinear model predictive control: theory and algorithms. Springer, London

Gu T, Dolan JM (2012) On-road motion planning for autonomous vehicles. In: Intelligent robotics and applications. Springer, Berlin, pp 588–597

Gu T, Snider J, Dolan JM, Lee J (2013) Focused trajectory planning for autonomous on-road driving. In: Intelligent vehicles symposium, IEEE, Gold Coast City (Australia), pp 547–552

Gutjahr B, Werling M (2014) Automatic collision avoidance during parking and maneuvering – an optimal control approach. In: Intelligent vehicles symposium, IEEE, Dearborn (USA), pp 636–641

Hargraves CR, Paris SW (1987) Direct trajectory optimization using nonlinear programming and collocation. J Guid Control Dyn AAS/AIAA Astrodyn Conf 10:338–342

Howard TM, Kelly A (2007) Optimal rough terrain trajectory generation for wheeled mobile robots. Int J Robot Res 26(2):141–166

Kang K (2012) Online optimal obstacle avoidance for rotary-wing autonomous unmanned aerial vehicles. PhD-thesis, Georgia Institute of Technology

Katzwinkel R, Auer R, Brosig S, Rohlfs M, Schöning V, Schroven F, Wuttke U (2012) Einparkassistenz. In: Handbuch Fahrerassistenzsysteme. Vieweg + Teubner Verlag, Wiesbaden, pp 471–477

Kelly A, Nagy B (2003) Reactive nonholonomic trajectory generation via parametric optimal control. Int J Robot Res 22(7–8):583–601

Koren Y, Borenstein J (1991) Potential field methods and their inherent limitations for mobile robot navigation. In: International conference on robotics and automation, IEEE, Sacramento (USA), pp 1398–1404

Krogh BH (1984) A generalized potential field approach to obstacle avoidance control. In: International robotics research conference, Bethlehem

Latombe JC (1990) Robot motion planning. Springer Science+Business Media New York

LaValle SM (2006) Planning algorithms. Cambridge University Press, Cambridge

Lawitzky A, Nicklas A, Wollherr D, Buss M (2014) Determining states of inevitable collision using reachability analysis. In: International conference on intelligent robots and systems, IEEE/RSJ, Chicago (USA), pp 4142–4147

Lewis F, Syrmos V (1995) Optimal control. Wiley, New York

Likhachev M, Ferguson D (2009) Planning long dynamically feasible maneuvers for autonomous vehicles. Int J Robot Res 28(8):933–945

McNaughton M (2011) Parallel algorithms for real-time motion planning. PhD thesis, Carnegie Mellon University

Montemerlo M et al (2008) Junior: the Stanford entry in the Urban Challenge. J Field Robot 25(9):569–597

Nocedal J, Wright SJ (2006) Numerical optimization. Springer Science+Business Media New York

Park JM, Kim DW, Yoon YS, Kim HJ, Yi KS (2009) Obstacle avoidance of autonomous vehicles based on model predictive control. Proc Inst Mech Eng D J Automob Eng 223(12):1499–1516

Pivtoraiko M, Knepper RA, Kelly A (2009) Differentially constrained mobile robot motion planning in state lattices. J Field Robot 26(3):308–333

Rawlings JB (2000) Tutorial overview of model predictive control. IEEE Control Syst 20(3):38–52

Reeds JA, Shepp LA (1990) Optimal paths for a car that goes both forwards and backwards. Pac J Math 145(2):367–393

Rouchon P, Fliess M, Levine J, Martin P (1993) Flatness, motion planning and trailer systems. In: Conference on decision and control, IEEE, San Antonio (USA), pp 2700–2705

Werling M, Liccardo D (2012) Automatic collision avoidance using model-predictive online optimization. In: Conference on decision and control, IEEE, Maui (USA), pp 6309–6314

Werling M, Kammel S, Ziegler J, Gröll L (2012) Optimal trajectories for time-critical street scenarios using discretized terminal manifolds. Int J Robot Res 31(3):346–359

Yoon Y, Shin J, Kim HJ, Park Y, Sastry S (2009) Model-predictive active steering and obstacle avoidance for autonomous ground vehicles. Control Eng Pract 17(7):741–750

Ziegler J, Stiller C (2009) Spatiotemporal state lattices for fast trajectory planning in dynamic on-road driving scenarios. In: International conference on intelligent robots and systems, IEEE/RSJ, St. Louis (USA), pp 1879–1884

Ziegler J, Werling M, Schröder J (2008) Navigating car-like robots in unstructured environments using an obstacle sensitive cost function. In: Intelligent vehicles symposium, IEEE, Eindhoven (Netherlands), pp 787–791

Ziegler J, Bender P, Dang T, Stiller C (2014) Trajectory planning for bertha – a local, continuous method. In: Intelligent vehicles symposium proceedings, IEEE, Dearborn (USA), pp 450–457

Anti Collision System PRORETA

57

Rolf Isermann, Andree Hohm, Roman Mannale, Bernt Schiele,
Ken Schmitt, Hermann Winner, and Christian Wojek

Contents

R. Isermann (✉)
Institut für Automatisierungstechnik, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: risermann@iat.tu-darmstadt.de

A. Hohm • K. Schmitt
Continental Teves AG & Co. oHG, Frankfurt am Main, Germany
e-mail: andree.hohm@continental-corporation.com; ken.schmitt@gmx.de

R. Mannale
Continental Automotive GmbH, Schwalbach, Germany
e-mail: roman.mannale@continental-corporation.com

B. Schiele
Computer Vision and Multimodal Computing, Max Planck Institute for Informatics, Saarbrücken, Germany
e-mail: schiele@mpi-inf.mpg.de

H. Winner
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

C. Wojek
Carl Zeiss AG, Oberkochen, Germany
e-mail: cwojek@mpi-inf.mpg.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_57

Abstract

This contribution describes the basic concept and practical evaluation of a driver assistance system, which early detects dangerous overtaking maneuvers on two-lane rural roads and helps to prevent accidents. A fusion of video and RADAR data combines a high-precision detection of far objects with accurate lateral position and velocity estimates for nearer objects. The detection of nearer objects is performed by a video based vehicle classifier. The fused environment data is the basis to identify a hazardous situation by combining a signal based detection of the driver's overtaking intention with a problematic constellation of the involved vehicles. When such a hazardous situation is detected, warnings are initiated, and an automatic brake intervention aborts the dangerous overtaking maneuver at the last possible moment, so that the driver can get behind the preceding vehicle by swerving to his/her own lane. The system was developed during the Proreta 2 research project by Technische Universität Darmstadt in cooperation with Continental AG.

1 Introduction

In 2006, the Proreta research cooperation between Technische Universität Darmstadt and Continental AG presented a collision avoidance system for one-way traffic, including automatic emergency braking and emergency evasion (Bender et al. 2007). In a second project, the cooperation developed a driver assistance system for two-way traffic, especially for overtaking maneuvers on rural roads (Isermann et al. 2009). Figure 1 shows the main functionalities of the system and the three involved research institutes of Technische Universität Darmstadt.

In this contribution, the basic concept of the developed system and results of driving experiments are presented. If the system's calculations indicate a conflict

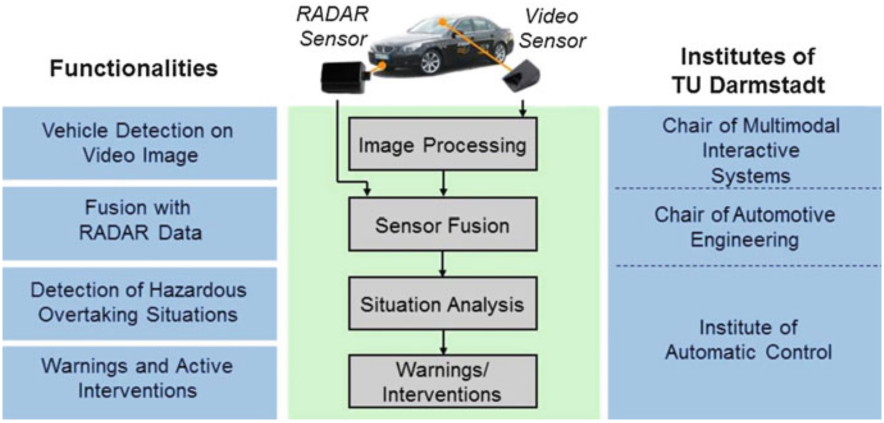


Fig. 1 Proreta 2 basic system architecture and involved research institutes#L

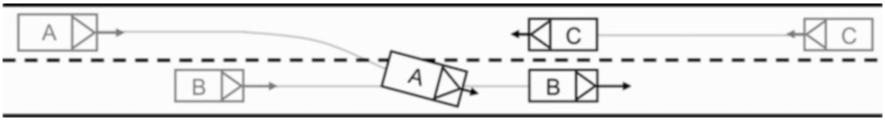


Fig. 2 Proreta 2 objective: abortion of hazardous overtaking maneuvers#L

with an oncoming vehicle during an overtaking maneuver, a collision avoidance strategy is initialized, which includes warnings and an automatic braking intervention, Fig. 2.

In Mannale et al. (2008), the required range of environmental sensors for such a system was estimated. For an initial speed of 90 km/h and an oncoming vehicle driving 120 km/h, even at a late time shortly before passing the preceding vehicle, the required detection range is 375 m. As that exceeds ranges of common ACC sensors by far, the Proreta 2 research vehicle was equipped with a modified 77 GHz RADAR with an extended range of 400 m.

2 Video-Based Scene Segmentation for Freespace Estimation

To get an image-based scene understanding in the close range (up to 50 m), a series CMOS color video camera (CSF200) is used. It is mounted on the car's windshield and determines an electronic representation for the situation in front of the car. The video image is segmented pixelwise into eight classes such as road, vehicle, grass or bush/tree. Besides local features at the pixel level, the developed method also incorporates the output of an object detector and the temporal dynamics of the video stream. Figure 3 shows the corresponding signal flow.

In the following, the method is presented in three stages. For a more detailed presentation, we refer the reader to Wojek and Schiele (2008). For a segmentation using so-called conditional random fields (CRF), groups of 8 × 8 pixels are built. For the resulting nodes, probability values are computed, which refer to the segmentation classes (road, vehicle, etc.). Thereby, local node potentials for larger pixel groups are computed based on filter bank responses and subsequent classification, see Torralba et al. (2004). Further, pairwise potentials allow modeling of neighborhood interactions.

Object detection methods (such as Dalal and Triggs (2005)) allow more robust results, as features are typically computed over a larger image area. To exploit and incorporate object detection confidences, the plain CRF model is augmented with additional random variables, resulting in the so-called object CRF. Finally, the expansion to the so-called dynamic CRF takes into account, that in case of highly dynamic overtaking scenarios, the dynamics for vehicles and for the scene background differ by far when projected to the image. Thus, vehicles are tracked with a Kalman filter, while the probability distribution of the segmentation at the current time step is propagated for the next input image.

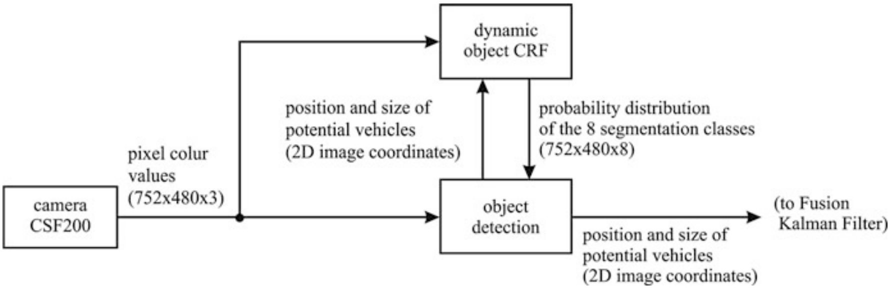


Fig. 3 Signal flow for video-based object detection and scene segmentation

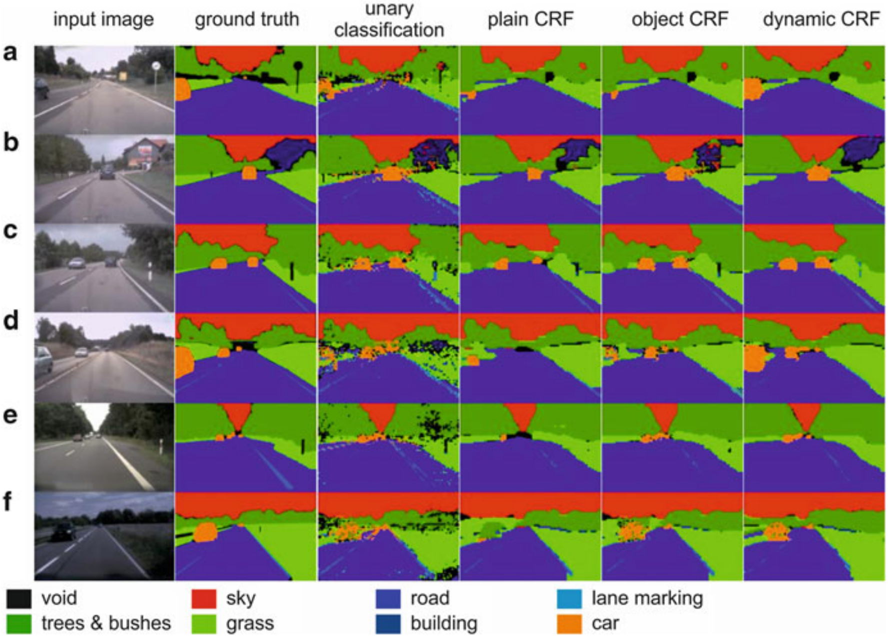


Fig. 4 Sample segmentation results with the presented models for highway road scenes

Figure 4 depicts some example results, while column 1 shows the input image, and column 2, being the result of a manual classification, shows the ground truth.

The plain CRF model (column 4) results in a much smoother result than unary filter based classification (column 3). However, unfortunately, it tends to eliminate pixels belonging to the class vehicle by overwriting them with other classes such as road. The additional random variables of the object CRF (column 5) improve the segmentation accuracy for vehicles substantially. Finally, the dynamic CRF model (column 6) can also improve the segmentation of vehicles, which are temporally not detected by the object detector.

This results in a segmentation of the entire scene on the video image and object detections from an image-based object detector being available for further processing.

3 Sensor Fusion of RADAR and Video Signals

In order to implement a driver assistance system for overtaking scenarios, the perception of objects in the environment is required. Within these, the oncoming traffic is crucial. To provide a detection range of about 400 m, which is required for an early analysis of overtaking situations, RADAR and LIDAR systems bear potential. In the project Proreta 2, preference was given to a RADAR sensor, due to certain advantages concerning the signal attenuation while detecting far away objects. The detection range of about 400 m could be achieved by modifying a series sensor.

To achieve a precise estimation and interruption-free detection of the state-vector for observed cars, object-tracking is used, applying an extended Kalman filter algorithm (EKF) (Winner et al. 2009; Darms and Winner 2006). The fusion is realized with the data of the object detector explained above, which has a better lateral detection performance than the RADAR in ranges up to 50 m. The detection areas of both sensors overlap within this range. Utilizing a sensor-fusion, this leads to a combination of both sources in a cooperative fusion approach.

It is obviously necessary to observe far objects continuously during lane-change maneuvers of the EGO-car, this situation occurs frequently at the beginning of an overtaking scenario, Fig. 5. The plots show that this continuity can be achieved, and there is no object-loss due to the consideration of the expected lateral shift of the observed car in the association process. This shift can be calculated out of the measurement of the yaw-movement of the EGO-car (A). The outcome shows, that

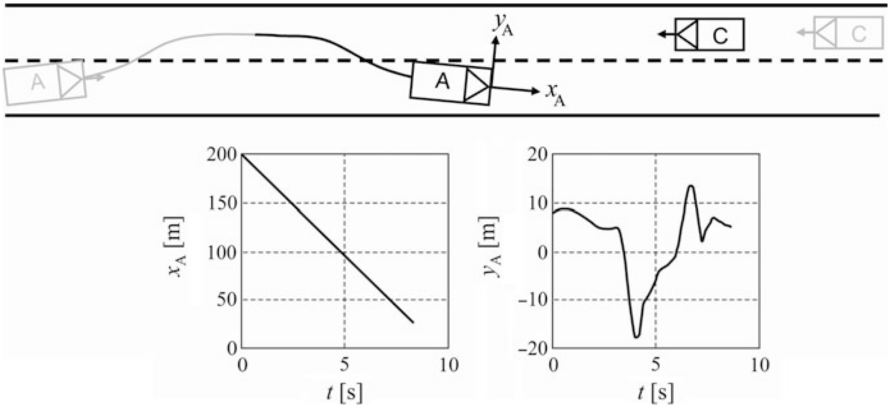


Fig. 5 Tracking of an oncoming object during a lane change maneuver; the plots show the x- and y-positions of the observed object within a coordinate system fixed on the EGO-car; no object-loss happens

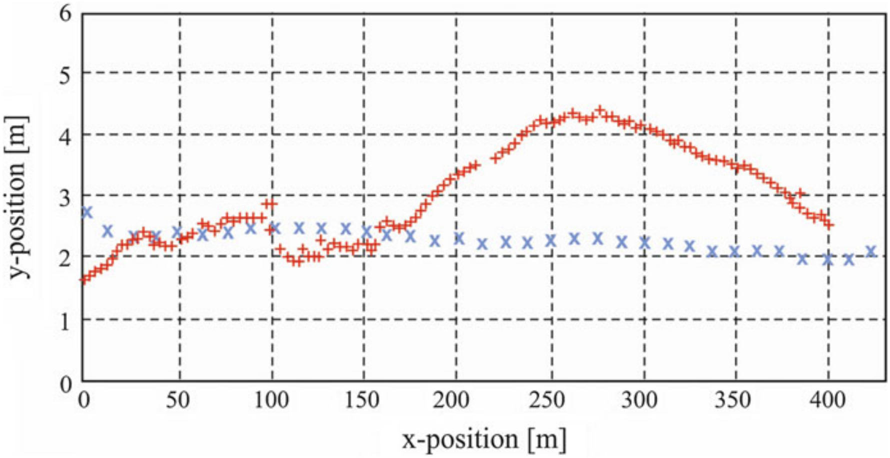


Fig. 6 Estimation of the position of an oncoming object (+) compared with corresponding ground-truth data (x); relative speed is 140 kph and $v_A = 0$; the maximum lateral deviation is on the order of 2 m

the realized environment perception system provides a valid basis for the reliable operation of algorithms based on its output. This is illustrated in Fig. 6, the plot represents the estimated position of an oncoming car.

The largest deviation in y-direction is on the order of about 2 m in a distance of about 260 m, this value corresponds to an angle-deviation even below 0.5°.

In addition to this, common variations of the position of the actual reflection point of the RADAR target also cause a range of deviations of about 1 m. A further particularity related to the detection of oncoming objects is the large value of the considered relative speed. In this area, the tracking-system could be verified experimentally up to relative speeds of -265 kph (approach), this indicates sufficient reserves for the observation of oncoming traffic on rural roads.

4 Situation Analysis for Overtaking Maneuvers

To assist the driver in dangerous overtaking situations, it is necessary to detect that an overtaking maneuver is being conducted as well as to detect that it is dangerous.

In a first step, the vehicle's position, orientation, and dynamics with respect to the road are determined in an odometry module. The state estimation is accomplished by coupling a vehicle model and a road model also in an EKF and fusion of signals from vehicle dynamics sensors and a camera based lane detection system. This allows a lane spanning ego localization and temporary breakdowns of the lane detection can be bridged. Based on the state estimates from the odometry module and environment sensor data for the leading vehicle B, longitudinal and lateral indicator variables are calculated. The maneuver detection is then accomplished in a state diagram, in which

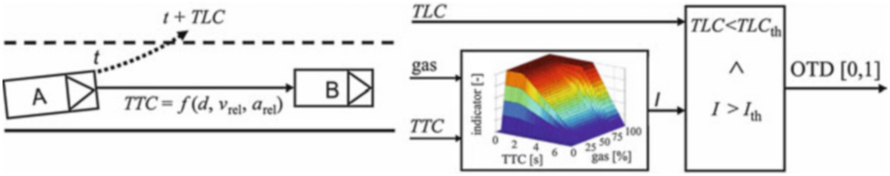


Fig. 7 Detection of overtaking maneuver based on lateral and longitudinal indicator variables

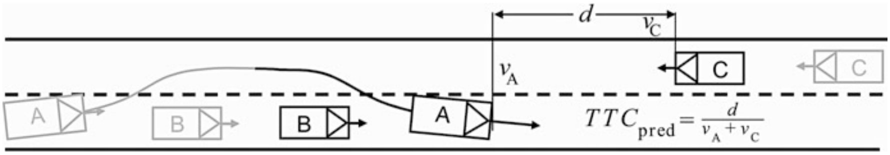


Fig. 8 Predicted time-to-collision (TTC_{pred}) with respect to oncoming traffic as a measure for safety distance d during completion of an overtaking maneuver

the maneuver transitions are modeled depending on the indicator variables. To be able to warn the driver in an early stage of critical overtaking maneuvers, additionally, an early detection of the overtaking start has been realized.

The system predicts an overtaking start, if the time-to-line-crossing (TLC) as well as a longitudinal overtaking indicator I reach corresponding threshold values, Fig. 7. (For a detailed description of the odometry and maneuver detection see Schmitt et al. (2009), Schmitt (2012)).

If an overtaking situation is detected it is continuously assessed to see whether the maneuver can be conducted or completed without a conflict with oncoming traffic. Considering the acceleration characteristics of the main vehicle, the relative kinematics of all relevant vehicles are pre-calculated. For the instant of time at which the main vehicle will have left the opposite lane after overtaking, the time-to-collision (TTC) with respect to vehicle C is predicted, Fig. 8.

The predicted TTC indicates the magnitude of the safety distance d between the main vehicle A and the oncoming vehicle C while completing the overtaking maneuver.

Based on the predicted TTC, the distance to the oncoming traffic can already be assessed when beginning the overtaking maneuver. If the predicted TTC is lower than a threshold, vehicle C is too near and the overtaking maneuver should be refrained or aborted.

5 Warnings and Automatic Emergency Braking

As soon as the module situation analysis indicates a dangerous overtaking maneuver, the system informs the driver by several warnings and starts planning a collision avoiding overtaking abort. Depending on distance and relative speed of the oncoming vehicle, the aborting maneuver has to be initialized sooner or later.

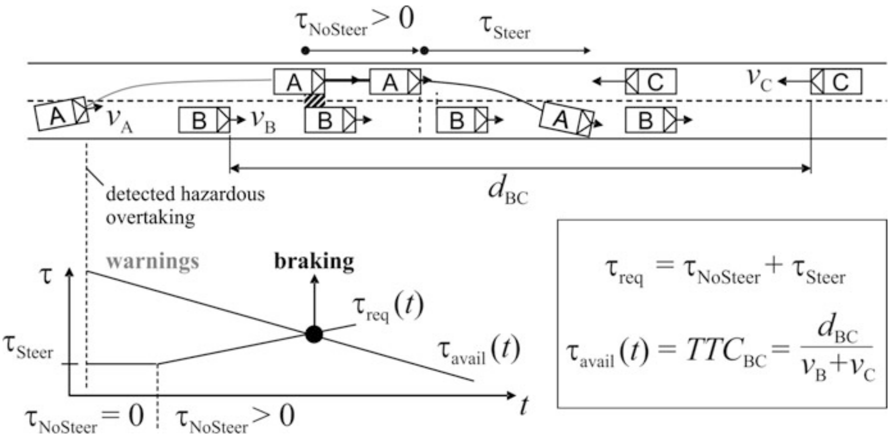


Fig. 9 Required and available time for an overtaking abort

If the vehicle has to fall behind the preceding vehicle, the system decelerates below the velocity of the preceding vehicle, but not below a minimal velocity, that allows dynamical lane changing.

Based on the vehicles' current distances and velocities, both the required and the available time for an aborting maneuver, τ_{req} and τ_{avail} respectively, are calculated and continuously updated.

The required time τ_{req} for the aborting maneuver is the estimated time until the vehicle will have left the overtaking lane again. When falling behind the preceding vehicle is needed, this time will increase. For updating the estimate after the preceding vehicle has left the detection range of the used sensors due to passing, the preceding vehicle is tracked using a model according to Schmitt and Isermann (2009), Schmitt (2012). In addition to the estimated time until a lane change is possible, a comfortable time reserve τ_{steer} for steering back to the nearside lane is included, e.g., τ_{steer} is set to 3 s.

The available time τ_{avail} is estimated by the time until the front edge of the oncoming vehicle meets the rear edge of the preceding vehicle. Both time estimates, τ_{req} and τ_{avail} are depicted in Fig. 9. The difference between τ_{req} and τ_{avail} can be used for an increasing warning level. If the difference becomes close to zero, the system initializes an automatic emergency braking until the driver can change to the nearside lane behind the preceding vehicle.

6 Results from Driving Experiments

Figure 10 illustrates a driving experiment on the test track of TU Darmstadt, simulating a dangerous overtaking maneuver. The preceding vehicle is detected in a distance of $d_{AB} \approx 30$ m and the main vehicle follows at a speed of $v_A \approx 60$ km/h. The system detects the beginning of an overtaking maneuver at $t \approx 4.3$ s. At

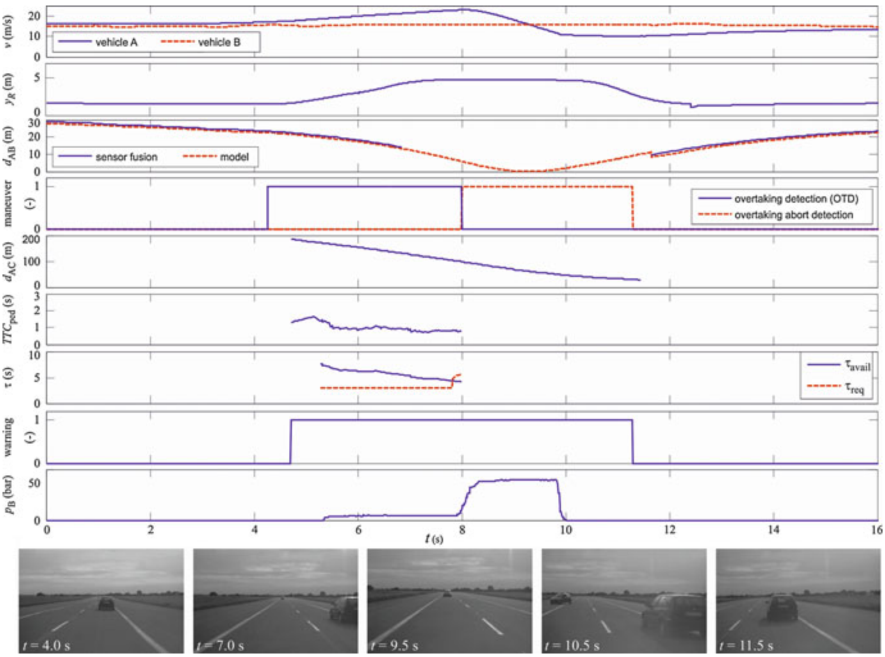


Fig. 10 Results from a driving experiment: system assists in aborting a dangerous overtaking maneuver

$t \approx 4.8$ s the oncoming traffic is no longer occluded by the preceding vehicle and detected in a distance of $d_{AC} \approx 185$ m. The system starts continuously predicting the time-to-collision at the end of the overtaking maneuver.

As the predicted TTC is beyond the corresponding warning threshold $TTC_{\min} = 2$ s, the system starts acoustic warnings, inducing the driver to abort the maneuver. The driver does not react and the system pre-fills the braking system in order to prepare a subsequent abort of the overtaking maneuver.

In the further course of the overtaking maneuver, the required time for an overtaking abort approaches the available time for collision avoidance, since the oncoming vehicle is approaching. At $t \approx 8$ s the latest possible time instant for overtaking abort is reached and the system automatically initiates an emergency braking until steering back behind the leading vehicle is possible. The acoustic warnings stop when the system detects the completion of the abort maneuver.

7 Summary

Severe traffic accidents in overtaking situations have motivated the development of a suitable driver assistance system. This contribution describes the concept and practical testing of a driver assistance system for overtaking situations.

This contribution describes the fusion of video and RADAR data, which combines a high-precision detection of far objects with accurate lateral position and velocity estimates for nearer objects. The detection of nearer objects is performed by a video based vehicle classifier. The fused environment data is the basis to identify a hazardous situation by combining a signal based detection of the driver's overtaking intention with a problematic constellation of the involved vehicles. When such a hazardous situation is detected, warnings are initiated, and an automatic brake intervention aborts the dangerous overtaking maneuver at the last possible moment, so that the driver can get behind the preceding vehicle and swerve into his/her own lane. The results of the research project Proreta 2 allow assisting drivers in timely abortions of hazardous overtaking maneuvers.

Acknowledgements This contribution results from the research cooperation Proreta between Technische Universität Darmstadt and Continental AG. The research project was carried out in cooperation of the Institutes of Automatic Control, Automotive Engineering and Multimodal Interactive Systems. The participating Institutes thank the Continental AG for the generous support and good cooperation.

References

Bender E, Darms M, Schorn M, Stählin U, Isermann R, Winner H, Landau K (2007) Anti collision system proreta – on the way to the collision avoiding vehicle. ATZ worldwide 109(4):20–23

Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: IEEE conference on computer vision and pattern recognition (CVPR), San Diego

Darms M, Winner H (2006) Validation of a baseline system architecture für sensor fusion of environment sensors. Fisita World Automotive Congress, Yokohama

Hohm A (2010) Umfeldklassification und identification von Überholzielen für ein Überholassistenzsystem. Dissertation Tu Darmstadt, Fortschritt-berichte VDI, Reihe 12, No. 727, VDI Verlag, Düsseldorf

Isermann R, Schiele B, Winner H, Hohm A, Mannale R, Schmitt K, Wojek C, Lüke S (2009) Elektronische Fahrerassistenz zur Vermeidung von Überholunfällen – Proreta 2 (Electronic Driver Assistance to avoid overtaking accidents – Proreta 2). VDI-Berichte, no 2075. Elektronik im Kraftfahrzeug, Düsseldorf

Mannale R, Hohm A, Schmitt K, Isermann R, Winner H (2008) Ansatzpunkte für ein System zur Fahrerassistenz in Überholsituationen (System approaches for Driver Assistance in overtaking situations). 3. Tagung Aktive Sicherheit durch Fahrerassistenz, Garching

Schmitt K (2012) Situationsanalyse für ein Fahrerassistenzsystem zur Vermeidung von Überholunfällen auf Landstraßen (Situationanalysis for a Driver Assistance System to avoid overtaking accidents on rural roads). Dissertation Technische Universität Darmstadt, Fortschr.-Ber. VDI Reihe 12, no 763. VDI Verlag, Düsseldorf

Schmitt K, Isermann R (2009) Vehicle state estimation in curved road coordinates for a driver assistance system for overtaking situations. In: 21st international symposium on dynamics of vehicles on roads and tracks (IAVSD), Stockholm

Schmitt K, Habenicht S, Isermann R (2009) Odometrie und Manövererkennung für ein Fahrerassistenzsystem für Überholsituationen (Odometry and maneuver detection for a Driver Assistance System for overtaking situations). 1. Automobiltechnische Kolloquium, München

Torralba A, Murphy KP, Freeman WT (2004) Sharing features: efficient boosting procedures for multiclass object detection. In: IEEE conference on computer vision and pattern recognition (CVPR), Washington, DC

Winner H, Danner B, Steinle J (2009) Adaptive cruise control. In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme (Handbook of driver assistance systems). Vieweg + Teubner, Wiesbaden

Wojek C (2010) Monocular visual scene understanding from mobile platforms. Dissertation Tu Darmstadt

Wojek C, Schiele B (2008) A dynamic conditional random field model for joint labeling of object and scene classes. In: European conference on computer vision (ECCV), Marseille

PRORETA 3: Comprehensive Driver Assistance by Safety Corridor and Cooperative Automation

58

Hermann Winner, Felix Lotz, Eric Bauer, Ulrich Konigorski, Matthias Schreier, Jürgen Adamy, Matthias Pfromm, Ralph Bruder, Stefan Lüke, and Stephan Cieler

Contents

H. Winner (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

F. Lotz
Continental AG, Frankfurt am Main, Germany
e-mail: lotz@fas-buch.de

E. Bauer • U. Konigorski
Control Systems and Mechatronics Lab, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: ebauer@rtm.tu-darmstadt.de; ukonigorski@iat.tu-darmstadt.de

M. Schreier • J. Adamy
Control Methods and Robotics Lab, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: schreier@rtr.tu-darmstadt.de; adamy@rmr.tu-darmstadt.de

M. Pfromm • R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: m.pfromm@iad.tu-darmstadt.de; bruder@iad.tu-darmstadt.de

S. Lüke
Continental Division Chassis & Safety, Advanced Engineering, Frankfurt am Main, Germany
e-mail: stefan.lueke@continental-corporation.com

S. Cieler
Interior Electronics Solutions, Continental Automotive GmbH, Babenhausen, Germany
e-mail: stephan.cieler@continental-corporation.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_19

Abstract

Instead of a multitude of single assistance functions, the PRORETA 3 concept presents just two functional assistance modes: Firstly, the so-called Safety Corridor in the case, that the vehicle guidance is carried out by the human driver and secondly, Cooperative Automation, offering partial automation of driving in cooperation with the driver. In the Safety Corridor mode, the system has to permanently monitor driving situations and assess concerned potential hazards. As a result, the driver will be informed in the first instance, then in the next stage, a warning is given, and in the last instance, an autonomous collision avoidance trajectory is generated.

Cooperative Automation is a concept of shared vehicle guidance. The execution of driving is automated; the driver interacts and supervises the execution, close to the concept of Conduct-by-Wire.

The functional architecture of PRORETA 3 integrates concepts for human guidance as well as for (full) automation. A multimodal Human–Machine Interface provides information, warnings, and action recommendations, if necessary, in order to make the PRORETA 3 Safety Corridor clear and understandable. A maneuver interface makes it possible to delegate maneuvers in the Cooperative Automation Mode to the PRORETA system. The traffic environment is represented by a Parametric Free Space (PFS) map. The Trajectory Planning uses a predictive control model applied on a risk potential field.

The research vehicle was demonstrated and tested on a test track in Griesheim, Germany. The system acceptance and driving experience were evaluated by questionnaires. The overall assessment of the Safety Corridor, the Cooperative Automation, and the entire system reflects a high acceptance of the PRORETA3 assistance concept.

1 The PRORETA 3 Approach

After 20 years of innovations a manifold of Advanced Driver Assistance Systems (ADAS) is available in the automotive market. Many assistance aspects are addressed with them. With informing and warning systems, with systems intervening soft or hard in the vehicle dynamics, and with systems taking over driving functions from the driver, a high degree of driving situations can be assisted. The price for that manifold is the increasing complexity. This can be seen in the effort for development and validation or in the complexity of use. Further assistance innovations will put additional pressure to that development. By this the question rises, whether an integral assistance concept may open an escape from this complexity trap.

That is exactly the point of research at which PRORETA 3 has started. This third cooperation project with four institutes at Technische Universität Darmstadt and research departments of Continental pursues the research strategy for accident-free driving. After common research concerning emergency braking and evasion (PRORETA 1, 2002–2006) and the avoidance of overtaking accidents (PRORETA 2, 2006–2009), the third step started 2011 with the cooperation of the Institutes of Automatic Control and Mechatronics (Control Systems and Mechatronics Lab / Control Methods and Robotics Lab), Institute of Ergonomics and Human Factors as well as the Institute of Automotive Engineering on the TU Darmstadt side and the Advanced Engineering Department of the Chassis & Safety Division as well as the Interior Division within Continental. The project was finalized with a prototype demonstration in September 2014.

Instead of a manifold of single assistance functions the PRORETA 3 concept presents just two functional assistance modes, ref. Fig. 1: A so-called Safety Corridor for the case that the vehicle guidance is carried out by the human driver and Cooperative Automation, offering a partial automation of driving in cooperation with the driver.

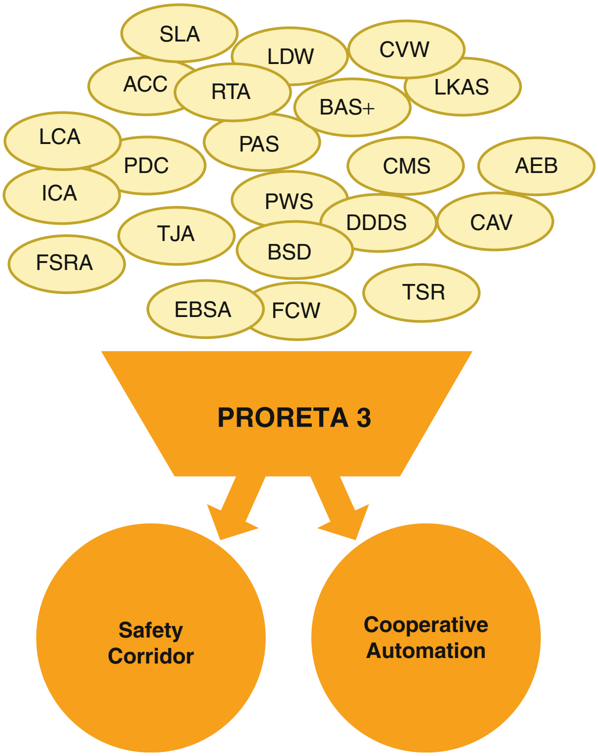


Fig. 1 PRORETA 3 – Approach: aggregation of known assistance functions (represented by typical abbreviations) to the two functional modes Safety Corridor and Cooperative Automation

1.1 Safety Corridor

This functional mode does not appear in the case of normal and safe driving. Only when a dangerous situation evolves it will support the driver to avert the danger. Nevertheless, the system has to monitor the situation permanently, and to assess concerned potential hazards. The basic criterion is the necessary maximum friction coefficient for a collision avoidance trajectory. Depending on the value of this criterion the driver will be informed in the first instance, then warning is given at the next level, and in the last instance, an autonomous collision avoidance trajectory is generated. Additionally, driver monitoring offers a differentiation of the actions depending on the state of attentiveness. By this, only necessary information or intervention will be given.

1.2 Cooperative Automation

Cooperative Automation is a concept of shared vehicle guidance. The execution of driving is automated; the driver interacts and supervises the execution (a survey concerning the different approaches to Cooperative Automation is given in ▶ Chap. 59, “Cooperative Guidance, Control, and Automation”). The variant chosen for PRORETA 3 is close to the concept of Conduct-by-Wire (CbW) (▶ Chap. 60, “Conduct-by-Wire”), even though the commands are generated by conventional operating elements, as described in more in detail in the Sect. 3. Like CbW, the core of that concept is discrete commanding of maneuvers and their execution without continuous sharing of control. With respect to the definitions of automated driving of the German Federal Highway Research Institute BASt (Gasser 2012) this follows the level of partially automated driving.

2 Functional Architecture and Behavior Planning

PRORETA 3’s architecture is based on software modules that are functionally encapsulated in order to enhance the system expandability. Two hierarchically layered modules, the behavior and trajectory planning modules form the architectural core (cf. Fig. 2). The result is a separation between (deliberate) behavioral planning and (reactive) behavioral execution, a concept that is known from behavior-based layered robot control architectures. For each of the two layers, an appropriate interface for the world model that includes environment representation and interpretation states as well as for the human–machine interface (HMI) is available. Within the coordination layer, the task of the behavior planning module is to supervise the system’s capabilities in terms of monitoring the correct functionality of all other modules. Depending on these capabilities, the driver is offered the modes Safety Corridor (SC) and/or Cooperative Automation (CA) which the driver can activate if desired.

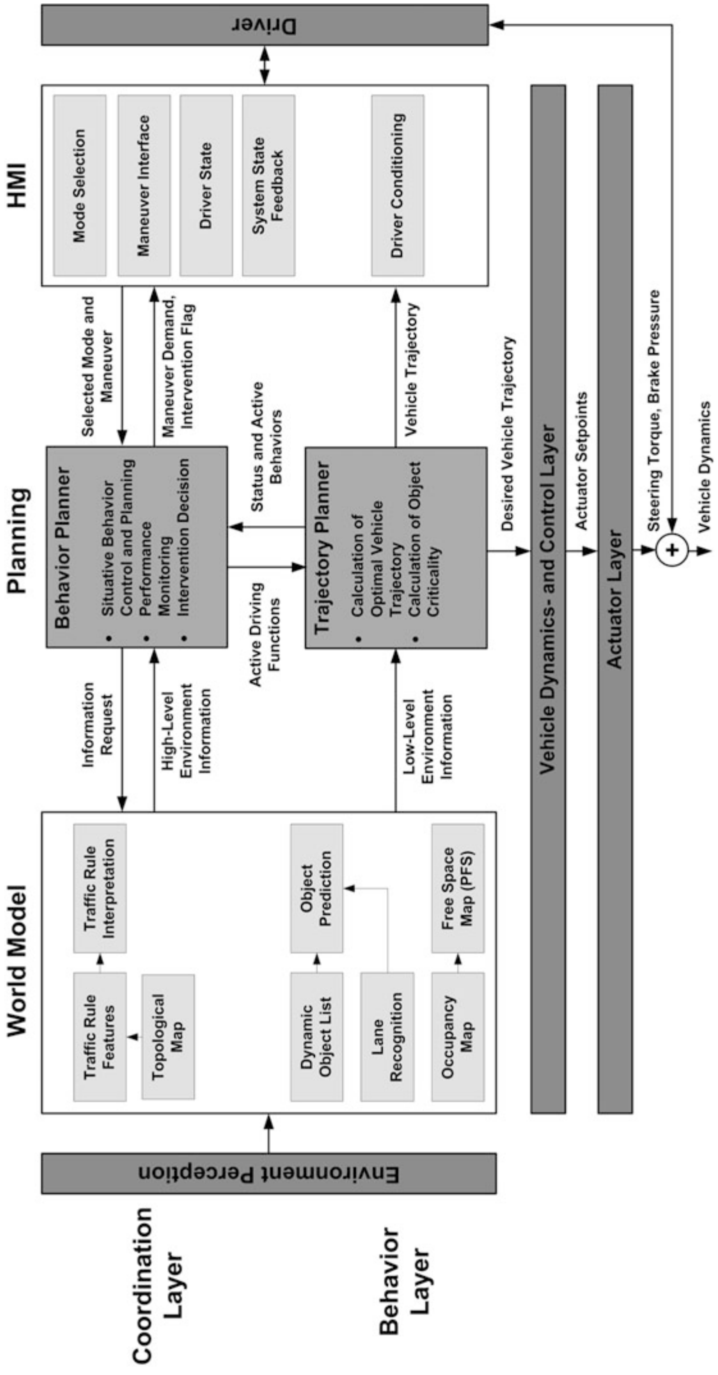


Fig. 2 Top-level view of the PRORETA 3 functional system architecture based on (Hohm et al. 2014)

Should SC be chosen, the behavior planner supervises the driver's focus of attention which is provided by HMI in the form of a driver monitoring camera. As soon as a potentially critical situation arises, the behavior planner obtains continuous information concerning the criticality level from the underlying trajectory planner as one result of the potential field-based planning process (cf. Sect. 5) and compares it to the driver's state of attention. As a consequence, the behavior planner can adapt the warning and intervention mechanism so that a distracted driver obtains assistance at an earlier point in time compared to an attentive driver. This helps in reducing situational-inappropriate interference of the system and avoiding false positives.

In CA mode, the behavior planner is responsible for a traffic rule compliant conduct of driver-requested vehicle maneuvers. Therefore, it offers the driver possible maneuvers based on the current (static) vehicle environment, e.g., lane changes when driving on highways and turning maneuvers in case of an upcoming intersection. The driver then can choose the maneuver which becomes delegated to the underlying trajectory planner in the case that the maneuver is safe to execute. The behavior planner is implemented via a rule-based finite state machine which leads to a deterministic vehicle behavior – a very important requirement since traffic rules have to be obeyed under all circumstances. More information about the implementation details and maneuver delegation strategy of the behavior planner are given in the Sect. 6 later on.

Within the coordination layer, a topological map as part of the world model provides the behavior planner with information about the static vehicle environment like the lane configuration in road segments as well as intersections, including the rule of precedence that allows to reason about the currently active traffic rules. Within the behavior layer, the world model consists of a parametric free space (PFS) map derived from an occupancy grid that describes the drivable maneuvering space in a compact way (cf. Sect. 4). Besides a lane recognition module that extracts the shape, type and position of the lane markings relative to the vehicle, an object list as well as a object prediction algorithm completes the necessary environment representation and -interpretation modules.

A more detailed discussion of the system architecture, including an outlook on how full vehicle automation can be supported within this concept, can be found in Lotz (2013), Lotz (2016) and Hohm et al. (2014). Within the next section, the human–machine interface as depicted in Fig. 2 is explained.

3 Human–Machine Interface

In order to ensure acceptance of every characteristic of assisted and automated driving it is important to always keep the driver informed about the current functional modes. Users will only trust the automation if the HMI shows clearly the change of role from an active operator to a supervisor and if it explains the distribution of tasks between the vehicle and driver. Here it is necessary to take into

Fig. 3 The safety corridor displayed on the instrument cluster

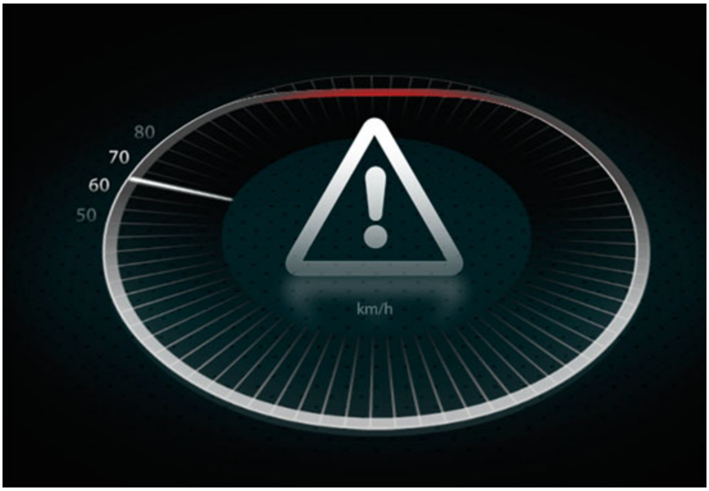


Fig. 4 Warning when approaching a critical object

account what information a driver needs, depending on the role he is currently performing.

A multimodal HMI provides information, warnings, and action recommendations, if necessary, in order to make the PRORETA 3 Safety Corridor clear and understandable. A maneuver interface makes it possible to delegate maneuvers in the Cooperative Automation Mode to the PRORETA 3 system.

- An innovative, simplified display in the instrument cluster portrays the safety corridor surrounding the vehicle and offers possible maneuvers, Figs. 3, 4, and 5.

Fig. 5 Maneuver interface in cooperative automation mode

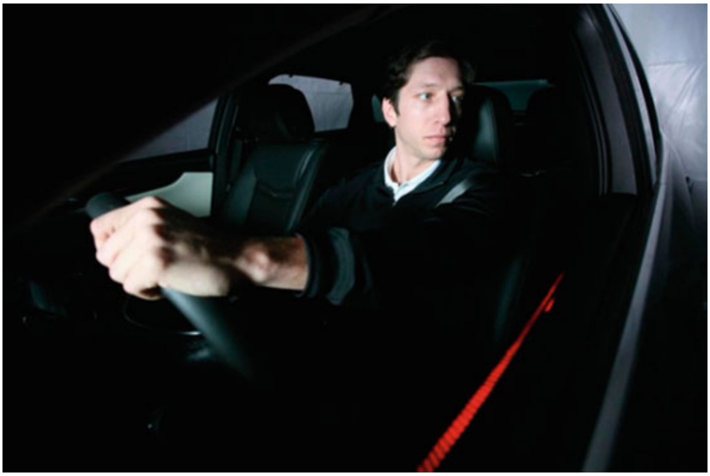


Fig. 6 LED lights attract and direct the driver's attention

- LED-strips (Pfromm et al. 2013) in the car's interior, plus directional audible warning signals, instantly direct the driver's attention to potential hazardous objects, cf. Fig. 6. An infrared camera constantly monitors the driver's viewing direction in order to avoid unnecessary and annoying attentional cueing.
- An Accelerator Force Feedback Pedal (AFFP) can also give feedback to the driver. For example, it can advise drivers discreetly but effectively to remove their foot from the accelerator pedal should the driver's visual or audible perception channel become intensively occupied and is thus capable of processing the signal very quickly.
- Maneuvers can be selected by a modified turn-signal-lever

4 Environment Representation and Sensors

An environment representation is an abstraction of the real word and has to be adapted not only to the type of environment a mobile robot, such as the PRORETA 3 vehicle, is located in but also to the tasks that need to be carried out. In principle, an environment representation should be as compact as possible and as general as needed for the given tasks. Solely *sparse* object-based models – as used in today's adaptive cruise control or emergency braking systems – are, however, not sufficient for the proposed assistance concept as they cannot represent unstructured or only slightly structured urban and rural environments. These include, for example, parked vehicles, trees, or traffic islands as obstacles that the safety corridor must take into account. Therefore, the requirement of a *dense* representation becomes obvious that is not based on predefined geometric shapes any more.

Ideally, such an environment model should be compact enough for transmission between electronic control units with limited transmission bandwidth, be sufficiently general to permit the realization of the PRORETA 3 functions, suppress irrelevant environment details to facilitate situation interpretation and planning, represent free space explicitly to permit safety-related trajectory planning, be generable online in a computationally inexpensive, robust way with all exteroceptive sensors used in the automotive domain, be able to cope with dynamic objects and to represent them adequately, and allow to incorporate sensor uncertainties.

In recent times, occupancy's grid-based dense representations (Elfes 1987) have increasingly been used in the ADAS domain. These tessellate the environment into a finite number of cells, each of which is assigned a probability of being occupied, depending on the sensor readings. Occupancy grid maps do not have to cope with the correspondence problem, can be constructed with limited computational resources by all automotive environment sensors, and are able to handle free space information explicitly. On the downside, they suffer from high bandwidth and memory requirements in their standard form, contain irrelevant details such as unreachable free space areas, suffer from discretization effects, and are not directly suitable for representing dynamic environments.

Therefore, a novel metric representation has been introduced within PRORETA 3 that copes with these deficits. It consists of a combination of dynamic object maps for the representation of dynamic entities such as other vehicles and so-called Parametric Free Space (PFS) maps for encoding relevant static parts of the environment (Schreier and Willert 2012; Schreier et al. 2013), see Figs. 7 and 8. This separation is beneficial for subsequent trajectory planning and situation analysis algorithms. The PFS map itself is a continuous 2D bird's-eye view of the local static environment around the ego vehicle that does not model the world by discrete cells but by a combination of a closed B-spline curve and geometric primitives. An important difference to other parametric maps is that not objects are described explicitly, but rather the opposite, i.e., relevant free space. Free space areas are considered relevant in this context if they are reachable by the PRORETA 3 vehicle. Consequently, areas behind guard rails or construction sites are not included in the PFS map.

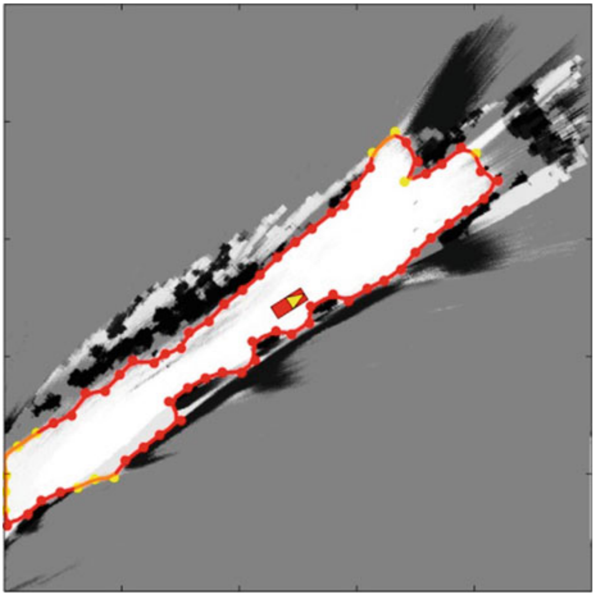


Fig. 7 A Parametric Free Space (*PFS*) map (overlaid over an occupancy grid map) describes the free space that the PRORETA vehicle can reach

Fig. 8 Parametric Free Space (*PFS*) map in an exemplary driving scenario. The map only represents relevant free space by a closed curve (*red*: obstacle boundary, *orange*: unknown environment boundary) around the ego vehicle and geometric primitives (*blue*) if enough free space has been mapped to drive around the obstacles

The complete representation is derived from a conventional occupancy grid, which has been modified to handle dynamic environments beforehand. This modification (Schreier et al. 2014a) minimizes undesirable dynamic object corruptions to generate a local, optimized grid map that only contains *static* environment structures while simultaneously detecting and tracking the corruption-causing *dynamic* objects. To this end, an ego-motion compensated temporal backward difference quotient grid is first built to separate newly free and newly occupied cells. These are subsequently clustered individually by a density-based approach, and rectangles are fitted to each cluster, which are in turn combined to reconstruct dynamic object hypothesis even in the vicinity of static environment structures by using knowledge about newly available free space. Thereafter, a nonlinear, adaptive Bayesian tracking filter in form of a combination of an Interacting Multiple Model (IMM) Filter with underlying Unscented Kalman (UK) Filters and Probabilistic Data Association (PDA) Filters – termed IMM-UK-PDAF – is used not merely to recursively estimate states of underlying motion models of extended objects but rather for checking the consistency of cell movement with motion characteristics of real objects to classify between real dynamic objects and false reconstructions. Tracks of real dynamic objects are integrated into the dynamic object map and cleared from the grid via a final flood fill procedure. The result is a dynamic object map with estimated state vectors and an optimized occupancy grid that (optimally) contains only the static driving environment.

The optimized grid constitutes the foundation for the extraction of PFS maps (Schreier and Willert 2012; Schreier et al. 2013), which can likewise be interpreted as a novel method of grid-based free space detection. Methodically, the optimized grid is treated as an image and methods well-known from the image analysis domain are employed to extract relevant free space boundaries. For this purpose, the grid is first median-filtered to get rid of noisy pixels (cells), followed by a pixel-based segmentation. Then, morphological erosion with a structuring element of the size of the vehicle’s width is applied, so that free space segments into which the vehicle does not fit are removed. Moreover, larger free space areas that are joined together only by narrow, impassable connections are separated from each other by the erosion. Afterwards, a connected components labeling is performed, and the segment in which the ego vehicle is currently located is selected and subsequently dilated with the same structuring element to bring the reachable free space back to its original size. A boundary tracing then extracts boundary cells of this final free space segment. These cells are used as measurements for a continuous, dynamic, probabilistic B-Spline free space contour tracking realized with the aid of an Information Filter. Free space holes are additionally represented by geometric primitives for an even more compact representation.

Due to the underlying grid, the complete metric representation can easily be generated by different automotive sensors and used by other existing grid-based systems. The main advantages of the PRORETA 3 environment representation are the suppression of irrelevant free space information, the compactness that allows the transmission within very limited bandwidth, the explicit free space representation vital for safety related evasive trajectory planning within the safety corridor, the robust separation between static and dynamic environment, and the sensor-independent, real-time capable generation process.

The sensor setup used to obtain the environment model is illustrated in Fig. 9. It consists of a stereo camera that additionally captures lane markings, traffic signs,

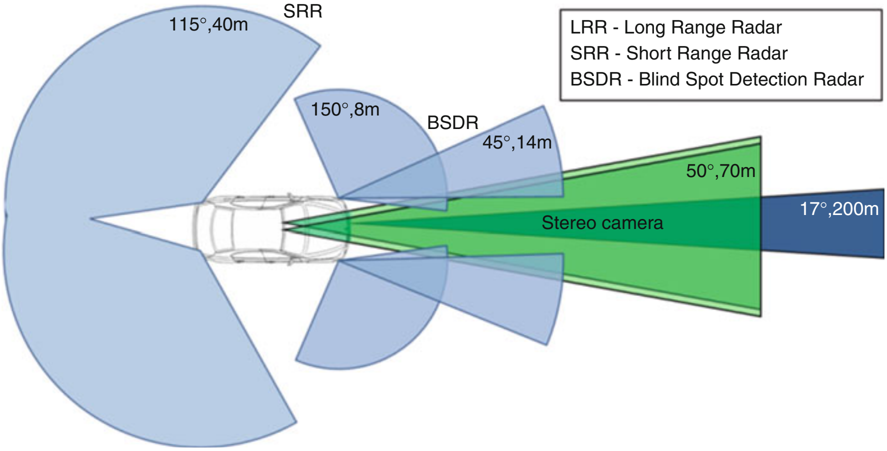


Fig. 9 Schematic diagram of built-in sensors

and traffic lights, a long-range RADAR (LRR, 77GHz) behind the front bumper as well as two short-range (SRR) and two Blind Spot Detection (BSDR) RADAR sensors (24 GHz) that cover the area alongside and behind the PRORETA vehicle. Together with a probabilistic, maneuver-based, long-term trajectory prediction for all vehicles in the traffic scene, see (Schreier et al. 2014b), the introduced environment representation provides the basis for trajectory planning described in the following. More detailed information about the environment representation and prediction can be found in Schreier (2016).

5 Trajectory Planning

The basic idea of PRORETA 3 described in the introduction consists of keeping a car permanently out of danger zones by means of longitudinal and lateral interventions. The planner needed for this purpose utilizes information obtained from the surroundings in the form of static and dynamic objects as well as markings on the road. Since it is not possible to plan trajectories based entirely on lane markings, especially in an urban environment, PRORETA 3 also utilizes a model of the free space around the vehicle. The particular surround information is mapped onto a potential-hazards field such as is seen in Fig. 10. The resulting potential field is a combination of the potential fields of the obstacles, of the road and of the PFS map and is described by a smooth two-dimensional spatial scalar field. The potential field values increase in spatial direction of the obstacles as well as of the boundaries of the road and the free space. Dynamic objects are treated by using the predicted object positions to build a time-dependent potential field of these obstacles. The so-modelled potential field is used for risk minimization.

Fig. 10 Potential field in an exemplary driving scenario

In the Cooperative Automation mode, the desired driving maneuvers (from behavior planning) vary the potential field so that this information is also implicitly contained. In case of a desired lane change maneuver the minimum of the road potential field is shifted to the desired lane. Another example is a desired turning maneuver at an intersection. The undesired arms of the intersection get closed in the potential field. A desired stopping maneuver is realized by a virtual obstacle in front of the vehicle.

The principle of model predictive control finds application here to plan a trajectory within the potential field. Hence, optimization problems are solved over finite moving prediction horizons. The minimum temporal length of the horizon depends on the sensor range and the maximum assisted velocity. In PRORETA 3, a horizon length of 4 s is chosen. The cost functional that is necessary for this purpose and that needs to be minimized contains the potential field on the one hand and an energy-optimized part on the other hand (Bauer et al. 2012). In case of the Cooperative Automation, the cost functional will be extended by some additional terms, which penalizes the deviation of a desired velocity. Furthermore, the steering wheel angle and angular velocity are limited and a nonlinear single track vehicle model is used within the optimization process. A real-time approach solves the resulting nonlinear model predictive control problem (Bauer and Konigorski 2013). The generated trajectory can be used for both Safety Corridor and Cooperative Automation. In the latter case, the trajectory for the steering wheel angle will be forwarded to the underlying control structure. In the Safety Corridor the planned trajectory is analyzed with respect to vehicle dynamic states. If some thresholds are exceeded, an emergency intervention is executed, i.e., the trajectory is forwarded to the underlying control structure. More detailed information about the trajectory planner can be found in Bauer (2016).

6 Implementation and Functional Demonstration

The functions of the research vehicle were demonstrated on a test track in Griesheim, Germany. In each of the Safety Corridor test situations, the PRORETA 3 vehicle supported the driver by intervening in the forward and lateral motion adequate to the situation. Figure 11 shows the car during automated braking in front of an obstacle.

Furthermore, with the implementation of Cooperative Automation mode in the PRORETA 3 test vehicle for the first time cooperative automation – including intersections – can be experienced in reality. In the following, a detailed description of scenarios implemented in the PRORETA 3 prototype is given. Corresponding videos can be found under www.proreta.de.

6.1 Safety Corridor Scenarios

1. **Suddenly appearing dynamic obstacle:** When the driver follows the lane and suddenly, a dynamic obstacle is shot into the way from the side, the PRORETA system performs full brake execution at the last possible moment to prevent the impending collision caused by inadequate reaction of the driver, cg. Fig. 11.
2. **Static obstacle with driver distraction detection:** If the vehicle critically approaches a static obstacle several seconds before the collision, LED-strips in the car's interior direct the driver's attention to the direction of the hazardous obstacle and audible warnings gain alertness. If the driver is distracted and does not look in the direction of the obstacle, an additional "Light Comet" is activated to guide the eye gaze. If the vehicle approaches further, the metaphoric "Safety Bubble" (cf. the circular visualization of the SC in Fig. 4) dents at the front and if

Fig. 11 The PRORETA 3 research vehicle during an emergency brake maneuver at the TUD August Euler Airfield in Griesheim, Germany

driver reactions are still absent, the AFFP is pressed against the driver's accelerator pedal foot and a strong emergency braking maneuver is activated.

3. **Speed adaption in road bends:** In case a driver is approaching a road bend with dangerously high speed, he/she is warned again by red LED-stripes, an acoustic signal and the AFFP. The vehicle speed is additionally reduced smoothly so that the driver can drive safely through the road bend.
4. **Collision avoidance in construction sites:** If the driver steers against construction walls, (cf. Fig. 3). the Safety Bubble dents at the corresponding side. An evasive steering torque is applied to prevent the collision – the driver can, however, always overrule the system if he/she feels the need to do so.
5. **Prevention of unintentional lane departure:** If the driver is about to leave the lane illegitimately without activating the indicator, the PRORETA system smoothly directs the vehicle back toward the center of the lane. The vehicle is directed back into the orientation of the lane which is expected to be easily controllable than a "simple" reactive additional steering torque.
6. **Prevention of wrong-way driving:** If the driver accidently drives into a one-way street the wrong way, he/she is warned by the LED-stripes, an acoustic signal, AFFP and an additional visualization of the corresponding wrong-way icon is shown in the instrument cluster. In case of no driver reaction, a smooth braking to standstill is triggered.
7. **Ignoring red traffic lights:** A similar warning and intervention strategy is applied in case of an ignorance of red traffic lights (cf. Fig. 12). The difference lies in the rigorousness of the intervention, which is much greater in this case – similar to emergency braking with respect to real obstacles.

Fig. 12 Testing of hazardous situations: ignoring a *red light*

6.2 Cooperative Automation Scenarios

In the Cooperative Automation-mode the vehicle is permanently centered in the middle of the lane. The controlled velocity is the minimum of speed limits indicated by traffic signs, road bends and the driver's desired velocity, which can be set via the PRORETA operating lever.

1. **Lane change maneuver on a two-lane road:** If a lane change maneuver is possible, the driver is informed by the freely programmable instrument cluster (FPC) and has the possibility to select the maneuver by pushing the indicator lever up or down, just like activating the right or left indicator in a conventional vehicle. Immediately after that, the vehicle performs an automated lane change in the desired direction.
2. **Lane merging scenario:** In this scenario the left lane is merged into the right lane. If the vehicle is approaching the lane merge on the left lane and the driver doesn't select a lane change by him/herself, the vehicle decelerates smoothly to a standstill. Afterwards, the driver can select a lane change maneuver in order to let the vehicle accelerate again to the desired velocity and to perform the lane change.
3. **Intersection with traffic signs (turn right):** If the vehicle approaches an intersection, the driver is informed about the possible turning directions via the FPC. He/she can then choose the desired turning maneuver, e.g., the "turn right" maneuver by pushing the maneuver lever up, just like activating the right indicator in a conventional vehicle. The vehicle subsequently performs the maneuver automatically. In case that the driver doesn't choose any maneuver, a "default maneuver" is activated which is individually determined by the intersection geometry and the specific rule of precedence. The PRORETA prototype vehicle has shown the capability to handle two different intersections regulated by traffic signs.
4. **Automated driving through construction site:** In analogy to the Safety Corridor scenario number 4, the automated system is able to handle construction sites and narrow road sections even without lane markings due to its flexible freespace detection and trajectory planning approach. This was demonstrated within a 150 m long stretch of road with boundaries to the left and right (cf. Fig. 3).
5. **Intersection with traffic lights:** Besides intersections with traffic signs, the system is also capable to successfully cover intersections controlled by traffic lights. Depending on the detected traffic light state and the desired driving maneuver, the PRORETA prototype is able to stop at a red traffic light and to continue driving after the signal switches back to green. Within this process, the driver can change the delegated maneuver as long as the vehicle is at a standstill.

Within the conceptualization of the Cooperative Automation mode, the functional philosophy had to be defined. The first question was when to offer possible maneuvers (e.g., a lane change) to the driver with the options "as soon as the static environment allows it" or "as soon as the static and dynamic environment allows it." As an example for the second option, in the case that the PRORETA car drove on the right lane of a busy two-lane highway (per direction), the system would offer the driver the lane change as soon as it detects a suitable gap that the vehicle can change into. As a result, the maneuver "lane change left" would be constantly offered and taken away again, which could annoy the driver and thus result in an acceptance problem. Offering the driver the maneuver as soon there is a suitable lane and delay the execution of the maneuver in the case that the target lane is blocked smoothens the delegation process and encourages driver confidence of the system. Therefore, we chose this option.

Another important question that had to be resolved was how the system behaves at intersections. One option is that the vehicle stops at every intersection and waits until the driver delegates a maneuver. This, however, results in possible unnecessary decelerations that also compromises following traffic participants and also results in ongoing maneuver inputs for the driver, even in case he/she just wants to continue travelling on the major road. Therefore, we decided that the default maneuver (which is performed if the driver does not select a specific maneuver) is always the maneuver that follows the current (major) road and does not result in turning into an inferior road segment (which most of the times means to go straight). This strategy, however, does not resolve all situations, e.g., when approaching a major road in a T-shaped intersection where there is no preferable turning option. In this case, the vehicle stops and waits for driver delegation.

7 System Acceptance and Driving Experience

A car clinic study was carried out to assess some basic usability and acceptance criteria of the PRORETA 3 assistance concept. Besides a summarizing judgment of the two operation modes Safety Corridor and Cooperative Automation the participants were asked to rate the ergonomic quality of the PRORETA 3 concept by means of scales addressing the *perceived workload*, the *perceived reliability*, the *perceived safety*, the *driving comfort,* and the *comprehensibility* of the PRORETA 3 functions. As the principles of spatial mapping and attentional cueing are central for the PRORETA 3 Safety Corridor and its HMI, special interest was given to the supportive qualities of the LED stripe and spatial sound effects.

7.1 Method and Sample Description

The study took place on the abovementioned test track over a 2-day period in September 2014. Altogether 84 people participated in the demonstration event, 68 of these were experts from the automotive industry and 16 persons from the field of research and media (The analysis of the data showed that both groups (Automotive and non-Automotive) generate similar results. Therefore the data of all subjects were pooled.). Two test drives had to be performed by the test persons. The first lap focused on the Safety Corridor the second on the mode Cooperative Automation. The specific scenarios that were experienced are described in the previous chapter. A short briefing on both system modes and the HMI principles of PRORETA 3 was undertaken just before the test drives. During both laps a test instructor was sitting on the front passenger seat. Each lap on the test course took approx. 10 min. After the test drive the test persons were asked to fill out a questionnaire including rating scales and open-ended questions. This feedback was submitted anonymously.

7.2 Results

The overall assessment of the Safety Corridor, the Cooperative Automation and the entire system reflects a high acceptance of the PRORETA3 assistance concept. The mean values on a 6-point grading scale (1 = excellent; 6 = fail) amount to 1.92 for the Safety Corridor (s = 0.93), 1.87 for the Cooperative Automation (s = 0.99), and 1.74 for the entire PRORETA 3 system (s = 0.86). The results do not reveal significant differences for the both operation modes and the entire system respectively.

In order to set up a usability profile for the both system functions a short semantic differential was used, cf. Fig. 13. The calculated mean values are based on a 6-point scale ranging from 1 (totally disagree) to 6 (totally agree). We can see that both operation modes come along with a good assessment of the five usability criteria.

The average values of scale 1 ("eases the burden on the driver") and scale 4 ("increases the driving comfort") vary at a significant level between the two

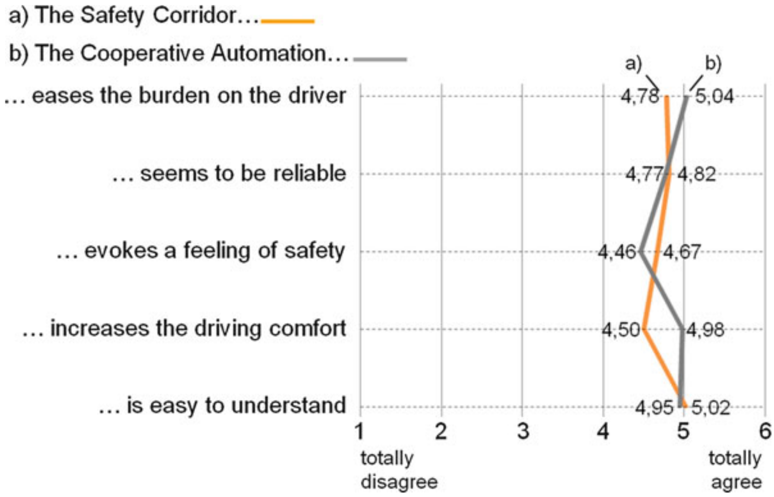


Fig. 13 Results of the subjective assessment of the PRORETA 3 "Safety Corridor" and "Cooperative Automation" mode. The scale ranges from 1 = totally disagree to 6 = totally agree

PRORETA 3 operation modes ($t_{(81)} = -2.05$, $p < 5$ % for scale 1 and $t_{(81)} = -3.32$, $p < 1$ % for scale 4). It is obvious that aspects regarding the ease of driving are much more related to the Cooperative Automation mode than the Safety Corridor.

A 6-point scale ranging from 1 = *no assistance* at all to 6 = *strong assistance* was also used to assess the perceived efficiency of the LED stripe and the spatial sound output in the Safety Corridor mode. Here, the sound device comes to an average of 4.28 (s = 1.18) compared to 4.65 (s = 1.09) for the LED stripe indicating that the visual device is more supportive from users' point of view ($t_{(78)} = 3.1$, $p < 1$ %). To which extent this result corresponds with objective measures will be in the focus of future experimental driving studies.

The feedback to the open-ended questions disclosed that the system is perceived as impressive and elaborate by most of the test persons. The system performance was described as "mature" and many subjects judged the HMI as "awesome" and "intuitive." The few negative statements concern among other things the braking behavior of the vehicle ("too late," "too abrupt"), the maneuver delegation concept ("annoying"), the Safety Bubble in the cluster instrument ("not intuitive"), and the LED stripe ("not easy to understand"). Nevertheless, eye-tracking measures gave evidence to the efficiency and good comprehensibility of the PRORETA 3 LED stripe concept (Pfromm et al. 2013).

A further evaluation of the PRORETA 3 system will be published in Pfromm (2016).

7.3 Summary of the Study

Altogether it can be summarized that the PRORETA 3 system performed well in the Griesheim car clinic. Both operation modes, the Safety Corridor as well as the Cooperative Automation, come along with good judgment regarding acceptance and usability criteria. According to the participants, the LED stripe is more efficient than the spatial sound output to become informed and warned about obstacles. As expected, the Cooperative Automation mode is associated with less workload and an increased driving comfort compared to the PRORETA 3 Safety Corridor.

8 Summary and Outlook

In PRORETA 3, a concept was developed that is able to handle an increasing number of assistance and automation functions in a compact and intuitive way. Different functions and functional modules are clustered into two different modes, the Safety Corridor for driver assistance and the Cooperative Automation that enables the driver to delegate driving maneuvers. Therefore, a modular structured system architecture combines one sensor setup, one environment representation and one control approach to cover many different functions and in order to enhance the system expandability.

The environment representation consists of a compact parametric free space map (PFS), containing a closed B-spline curve and geometric primitives, and a dynamic object map – both derived from a common underlying occupancy grid. Utilizing this information, a potential field is derived based on which a safe vehicle trajectory can be determined via a cost function optimization. The resulting trajectory is selectively given to the underlying vehicle dynamics controller in the case of the safety corridor mode or continuously in the case of the cooperative automation mode.

Part of the concept is a multimodal HMI using a freely programmable cluster instrument, an LED stripe, an acceleration pedal with force feedback as well as three-dimensional sound effects. It provides situational-appropriate information about the activated mode, warnings about possible collisions and action recommendations.

A test vehicle was equipped with the system concept. On a test track the vehicle successfully handled seven hazardous situations in the Safety Corridor mode and five situations in the Cooperative Automation mode.

These tests gave feedback regarding different system options. Especially the delegation process in the cooperative automation mode was insightful regarding the timing to provide the maneuver selection to the driver. In addition, the vehicle behavior in different situations could be tested and optimized, such as adding a traffic sign "fallback"– solution to shut down traffic lights.

In a car clinic study, participants were given the opportunity to judge the two operation modes and rate the ergonomic quality of the PRORETA 3 concept by means of scales addressing the perceived workload, the perceived reliability, the perceived safety, the driving comfort, and the comprehensibility of the PRORETA 3 functions.

The overall evaluation of the entire system with both modes reflects a high acceptance and usability. According to the participants, the LED stripe is more efficient than the spatial sound output to become informed and warned about obstacles. As expected, the Cooperative Automation mode is associated with less workload and increased driving comfort compared to the PRORETA 3 Safety Corridor.

Regarding the results of the tests and the car clinic study, the modular system architecture concept, function clustering and the Cooperative Automation provide possible answers to reduce development costs and driver workload. To verify these concepts and to bring them into products, further development steps are necessary.

Remark This chapter is a revised and extended version of (Winner et al. 2015).

References

Bauer E (2016) Modellprädiktive integrale Trajektorienplanung für Fahrerassistenzsysteme (Model predictive integral trajectory planning for driver assistance systems). PhD thesis, Technische Universität Darmstadt (manuscript in preparation)

Bauer E, Konigorski U (2013) Ein modellprädiktiver Querplanungsansatz zur Kollisionsvermeidung (Model predictive lateral planning approach to collision avoidance). In: 6. VDI Fachtagung AUTOREG 2013, Steuerung und Regelung von Fahrzeugen und Motoren, Baden-Baden, 5–6 June

Bauer E, Lotz F, Pfromm M, Schreier M et al (2012) PRORETA 3: an integrated approach to collision avoidance and vehicle automation. Automatisierungstechnik 60(12):755–765

Elfes A (1987) Sonar-based real-world mapping and navigation. J Robot Autom 3(3):249–265

Gasser TM (2012) Legal issues of driver assistance systems and autonomous driving. In: Eskandarian A(ed) Handbook of intelligent vehicles. Springer, London

Hohm A, Lotz F, Fochler O, Lueke S et al (2014) Automated driving in real traffic: from current technical approaches towards architectural perspectives. SAE technical paper 2014-01-0159. doi:10.4271/2014-01-0159

Lotz F (2013) System architectures for automated vehicle guidance concepts. In: Maurer M, Winner H (eds) Automotive systems engineering. Springer, Berlin

Lotz F (2016) Entwicklung einer Referenzarchitektur für die assistierte und automatisierte Fahrzeugführung mit Fahrereinbindung (Development of a reference architecture for assisted and automated driving with driver integration). PhD thesis, Technische Universität Darmstadt (manuscript in preparation)

Pfromm M (2016) Menschbezogener Umgang mit Systemstörungen bei manöverbasierter Fahrzeugführung (Human behavior and disfunctions of maneuver-based vehicle guidance systems). PhD thesis, Technische Universität Darmstadt (manuscript in preparation)

Pfromm M, Cieler S, Bruder R (2013) Driver assistance via optical information with spatial reference. In: Proceedings of the IEEE international conference on intelligent transport systems, The Hague, 6–9 Oct

Schreier M (2016) Bayesian environment representation, prediction, and criticality assessment for driver assistance systems. PhD thesis, Technische Universität Darmstadt (manuscript in preparation)

Schreier M, Willert V (2012) Robust free space detection in occupancy grid maps by methods of image analysis and dynamic B-spline contour tracking. In: Proceedings of the IEEE international conference on intelligent transportation systems, Anchorage, 16–19 Sept

Schreier M, Willert V, Adamy J (2013) From grid maps to parametric free space maps – a highly compact, generic environment representation for ADAS. In: Proceedings of the IEEE intelligent vehicles symposium, Gold Coast, 23–26 June

Schreier M, Willert V, Adamy J (2014a) Grid mapping in dynamic road environments: classification of dynamic cell hypothesis via tracking. In: Proceedings of the IEEE international conference on robotics and automation, Hong Kong, 31 May–7 June

Schreier M, Willert V, Adamy J (2014b) Bayesian, maneuver-based, long-term trajectory prediction and criticality assessment for driver assistance systems. In: Proceedings of the IEEE international conference on intelligent transportation systems, Qingdao, 8–11 Oct

Winner et al. (2015) PRORETA 3 – an integrated ADAS concept – comprehensive driver assistance by safety corridor and cooperative automation. In: First international ATZ-conference on driver assistance systems, Frankfurt a. M., Apr 2015

Cooperative Guidance, Control, and Automation

59

Frank Flemisch, Hermann Winner, Ralph Bruder, and
Klaus Bengler

Contents

F. Flemisch (✉)
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany

Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany
e-mail: f.flemisch@iaw.rwth-aachen.de; frank.flemisch@fkie.fraunhofer.de

H. Winner
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

R. Bruder
Institute of Ergonomics and Human Factors, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: bruder@iad.tu-darmstadt.de

K. Bengler
Fakultät für Maschinenwesen, Technische Universität München, Garching, Germany
e-mail: bengler@lfe.mw.tum.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_58

Abstract

The technological feasibility of more and more assistant systems and automation in vehicles leads to the necessity of a better integration and cooperation with the driver and with other traffic participants. This chapter describes an integrated cooperative guidance of vehicles including assisted, partially automated, and highly automated modes. Starting with the basic concepts and philosophy, the design space, parallel and serial aspects, the connections between abilities, authority, autonomy, control, and responsibility, vertical versus horizontal and centralized versus decentralized cooperation are discussed, before two follow-on chapters of H-Mode and Conduct-by-Wire describe instantiations of cooperative guidance and control.

1 Introduction

Driver assistance systems have been undergoing fascinating developments since making an initial appearance in the 1970s and 1980s, and the process is far from over. On the one hand, individual systems like ACC, LDWS, parking assistant technology, and others are already in mass production, supporting vehicle guidance by assisting with longitudinal or lateral control. On the other hand, there have been developments leading to increasingly autonomous driving capabilities, suggesting the possibility of fully automated driving without driver intervention. Another development is moving towards connected driving, whereby autonomous vehicles share data with one another and with infrastructure and cooperate with one another.

This trend towards more complex assistance and automation raises many questions:

- How could or should the trend progress?
- How can diverse individual systems be coherently incorporated as integrated wraparound systems?
- How can autonomous capabilities be used without stumbling on their limits and risks?
- How can human beings, who have been responsible for vehicle guidance thus far, remain intelligently integrated, with all of their limitations and capabilities?
- How can humans and assistance or automation systems work together effectively?
- How do we avoid paternalism and ensure sufficient driver autonomy and choice while offering greater usability, data protection, and an enjoyable driving experience?
- How do we ensure that this technology, which will initially have its constraints, can be developed into more powerful transport systems with controllable risks?

A series of DFG, EU, and industrial projects transcending institutional boundaries have led to a consistent picture of integrated assistance and automation that this chapter describes as integrated cooperative guidance of assisted, partially automated, and highly automated vehicles or cooperative vehicle guidance and control for short. Assisted, partially automated, and highly automated refers to the fact that the autonomous capabilities of assistance and automation systems are not used solely for fully automated driverless applications but in integrated cross-

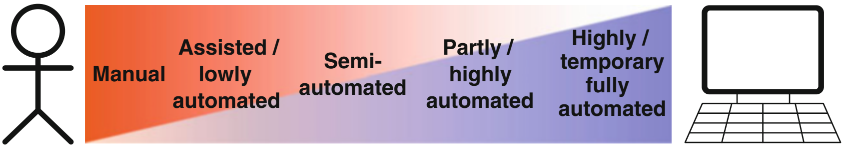


Fig. 1 Distribution of control between driver and automated system, simplified to one-dimensional cooperative vehicle guidance design space (Adapted from Flemisch et al. 2003; Hoeger et al. 2011; Gasser et al. 2012)

coordinated degrees of assistance and automation as described in (Gasser et al. 2012) (see also ▶ Chap. 3, “Framework Conditions for the Development of Driver Assistance Systems”). In such cases, drivers can choose to do the driving themselves with support from assistance systems or to permit the partially automated system to do more of the driving, for instance, by allowing maneuvers that the driver can still be sufficiently involved in and maintain control of. This approach also contains future migration stages in which humans can disengage themselves from guiding the vehicle for specific time periods or distances and allow the highly automated system to drive.

The word “integrated” refers to the fact that the driver perceives and uses the various individual assistance and automation systems with coherent degrees of assistance and automation as an integrated whole. Figure 1 breaks the control design space down into a greatly simplified design space where discrete degrees (modes) of assistance and automation are defined on a one-dimensional scale for distributing control. The assistance and automation stages sketched in Gasser et al. (2012) are an example of this. They range from a manual mode, in which the human driver has complete command over vehicle guidance and control, to a highly automated or temporary fully automated zone in which only the vehicle is in command for a time.

“Cooperative” refers to the most important quality of this transport system, namely, that the automated system does not work autonomously but rather primarily in cooperation with a human being, in this case the driver. This development can use multi-vehicle cooperation, such as vehicle-to-vehicle communication, but it goes beyond mere technology as it cooperatively incorporates humans and automation.

This overview chapter will sketch the basic concepts and basic philosophy of cooperative vehicle guidance and control. The two chapters that follow will go into the specifics of cooperative vehicle guidance (see ▶ Chaps. 60, “Conduct-by-Wire,” and ▶ 61, “H-Mode”).

2 Cooperation and Vehicle Guidance and Control

“Cooperation” is derived from the Latin words “co” (together) and “operatio” (work, activity) and is generally understood to mean “working together” (Duden 2014a) or “the action or process of working together towards common goals”

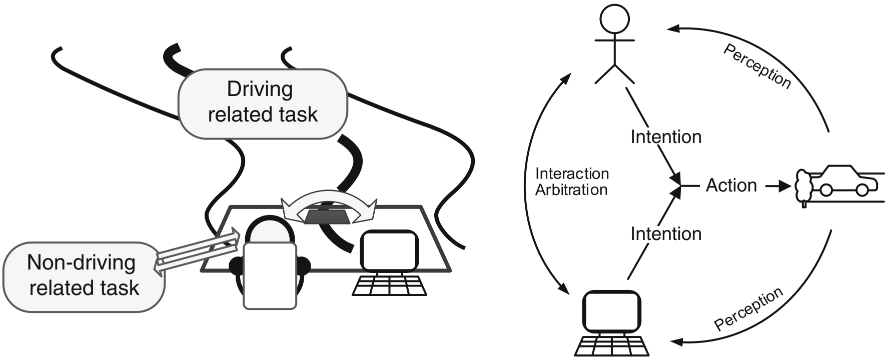


Fig. 2 *Left*: Prototypical switch between task types as a part of cooperation. *Right*: Cooperative vehicle guidance as integrally interconnected guidance and control loops, based on Flemisch et al. (2012)

(Oxford Dictionaries 2014a). Cooperative guidance and control (of vehicles) is understood in this chapter to mean cooperation between a human being and at least one computer when guiding one or more vehicles. It involves both the human and the automated system forming intentions based on their perception and then implementing those intentions in cooperative action (see Fig. 2). Cooperative control and guidance includes cases in which a human and a computer affect the same control path, which is also known as "shared control" (Griffiths and Gillespie 2004; Mulder et al. 2012) or "shared authority" (Inagaki 2008; Flemisch et al. 2011). But cooperative control also includes the option of fully or partially delegating tasks to different agents, as has been sketched in Rasmussen (1983) and elsewhere. In addition, cooperative control can comprise aspects of adaptivity and adaptability, as described in Sheridan and Parasuraman (2006) as "adaptive automation." The use of the word "cooperation" in the context of human-machine cooperation has already been sketched by Hollnagel and Woods (1983), Rasmussen (1983), Sheridan (2002), generalized into a framework for human-machine cooperation by Hoc (2000), for example, and applied to vehicle guidance by Flemisch et al. (2003), Hoc et al. (2006), Holzmann (2007), Biester (2008), Flemisch et al. (2008), and Hakuli et al. (2009), among others. Additional examples of cooperative and shared control are also described in Mulder et al. (2012).

3 Cooperative Guidance and Control as a Complex Concept or Cluster Concept

It is helpful to understand cooperative vehicle guidance as a cluster concept rather than as a clearly defined term. Cluster concepts go back to Ludwig Wittgenstein's fundamental critique of ostensive definition and describe a concept based on a list and description of attributes associated with it (see, e.g., Swartz 1997; Gottschalk-Mazouz 2007). In the following text, cooperative vehicle guidance is understood as comprising the following elements, which while not absolutely essential are nevertheless beneficial to cooperativeness in vehicle guidance:

- Autonomous guidance capabilities, both on the part of the machine and the human being
- Intuitive interaction with adequate external compatibility, i.e., satisfactory matching of external interfaces between human and machine
- Internal compatibility between human and machine, i.e., satisfactory matching of internal (typically cognitive) subsystems of human and machine, especially:
  - Compatible representation of movement through space
  - Compatible (but not necessarily explicit) goal and value systems
- Mutually comprehensible abilities and intentions
- Clear and potentially dynamic division of control
- Conflict prevention or arbitration
- Adaptivity and adaptability on the part of the machine for a good balance of stability and agility of the overall system

4 Design Space for Cooperative Vehicle Guidance and Control

Cooperative activities can be differentiated according to levels, such as the action level, planning level, and metalevel (Hoc 2001; Pacaux-Lemoine and Debernard 2007). If one begins by looking at the action and planning level, control can take place at different levels of the movement task. For this purpose, we took the three levels of navigation, guidance, and stabilization as described in Donges (1982) with further suggestions by Bernotat (1970), Sheridan (1976), Parasuraman et al. (2000) and developed a shared generic model of vehicle guidance with four levels (Fig. 3). Earlier approaches to cooperative vehicle guidance (Winner and Hakuli 2006; Flemisch et al. 2006) had already shown the benefit of further dividing the guidance level into maneuver guidance and trajectory guidance. Maneuver (executed motion, tactical movement (Duden 2014b)) is understood here, much as in Oxford Dictionaries (2014b) as a spatially and temporally connected pattern of moving a vehicle in relation to the environment. One example of a driving maneuver is changing lanes. The number of possible maneuvers for a driving mission is typically small compared to the many ways of representing a maneuver, such as a trajectory, that is, a vector of the place and time of a potential or real movement of a selected point on a moving object, such as the center of gravity.

Assuming adequate capabilities on the part of the human and the computer, cooperation can take place between driver and automation at all levels of the movement task. The distribution of roles in this cooperation can be static or it can dynamically change across the different layers. The control loops across the four levels influence each other but also differ in their temporal characteristics. Typically, the action frequency moves from comparatively low-frequency navigation to comparatively high-frequency control.

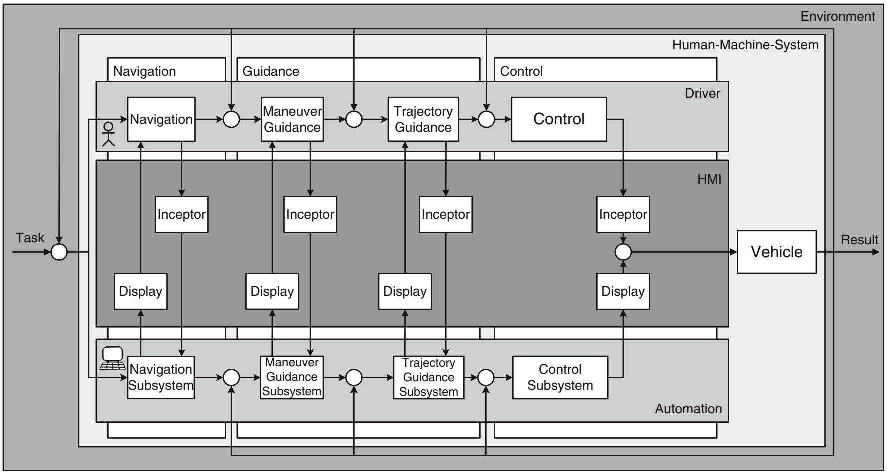


Fig. 3 Generic control flow diagram for cooperative vehicle guidance, based on Flemisch et al. (2014)

5 Parallel and Serial Aspects of Cooperative Vehicle Guidance and Control

Cooperative vehicle guidance and control can encompass and combine different forms of control flows and distributions across the levels of vehicle guidance. The most important characteristics for understanding this are serialism versus parallelism (see Fig. 4). The cooperating duo of human and machine can act serially, i.e., in sequence, such as when the human gives the machine a command that the machine then executes. One example is giving a maneuver command via a separate maneuver interface, realized, for example, in the conduct-by-wire concept (see ▶ Chap. 60, "Conduct-by-Wire"). The cooperating duo can also act in parallel, such as when both make contributions to the same task. An example is cooperation at the control level in H-Mode (see ▶ Chap. 61, "H-Mode"), in which both human and machine act simultaneously but to different degrees on a haptic operating element, such as an active steering wheel or an active sidestick. Serial and parallel aspects can also be combined, for example, in H-mode, which can be used to sequentially command a maneuver by making a motion on the operating element. The driver can then "accompany" and influence the maneuver as the automated system is acting. Aspects of serialism and parallelism can have a decisive influence on the reliability and robustness or resilience of the overall system, such as when subsystems fail.

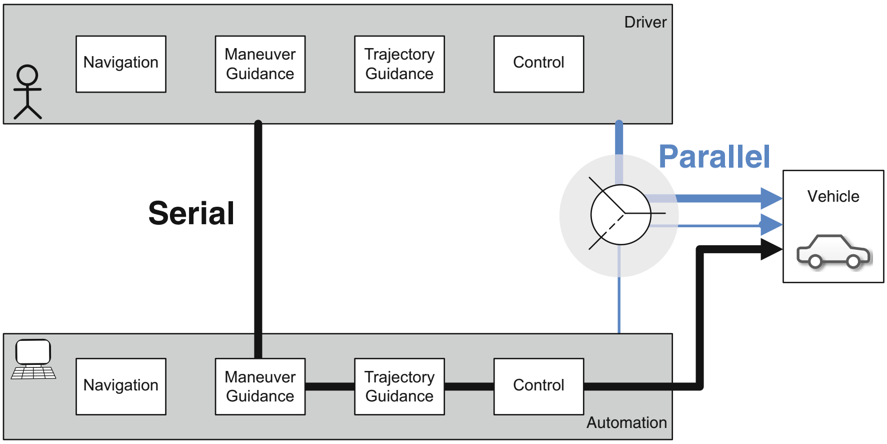


Fig. 4 Serial versus parallel vehicle guidance

6 Connections Between Abilities, Authority, Autonomy, Control, and Responsibility in Cooperative Vehicle Guidance

The cooperation dynamic is critically shaped by the duo's ability to influence the situation and to cooperate, by the authority and degree of autonomy given to one partner or the other by an outside party during the development phase or by the other partner, by the control over the determining factors of the situation, and by the responsibilities that are subsequently assessed and demanded (see Fig. 5).

This results in complex connections between abilities, authority, autonomy, control, and responsibility, such as that the authority granted should not be greater than the respective abilities, that control is only possible with adequate abilities and a minimum of autonomy and only sensible with a minimum of authority, and that the driver or the automated system should only to be held responsible if there was a minimum of abilities, autonomy, and therefore potential control over a situation. These dependencies are dynamic, so, for example, the driver can authorize control for the automated system, although this is only sensible when the automated system has the ability to control that particular situation. When abilities are no longer adequate, control should be returned to the other cooperating partner. An important factor here is situational awareness as regards the abilities of each partner and the division of authority and control, which can be drastically improved through appropriate human-machine interaction.

An overview of these connections can be found in Flemisch et al. (2011), while links to the legal situation are sketched in Gasser et al. (2012).

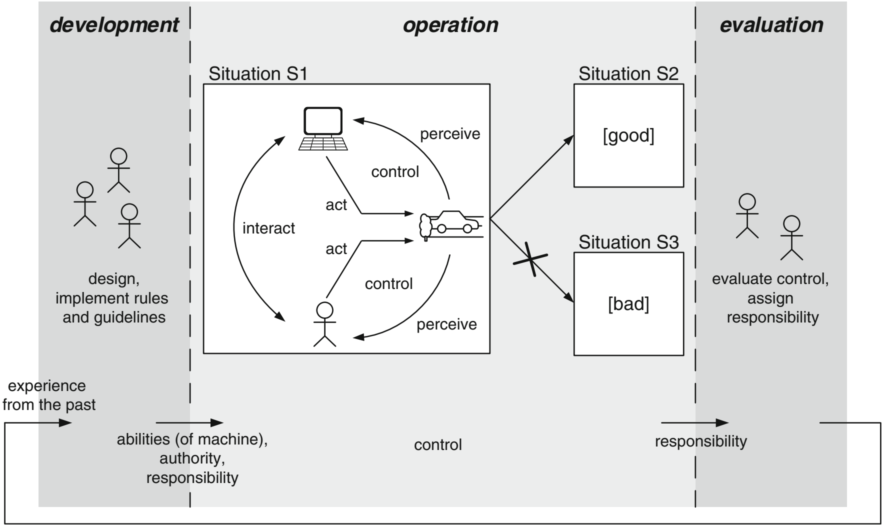


Fig. 5 Cooperation in abilities, authority, autonomy, control, and responsibility in the life cycle of a cooperative vehicle guidance system, based on Flemisch et al. (2011)

7 Outlook: Vertical and Horizontal, Centralized and Decentralized Aspects of Cooperative Vehicle Guidance

In addition to the type of cooperation between driver and vehicle explained here, actual street traffic also needs to consider cooperation between traffic participants (see Fig. 6), without which vehicle guidance would be incomplete. To distinguish the two types, driver-vehicle cooperation or driver-automation cooperation, respectively, is referred to as vertical cooperation while cooperation between traffic participants is referred to as horizontal cooperation. Until now the latter has taken place exclusively via human-to-human negotiation, but automation also needs to be appropriately integrated in horizontal cooperation. This leads to a large number of new questions as well as new potential solutions. Vertical and horizontal cooperation can be seen as connected in the sense of cooperation networks that complement each other and enable cooperation schemata (Flemisch and Lüdke 2009, personal communication; Zimmermann and Bengler 2013). Another important degree of freedom for cooperation networks is that they can comprise aspects of decentralized cooperation, such as cooperation between individual vehicles, and centralized or centrally mediated cooperation, via a traffic control room, for instance. Completely centralized control via a traffic control room represents an extreme variant in the design space of what would normally be a cooperative traffic system, and it is one that might be deployed for special situations. In general, thinking in networks of horizontal and vertical cooperation enables dynamic

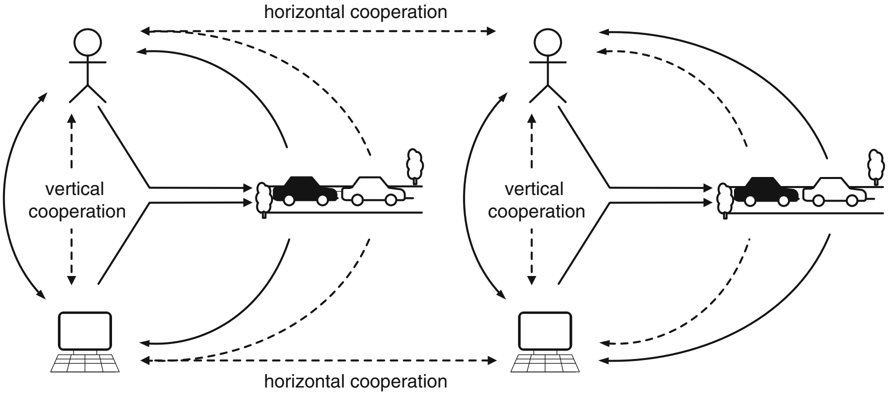


Fig. 6 Horizontal and vertical cooperation in vehicle guidance

negotiation of the division of control between the human and machine actors and supports a clear design thinking even as the number of cooperation partners in future traffic systems increases.

References

Bernotat R (1970) Anthropotechnik in der Fahrzeugführung (human engineering in vehicle guidance). Ergonomics 13(3):353–377

Biester L (2008) Cooperative automation in automobiles. Thesis, Humboldt-Universität zu Berlin

Donges E (1982) Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen (aspects of active safety for the guidance of a passenger car). Automob Ind 27(2):183–190

Duden (2014a) Kooperation (cooperation). http://www.duden.de/rechtschreibung/Kooperation. Accessed 16 June 2014

Duden (2014b) Manöver (maneuver). http://www.duden.de/rechtschreibung/Manoever. Accessed 16 June 2014

Flemisch FO, Adams CA, Conway SR, Goodrich KH, Palmer MT, Schutte MC (2003) The H-Metaphor as a guideline for vehicle automation and interaction. Report no. NASA/TM-2003-212672, NASA Research Center, Hampton

Flemisch F, Kelsch J, Schieben A, Schindler J (2006) Stücke des Puzzles hochautomatisiertes Fahren: H-Metapher und H-Mode (pieces of the puzzle highly automated driving: H-Metaphor and H-Mode). 4. Workshop Fahrerassistenzsysteme, Löwenstein, 4–6 Oktober 2006

Flemisch F, Kelsch J, Löper C, Schieben A, Schindler J (2008) Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In: de Waard D, Flemisch F, Lorenz B, Oberheid H, Brookhuis K (eds) Human factors for assistance and automation. Shaker, Maastricht

Flemisch F, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2011) Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cogn Tech Work 14(1):3–18

Flemisch F, Meier S, Baltzer M, Altendorf E, Heesen M, Griesche S, Weißgerber T, Kienle M, Damböck D (2012) Fortschrittliches Anzeige- und Interaktionskonzept für die kooperative Führung hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse mit H-Mode 2D 1.0 (advanced display and interaction concept for cooperative guidance and control of highly automated vehicles: selected results with H-Mode 2D 1.0). 54. Fachausschusssitzung Anthropotechnik: Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung

Flemisch FO, Bengler K, Winner H, Bruder R (2014) Towards a cooperative guidance and control of highly automated vehicles: H-Mode and conduct-by-wire. Ergonomics 57(3):343–360

Gasser TM, Arzt C, Ayoubi M, Bartels A, Eier J, Flemisch F, Häcker D, Hesse T, Huber W, Lotz C, Maurer M, Ruth-Schumacher S, Schwarz J, Vogt W (2012) Project group "Rechtsfolgen zunehmender Fahrzeugautomatisierung" (legal consequences of increasing vehicle automation) Federal Highway Research Institute (BASt). Wirtschaftsverlag NW, Bergisch Gladbach

Gottschalk-Mazouz N (2007) Was ist Wissen? Überlegungen zu einem Komplexbegriff (What is knowledge? Reflections of a complex concept). In: Ammon S, Heineke C, Selbmann K (eds) Wissen in Bewegung (knowledge in motion). Velbrück, Weilerswist

Griffiths P, Gillespie RB (2004) Shared control between human and machine: haptic display of automation during manual control of vehicle heading. In: Proceedings of the 12th international symposium on haptic interfaces for virtual environment and teleoperator systems, IEEE, Chicago

Hakuli S, Bruder R, Flemisch F, Löper C, Rausch H, Schreiber M, Winner H (2009) Kooperative Automation (cooperative automation). In: Winner H, Hakuli S, Wolf G (eds) Handbuch Fahrerassistenzsysteme (Handbook of driver assistance systems). Vieweg + Teubner, Wiesbaden

Hoc JM (2000) From human – machine interaction to human – machine cooperation. Ergonomics 43(7):833–843

Hoc JM (2001) Towards a cognitive approach to human-machine cooperation in dynamic situations. Int J Hum Comput Stud 54:509–540

Hoc JM, Mars F, Milleville-Pennel I, Jolly E, Netto M, Blosseville JM (2006) Evaluation of human-machine cooperation modes in car driving for safe lateral control in bends: function delegation and mutual control modes. Le Trav Hum 69:153–182

Hoeger R, Zeng H, Hoess A, Kranz T, Boverie S, Strauss M, Jakobsson E, Beutner A, Bartels A, To TB, Stratil H, Fürstenberg K, Ahlers F, Frey E, Schieben A, Mosebach H, Flemisch FO, Dufaux A, Manetti D, Amditis A, Mantzouranis I, Lepke H, Szalay Z, Szabo B, Luithardt P, Gutknecht M, Schömig N, Kaussner A, Nashashibi F, Resende P, Vanholme B, Glaser S, Allemann P, Seglö F, Nilsson A (2011) Final report, deliverable D61.1. Highly automated vehicles for intelligent transport (HAVEit), 7th framework programme

Hollnagel E, Woods DD (1983) Cognitive systems engineering: new wine in new bottles. Int J Man Mach Stud 18:583–600

Holzmann F (2007) Adaptive cooperation between driver and assistant system. Springer, Berlin

Inagaki T (2008) Smart collaboration between humans and machines based on mutual understanding. Annu Rev Control 32:253–261

Mulder M, Abbink DA, Boer ER (2012) Sharing control with haptics: seamless driver support from manual to automatic control. Hum Factors 54(5):786–798

Oxford Dictionaries (2014a) Cooperation. http://www.oxforddictionaries.com/definition/english/cooperation. Accessed 16 June 2014

Oxford Dictionaries (2014b) Manoeuvre. http://www.oxforddictionaries.com/definition/english/manoeuvre. Accessed 16 June 2014

Pacaux-Lemoine MP, Debernard S (2007) Common work space or how to support cooperative activities between human operators: application to fighter aircraft engineering. In: Harris D (ed) Engineering psychology and cognitive ergonomics. Springer, Berlin

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Syst Hum 30(3):286–297

Rasmussen J (1983) Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans Syst Man Cybern 13(3):257–266

Sheridan TB (1976) Toward a general model of supervisory control. In: Sheridan TB, Johannsen G (eds) Monitoring behaviour and supervisory control. Springer, New York

Sheridan TB (2002) Humans and automation: system design and research issues. Human Factors and Ergonomics Society, Santa Monica

Sheridan TB, Parasuraman R (2006) Human-automation interaction. In: Nickerson RS (ed) Reviews of human factors and ergonomics. Human Factors and Ergonomics Society, Santa Monica, pp 89–129

Swartz D (1997) Culture and power: the sociology of Pierre Bourdieu. The University of Chicago Press, Chicago

Winner H, Hakuli S (2006) Conduct-by-wire – following a new paradigm for driving into the future. FISITA 2006 World Automotive Congress, Yokohama

Zimmermann M, Bengler K (2013) A multimodal interaction concept for cooperative driving. In: IEEE intelligent vehicles symposium (IV), Gold Coast

Conduct-by-Wire

60

Benjamin Franz, Michaela Kauer, Sebastian Geyer, and
Stephan Hakuli

Contents

Abstract

In this article, the development of the maneuver-based vehicle guidance concept Conduct-by-Wire is described. After a brief introduction, the function allocation between driver and vehicle is presented in detail. A promising approach is the separation of decision making and execution. Therefore, while driving with Conduct-by-Wire, the driver passes maneuver commands to the vehicle which are then translated into driving functions. In this article, the development and verification of both, maneuvers and driving functions, are described. The article

---

Contribution to this publication was developed during research assistant work at the Institute of Automotive Engineering at the Technische Universität Darmstadt.

B. Franz (✉) • M. Kauer
Institut für Arbeitswissenschaft, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: b.franz@custom-interactions.com; kauer@iad.tu-darmstadt.de

S. Geyer
Mercedes-Benz Cars, Daimler AG, Sindelfingen, Germany
e-mail: sebastian.geyer@daimler.com

S. Hakuli
Continental Engineering Services GmbH, Frankfurt am Main, Germany
e-mail: stephan.hakuli@ipg.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_59

closes with the development and evaluation of an interaction concept for maneuver inputs and an outlook on future work.

1 Introduction

One possible form of cooperative vehicle guidance (▶ Chap. 59, "Cooperative Guidance, Control, and Automation") is represented by the Conduct-by-Wire (Winner and Heuss 2005) vehicle guidance paradigm. When driving with Conduct-by-Wire, the driver gives maneuver (e.g., "lane change to the left") and parameter commands (e.g., desired speed) to the vehicle, which are subsequently reviewed and independently carried out with the help of driving functions (Schreiber et al. 2009). From the driver's point of view, this transforms the continuous input of control variables (e.g., steering, braking) at the stabilization level into discrete input of commands at the guidance level (Fig. 1; for an explanation of the three-level model of vehicle guidance, see ▶ Chap. 2, "Driver Behavior Models" and (Donges 1982)). To successfully shift interaction to the guidance level, several key topics had to be examined during the development of

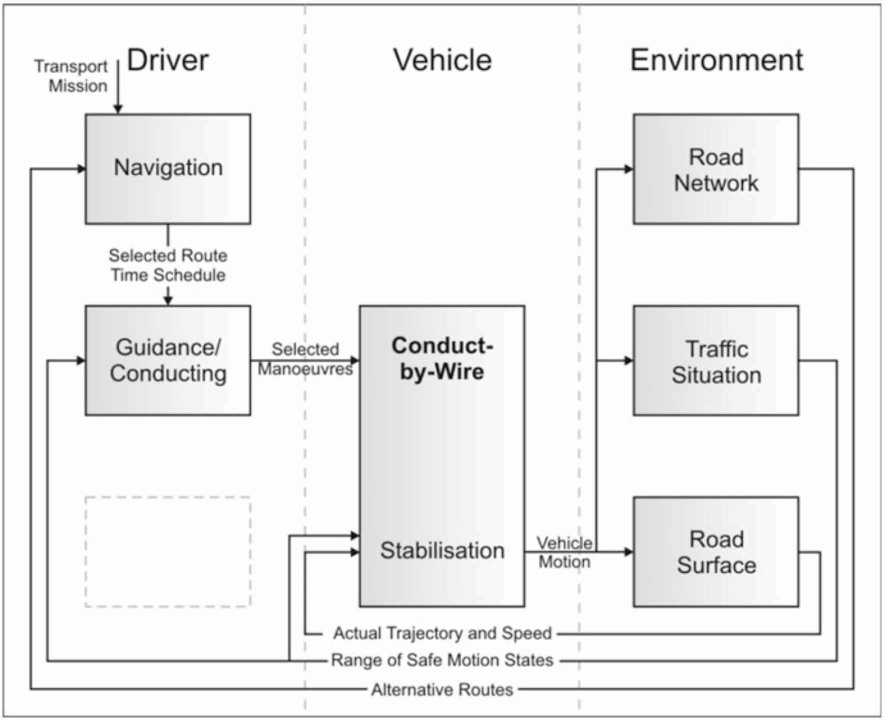


Fig. 1 Three-level model of vehicle guidance for Conduct-by-Wire (From Winner and Hakuli (2006))

Conduct-by-Wire. The research study results presented in the following sections were obtained from several projects supported by the Deutsche Forschungsgesellschaft (DFG). Divisions of tasks between driver and vehicle will be used as a starting point (Sect. 2). Section 3 describes the maneuvers and their driving functions, along with maneuver input. The chapter concludes with a summary and an overview of future research topics (Sect. 4).

2 Division of Tasks Between Driver and Vehicle

The division of tasks between human and machine is a central element in the automation of functions (Chapanis 1965; Parasuraman et al. 2000). In order to provide the driver with the best possible support, simple human information processing systems were taken into account during the development of Conduct-by-Wire (Luczak 1975).

Simple information processing (Fig. 2) can be organized into four successive steps: stimulus detection (e.g., detection of a light stimulus), stimulus recognition (e.g., a red traffic light), decision making (e.g., stopping at the red light), and taking action (e.g., stepping on the brakes). Here it is assumed that, for some stages, humans will have advantages over automated systems, while in other stages, the automated systems will have advantages over humans (cf. Fitts 1951; Chapanis 1965; Edwards and Lees 1973; Price 1985; Kraiss and Schmidtke 2002; Sheridan 2006).

Throughout the development of Conduct-by-Wire, these differences between human and technological strengths were used as the basis for the division of tasks between driver and automation (cf. Franz 2014). It was assumed that humans have advantages over automation during stimulus detection (e.g., high sensitivity to visual and auditory stimuli), while automation has access to a larger range of sensors (e.g., RADAR waves) (Kraiss and Schmidtke 2002). For this reason, Conduct-by-Wire assigns stimulus detection as a cooperative task between human and vehicle (see Fig. 3). When it comes to recognizing stimuli, humans have partial advantages over automation (e.g., excellent pattern recognition for visual and

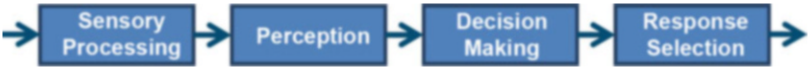


Fig. 2 Simple information processing system (Based on Luczak (1975). Representation from Parasuraman et al. (2000))

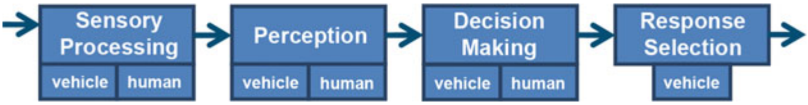


Fig. 3 Fundamental division of tasks in Conduct-by-Wire based on simple information processing systems as outlined by Luczak (1975) (Representation from Franz (2014))

auditory stimuli), while automation has partial advantages over humans (e.g., rapid reaction times) (Kraiss and Schmidtke 2002). For this reason, Conduct-by-Wire allows the driver and vehicle to jointly undertake stimulus recognition (see Fig. 3). Similarly, both humans and machines have advantages during decision-making processes. In general, machines make decisions more rapidly than humans and are more robust against errors when operating under a fixed decision tree (Kraiss and Schmidtke 2002). On the other hand, humans have advantages when decisions must be made on the basis of incomplete information (Kraiss and Schmidtke 2002). In Conduct-by-Wire, therefore, rule-based decisions are made by the vehicle, while all other decisions are made by the driver (see Fig. 3). When it comes to executing actions, the advantages of automation prevail (e.g., precise adjustment of lane drift) (Kraiss and Schmidtke 2002); hence, the execution of actions is biased toward automation in Conduct-by-Wire. For a more in-depth description of this allocation, see Franz (2014).

Concretely, this means (cf. Franz 2014): vehicle and system jointly take part in detecting and recognizing stimuli from the environment. Based on the recognized stimuli, the vehicle specifies a decision space for the human containing those maneuvers which the vehicle can safely execute in the given traffic situation (e.g., changing lines to the right will not be offered when there is no right lane). The driver then chooses an option from this decision space and delivers it to the vehicle with the help of the human-machine interface. The vehicle translates this decision into control variables with the help of driving functions and carries out the action.

Maneuver commands delivered to the vehicle can have additional parameters applied with the help of the driver. A total of three parameters are available for the driver: desired speed, eccentricity within the lane, and time gap to the vehicle in front (Schreiber et al. 2009). After the parameters are given, they are scanned by the vehicle for feasibility and executed independently. For example, the desired speed a driver enters will not always correspond to the vehicle's actual driving speed, but will instead be determined according to legal and physical limits on speed.

3 Maneuver and Driving Functions

Within the Conduct-by-Wire concept, there is a differentiation between explicit and implicit maneuvers (cf. Schreiber et al. 2009, 2010; Schreiber 2012). Explicit maneuvers are self-contained units of activity initiated by driver action and subsequently carried out by the vehicle (e.g., changing lanes to the right). As described in the introduction, current feasibility is examined before the execution of an explicit maneuver (e.g., in the case of a lane change to the right, the vehicle checks whether other vehicles are occupying the target lane within the relevant area). After executing the explicit maneuver, the vehicle independently carries out an implicit maneuver (e.g., following the lane). In contrast to explicit maneuver, which has a

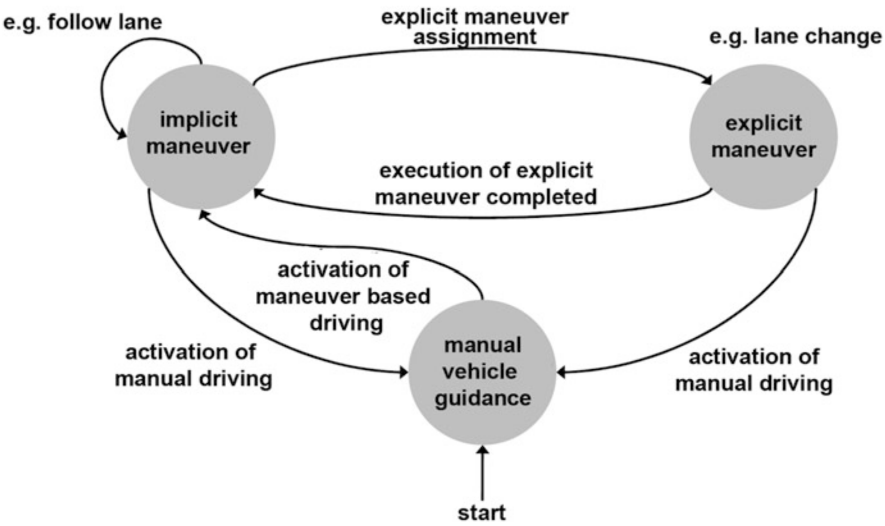


Fig. 4 Representation of the transition between an explicit and implicit maneuver, as well as the transition to and from normal vehicle guidance (Based on Schreiber (2012))

Table 1 Maneuvers and parameters within Conduct-by-Wire (Based on Schreiber et al. (2010), Franz (2014))

Maneuver	
Following the course of the road (including braking, stopping, and starting)	Implicit
Straight	Explicit
Lane change left/right	Explicit
Turning (half) left/right	Explicit
Parameter	
Desired speed	
Time gap to the vehicle in front (1, 1.5, 2, 2.5 s)	
Eccentricity within the lane (20 % of lane width to the left, 10 % left, none, 10 % right, 20 % right)	

defined starting and ending point, implicit maneuvers are units of actions which are not initiated by the driver and which are unrestricted in duration (see Fig. 4).

Maneuvers relevant for Conduct-by-Wire have been identified on the basis of existing studies (Schreiber et al. 2009) using decision point analysis (cf. Schreiber et al. 2010; Schreiber 2012) and defined for various contexts of use (highway, country, and city driving). Driving functions are allocated to the various maneuvers, by means of which the vehicle puts together the desired maneuver. An overview of all maneuvers and parameters within Conduct-by-Wire can be extracted from Table 1. The development and evaluation of driving functions are described in the following paragraph.

3.1 Development and Evaluation of Driving Functions

Explicit maneuver commands given by the driver to the vehicle, along with maneuvers implicitly selected by the vehicle, are interpreted by a maneuver control system implemented as a state machine and allocated to the execution of driving functions. The catalog of driving functions consists of elementary functions, concatenating functions, and functions which take effect either longitudinally or laterally, such as maintaining speed, preparing to merge, targeted braking, avoiding obstacles within the lane limits, changing lanes, turning, and many others. At any given instant, exactly one pair of functions is active, one lateral and one longitudinal; the selection, activation, deactivation, and parameterization of these functions are the task of the higher-level maneuver control system. In order to satisfy its task as interpreter of maneuver inputs, the catalog of driving functions must fulfill various requirements, of which several will be presented in the following section as examples.

3.1.1 Completeness

The catalog of driving functions must prepare a driving function for every situation occurring within the admitted area of discourse. If this is not possible, controlled degradation to a manual form of steering must occur. Since it is inadmissible to draw conclusions about suitability in any given scenario from a successfully tested array of traffic scenarios, function development follows a falsification approach: the universal hypothesis of the completeness of the available catalog of functions and its accompanying set of regulations must be refuted. In other words, it is necessary to search for those traffic situations which cannot be resolved using the currently available functional range.

Figure 5 depicts the associated iterative development process. Starting at the upper right, the respective current state of the function catalog and the maneuver control system are reviewed in relevant test scenarios which are reduced to the necessary details for simulation and then implemented as simulation cases. For each test case, there is a solution strategy in the form of simulated event-dependent or path-dependent driver input. The combination of test case and solution strategy results in a simulatable experimental procedure which is either passed or not passed. Successful completion of a test does not allow conclusions to be drawn about the completeness of the catalog of functions or set of regulations; it merely fails to prove incompleteness, resulting in an increase in complexity for the scenario or the selection of a new scenario for the next test. In the case of an unsuccessfully completed test scenario, it must be examined whether the cause can be attributed to an insufficiently implemented function or to the absence of a necessary driving function or a transition between functions.

3.1.2 Safety

During active Conduct-by-Wire functionality, the driver and the driving functions are arranged in series. Contrary to a parallel arrangement, the driver lacks

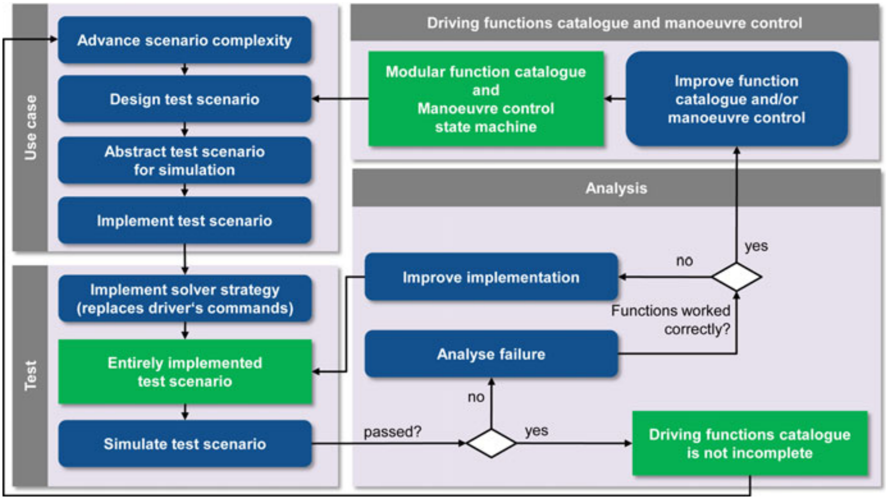


Fig. 5 Iterative falsification approach (From Hakuli et al. (2010))

mechanical access to the actuators. For this reason, the implementation must satisfy functional safety requirements of by wire systems according to ISO 26262.

3.1.3 Performance and Quality of Execution

Compared to a conventionally operated vehicle, the quality of maneuver execution may not lead to acceptance problems. While a diminished quality of execution compared to manual driving might be acceptable in automated vehicles because of the added value provided by a complete decoupling of responsibility for vehicle guidance, quality of execution during maneuver-based vehicle guidance occurs in constant competition with the driver, who occupies the role of observer and decision maker. The significant acceptance factors concerning the driving functions are availability, predictability of behavior, and the possibility of influencing the execution by means of parameterization.

3.1.4 Evaluation of Alternatives

Unlike fully automated vehicles, which reach all relevant driving decisions independently, Conduct-by-Wire generally leaves the task of evaluating alternatives to the driver (Sect. 2). Nevertheless, depending on the situation or known driver preferences (e.g., frequently traveled routes), prioritized alternatives can be predetermined for the maneuver to be executed (e.g., in a form which then only needs to be enabled for execution).

The preceding examination describes Conduct-by-Wire's fundamental interaction concept in the form of maneuvers assigned by the driver and executed by the delineated maneuver control system and the catalog of driving functions. In addition to simple maneuver commands, scenarios with increased collision risk involving a cross of trajectory with other road users require further decision making in

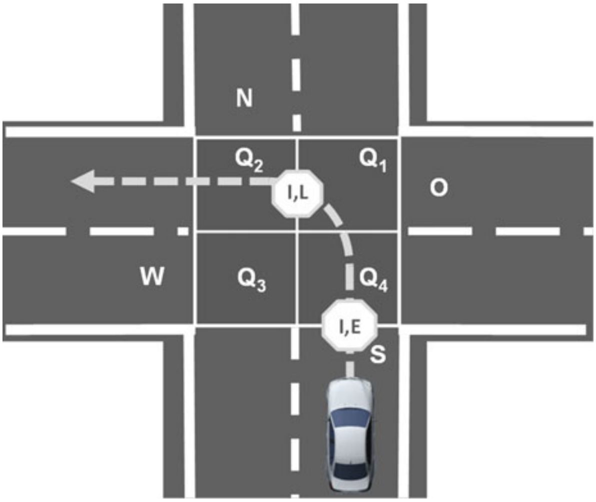


Fig. 6 Example scenario with identified gates (intersection quadrants Qi, gate positions I,E and I,L, and compass directions of all paths through the intersection) (From Geyer (2013))

order to safely execute the maneuver. While this task would quickly lead the automation to its technical limits if it were undertaken independently, a cooperative interaction approach to decision making between driver and automation during the execution of maneuver affords new design possibilities for the technical implementation of an interaction concept which takes driver and automation requirements equally into account. One interaction concept derived from these considerations is the "gate concept" first proposed by Geyer et al. (2011), which consists of segmented maneuver execution. In this concept, the gates mark the points along a planned trajectory where a decision on the continuation of the driving mission has to be made. Each gate is assigned to an information cluster, which comprises different information needed at that point.

Figure 6 shows the gate concept using the example of an X-intersection where the right of way is "right before left." The Conduct-by-Wire vehicle approaches the intersection from the south and turns left. During maneuver execution, a sequence of two gates has to be passed. The first gate, "intersection entry (I,E)," is positioned at the entrance to the intersection. In order to decide whether it is possible to safely continue executing the left turn maneuver until the next gate is reached, road users approaching the intersection from the east, and who therefore have the right of way, must be taken into account.

Furthermore, the free space between the two gates needed for maneuver execution has to be examined. In order to pass the second gate, "intersection left (I,L)," it is necessary to take into account oncoming vehicles from the north, as well as the area remaining until leaving the intersection. Studies on the feasibility of the gate concept show its theoretical application to 400 representative scenarios. The position of the gates as well as the required information allocated to each gate can always be clearly defined (Geyer 2013).

When designing an interaction concept for partially automated vehicle guidance, various system designs with increasing degrees of automation can be derived from the gate concept, ranging from automation, which displays the next gate while the driver reaches a decision, to automation which suggests a decision, all the way to automation which reaches a decision independently. The gate concept must be expanded independently of the respective system design to include a safety concept which transfers the vehicle to a safe state (stopping at the gate) if no decision is made by the driver or the automation. Technical implementation of this safety maneuver affords the possibility of integrating the human driver according to the fundamental design principle of a cooperative interaction by allowing sufficient time for decision making. “Signalizing deceleration” is one preferred control strategy (Geyer 2013). By means of a slight initial deceleration, the driver is informed about the start of the approach maneuver and thus the need to make a decision, without greatly disrupting execution of the current maneuver. The second level of deceleration corresponds to the required deceleration for stopping the vehicle at the gate. A simulator study with 42 subjects shows differences in evaluating various system designs of the gate concept by potential users (Geyer 2013). In this study, all subjects were able to complete the selected representative scenarios for the lowest (automation displays the next gate and the driver decides) and highest (automation reaches a decision independently) degrees of automation. However, there is a room for critical discussion regarding the middle degree of automation, in which the driver receives a suggested decision from the automation in combination with the next gate being displayed, due to poorer results compared to the other system designs. This degree of automation leads to longer decision times compared to the lowest degree of automation, in which the driver makes a decision without support from the automation. On the other hand, this system design was not really accepted by some test subjects, as shown while examining the moment of decision, or even led to safety critical irritations, which led to collisions with other road users in a few cases. These results suggest that a clear division of tasks between driver and automation in the form of “display” and “decision” system models is preferable. Taken as a whole, the gate concept provides a basis for transferring the Conduct-by-Wire concept into more complex, inner-city scenarios. The various system designs can be seen as potential development stages of the described interaction concept. Starting with a low level of automation, migration from today’s assisted driving to partially automated driving could proceed.

Depending on the user’s experience with the system and its technical development, particularly with respect to the environment perception systems, it would be conceivable to enhance the degree of automation and thus expand the functional scope of automation. Based on the gate concept introduced in this section, the next section will envision the development and evaluation of a concrete maneuver interface for Conduct-by-Wire. This development and evaluation are based on the highest automation level in the gate concept (decisions reached by the automation), so that the driver only has to make decisions about an upcoming driving maneuver (e.g., “turn left”) which is then executed by the vehicle.

3.2 Development and Evaluation of the Maneuver Interface

Particularly when it comes to situations at complex intersections (e.g., where there are several options for the turning right maneuver), it is no longer possible to transmit maneuvers to the vehicle with the help of usual control elements (steering wheel and pedals) (Franz 2014). For this reason, since 2008, various interaction concepts have been developed iteratively for Conduct-by-Wire which enable maneuver input, initially on highways and later in the previously described situations at complex intersections (Kauer et al. 2010; Franz et al. 2012a, b; Schreiber 2012; Franz 2014). Starting from standards and norms, recommendations and measured values are transmitted to Conduct-by-Wire and formulated as requirements (cf. Schreiber 2012; Franz 2014). Here it can be shown that the number of input errors, along with eye behavior, is the most suitable variables for evaluating interactions in maneuver-based vehicle guidance concepts.

In the first prototypical implementation, a tactile touch display was positioned on the steering wheel and the inputs delivered to the vehicle (Kauer et al. 2010; Schreiber 2012). By means of predefined buttons, maneuvers and parameters required on the highway could be delivered to the vehicle (see Fig. 7). With the help of this implementation, Conduct-by-Wire's feasibility for highway driving could be demonstrated from the driver's point of view (Schreiber 2012). However, it was shown that the percentage requirement for identical eye behavior could not be fulfilled (Franz et al. 2012a; Franz 2014). While driving with the tactile touch

Fig. 7 Content depicted on the tactile touch display (From Franz et al. (2011), derived from Kauer et al. (2010)). In this example, the maneuver “lane change left/right” is available, and the maneuver “follow the course of the road” is active. Desired speed is set to 100 km/h, while the vehicle is currently driving at about 85 km/h

display, drivers looked at the input device significantly longer and more frequently than during a comparable journey with usual control elements.

In order to improve glance behavior, in the next stage of development, the controls were separated from the display (Franz et al. 2011, 2012a). All information needed by the driver (e.g., available maneuvers) were presented in a (simulated) head-up display. Through a touchpad integrated in the right armrest of the driver's seat, the driver could input desired maneuvers and parameters in the form of gestures. During evaluation, glance behavior was improved compared to the tactile touch display, but ranked far poorer than glance behavior while driving with the usual control elements (Franz et al. 2012a; Franz 2014). Furthermore, the amount of input errors with gesture recognition increased to an undesirable degree, since the driver's gestures had to be first properly performed and then correctly recognized. The third stage of development consisted of developing the current up-to-date *pieDrive* interaction concept, which combines the low amount of input errors from the tactile touch display with the improved eye behavior of gesture recognition (Franz et al. 2012b; Franz 2014).

In order to arrive at a higher percentage of eye focus on the road, the *pieDrive* interaction concept is also based on the separation of controls from display. As with gesture recognition, the driver receives all necessary information through the head-up display and performs maneuver and parameter commands through a touchpad integrated into the right armrest of the driver's seat. The following section will first describe the head-up display and then the control concept.

In the head-up display, available and active maneuvers are arranged in a semicircular menu (see Fig. 9). If several maneuvers are available, the semicircle is subdivided into segments, where each segment represents a maneuver (see Fig. 8). The organization of maneuvers within the semicircle is based on their respective directions, e.g., a "turn right" maneuver is displayed on the right, and a

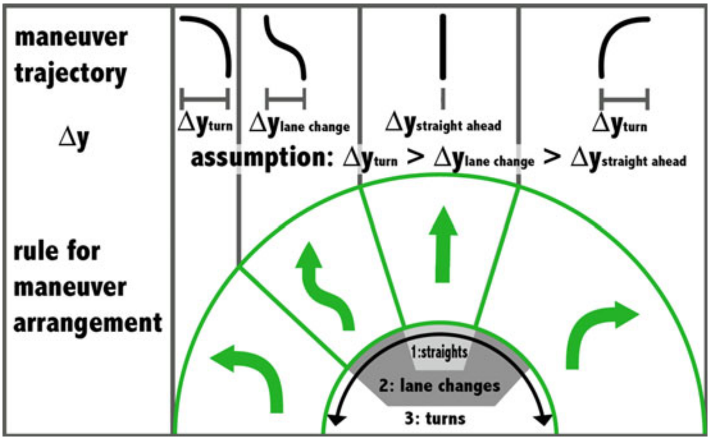


Fig. 8 Organization of maneuvers in the semicircular *pieDrive* menu (From Franz et al. (2012b))

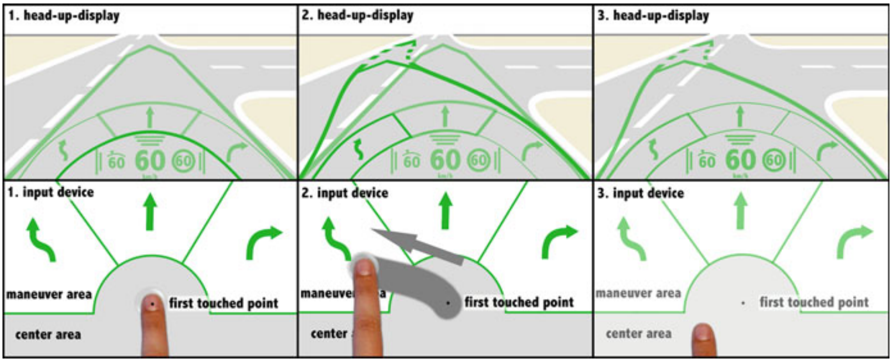


Fig. 9 Contact-analog head-up display for the *pieDrive* design solution (From Franz et al. (2012b)). Content displayed on the input device is for clarification purposes and is not visible in the implemented human-machine interface. (*1*) The maneuver "follow the course of the road" is active, and the driver begins maneuver input. (*2*) The driver has selected the maneuver "lane change left." (*3*) The driver has engaged the lane change, which is now active

"turn left" maneuver is displayed on the left. Additionally, the maneuver segment for the active maneuver is highlighted in bright green.

Parameters are displayed in the inner area of the semicircular maneuver menu (see Fig. 9). The driver's desired speed is shown on the left, the actual vehicle speed is shown on top, and the speed limit is shown on the right. In order to better differentiate between the three speeds, the representation of desired speed is supplemented with a circular segment and a triangle. Additionally, a circular symbol is superimposed on the numerical value for the speed limit.

The time gap to the vehicle in front is displayed between the three speeds with the help of at most four horizontal bars and a stylized vehicle symbol. Here the number of bars represents the configured time gap (1 bar, 1 s, up to 4 bars, 2.5 s). The stylized vehicle symbol is also used to represent the vehicle's eccentricity within the lane. The preset level of eccentricity is clarified by four vertical bars and the position of the stylized vehicle symbol.

In order to clarify the active maneuver beyond the highlight on the maneuver menu, the vehicle's future trajectory is represented by an arrow lying on the road (see Fig. 9). The represented trajectory corresponds to the correct location.

In order to deliver a maneuver command to the vehicle, the driver places a finger on the touchpad (see Fig. 9 left). This causes the inner circle on the maneuver menu in the head-up display to be highlighted in bright green. Then the driver moves his or her finger in the direction of the desired maneuver segment (see Fig. 9 middle). Once a maneuver segment is reached, it is highlighted bright green in the head-up display. Additionally, a second arrow in the head-up display drawn with a dotted line shows the trajectory the vehicle would follow when carrying out the selected maneuver. To assign the selected maneuver, the driver then lifts his or her finger above the corresponding maneuver segment (see Fig. 9 right). A selection can be corrected by choosing another maneuver segment or canceled if the driver lifts his or her finger within the inner semicircle (start zone).

The *pieDrive* interaction concept has been validated in multiple driving simulation studies (see Franz 2014). Among other tests, the concept was screened in a study which took place over four test days per participant in order to investigate changes in driver behavior over time. This examination showed that, despite a thorough theoretical introduction to the concept, subjects only required very short learning periods to master the controls of the interaction concept. This was evident particularly in the significant reduction of the number of false inputs by the subjects during maneuvers necessary for following the predetermined target route (errors at decision points). Even during the first test drive, a significant reduction was seen. All subjects performed their last test drive without errors. Furthermore, with *pieDrive*, there were no input errors due to an incorrect assignment of the driver inputs.

All drivers were able to manage all route elements and simulated traffic situations with Conduct-by-Wire and *pieDrive* within the context of the tests. This was the case for highway, country, and city driving.

Analysis of eye movement further showed that separating output (head-up display) and input (touchpad) elements was extremely effective at reducing the time during which the eyes were diverted from the road. Thus, the driver keeps traffic events in view even while inputting new maneuvers and can react to them as necessary.

4 Conclusion and Outlook

All existing results indicate that Conduct-by-Wire is a promising concept for partly automated vehicle guidance. The concept's strength lies in relieving the driver from monotonous and undemanding tasks without allowing complete withdrawal from the task of driving.

Though previous studies provide enough for an initial appraisal, it is evident that further research into the concept is necessary (cf. Franz 2014). Until now, only shorter driving units or low numbers of journeys have been examined, without investigating how driving behavior changes over a longer period of time during exclusive use of Conduct-by-Wire. It is possible that drivers could develop strategies for traveling as far as possible without participating in the task of driving. This would reduce the system's overall level of safety, since the driver would no longer be available as a monitoring authority.

Furthermore, since all results presented here are based on driving simulation studies, no conclusions can be drawn regarding the functionality of the Conduct-by-Wire system in real traffic. This applies to both the achievable reliability of such a system and the flexibility necessary for executing actions under varying traffic conditions.

Although existing studies have attempted to replicate actual traffic situations as broadly as possible, until now, the focus has been on developing driver-vehicle interaction and driving functions in standard cases. Special scenes (e.g., roundabouts with multiple lanes) or critical situations (e.g., another vehicle takes right of way) have only marginally been considered. Nevertheless, dealing with nonstandard situations makes up a substantial part of the system's feasibility for real street traffic.

In summary, Conduct-by-Wire is a promising vehicle guidance concept which in its most current implementation solves some of the interaction problems of previous assistance systems (e.g., prioritization of system feedback is no longer necessary, no diversion of the eyes due to an uneven control concept). At the same time, existing research results indicate that Conduct-by-Wire could represent a suitable step in the migration toward fully automated driving.

References

Chapanis A (1965) On the allocation of functions between men and machines. Occup Psychol 39:1–11

Donges E (1982) Aspekte der aktiven Sicherheit bei der Führung von Personenkraftwagen (Aspects of active safety while driving passenger cars). Automobilindustrie 27(2):183–190

Edwards E, Lees FP (1973) Man and computer in process control. Institution of Chemical Engineers, London

Fitts PM (1951) Human engineering for an effective air navigation and traffic control system. National Research Council, Washington, DC

Franz B (2014) Entwicklung und Evaluation eines Interaktionskonzepts zur manöverbasierten Führung von Fahrzeugen (Development and evaluation of an interaction concept for maneuver based vehicle guidance). Technische Universität Darmstadt, Darmstadt

Franz B, Kauer M, Schreiber M, Blanke A, Distler S, Bruder R, Geyer S (2011) Maneuver-based driving today and in the future – development of a new human-machine interface for conduct-by-wire. In VDI-Bericht. Fahrer, Fahrerunterstützung und Bedienbarkeit, vol 2134. Braunschweig

Franz B, Kauer M, Blanke A, Schreiber M, Bruder R, Geyer S (2012a) Comparison of two human-machine-interfaces for cooperative maneuver-based driving. Work J Prev Assess Rehabil 41 (1):4192–4199

Franz B, Kauer M, Bruder R, Geyer S (2012b) pieDrive – a new driver-vehicle interaction concept for maneuver-based driving. In: Toledo-Moreo R, Bergasa LM, Sotelo MÁ (eds) Proceedings of the 2012 international IEEE intelligent vehicles symposium workshops (W2: workshop on human factors in intelligent vehicles), Alcala de Henares (Spain)

Geyer S (2013) Entwicklung und Evaluierung eines kooperativen Interaktionskonzepts an Entscheidungspunkten für die teilautomatisierte, manöverbasierte Fahrzeugführung (Development and evaluation of a cooperative interaction concept on decision points for semi automated maneuver based vehicle guidance). Fortschritt-Berichte VDI Reihe 12, vol 770. Düsseldorf

Geyer S, Hakuli S, Winner H, Franz B, Kauer M (2011) Development of a cooperative system behavior for a highly automated vehicle guidance concept based on the conduct-by-wire principle. In: Intelligent vehicles symposium (IV), 2011 IEEE. IEEE, BadenBaden, pp 411–416

Hakuli S, Kluin M, Geyer S, Winner H (2010) Development and validation of manoeuvre-based driver assistance functions for conduct-by-wire with IPG CarMaker. In: Proceedings of the FISITA 2010 world automotive congress, 30 May–4 June 2010. FISITA, Budapest

Kauer M, Schreiber M, Bruder R (2010) How to conduct a car? A design example for maneuver based driver-vehicle interaction. In: 2010 I.E. intelligent vehicles symposium (IV). IEEE, San Diego, CA, pp 1214–1221

Kraiss K, Schmidtke H (2002) Funktionsteilung Mensch-Maschine (Function distribution human-machine). In: Bundesamt für Wehrtechnik und Beschaffung (ed) Handbuch der Ergonomie: Erg.-Lfg. 7. Carl Hanser, München

Luczak H (1975) Untersuchungen informatorischer Belastung und Beanspruchung des Menschen (Investigations of informational stress and strains of humans), Fortschrittberichte der VDI-Zeitschriften. VDI-Verlag, Düsseldorf

Parasuraman R, Sheridan T, Wickens C (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern A Syst Hum 30(3):286–297

Price HE (1985) The allocation of functions in systems. Hum Factors J Hum Factors Ergon Soc 27 (1):33–45. doi:10.1177/001872088502700104

Schreiber M (2012) Konzeptionierung und Evaluierung eines Ansatzes zu einer manöverbasierten Fahrzeugführung im Nutzungskontext Autobahnfahrten (Concept and evaluation of an approach for a maneuver based vehicle guidance in the context of motorway driving). TU Darmstadt

Schreiber M, Kauer M, Bruder R (2009) Conduct by wire – maneuver catalog for semi-autonomous vehicle guidance. In: Intelligent vehicles symposium, 2009 IEEE: intelligent vehicles symposium. IEEE, Xi'an, China, pp 1279–1284

Schreiber M, Kauer M, Schlesinger D, Hakuli S, Bruder R (2010) Verification of a maneuver catalog for a maneuver-based vehicle guidance system. In: Systems man and cybernetics (SMC), Istanbul, pp 3683–3689

Sheridan TB (2006) Supervisory control. In: Handbook of human factors and ergonomics. Wiley, Hoboken, pp 1025–1052

Winner H, Hakuli S (2006) Conduct-by-wire-following a new paradigm for driving into the future. In: Proceedings of FISITA world automotive congress, Yokohama

Winner H, Heuss O (2005) X-by-Wire Betätigungselemente-Überblick und Ausblick (X-by-Wire actuating elements – overview and outlook). In: Darmstädter Kolloquium Mensch und Fahrzeug. Cockpits für Straßenfahrzeuge der Zukunft, pp 79–115

H-Mode

61

A Haptic-Multimodal Interaction Concept for Cooperative Guidance and Control of Partially and Highly Automated Vehicles

Eugen Altendorf, Marcel Baltzer, Matthias Heesen,
Martin Kienle, Thomas Weißgerber, and Frank Flemisch

Contents

E. Altendorf (✉)
Institut für Arbeitswissenschaft (IAW), RWTH Aachen, Aachen, Germany
e-mail: e.altendorf@iaw.rwth-aachen.de

M. Baltzer • M. Heesen
Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany
e-mail: marcel.baltzer@fkie.fraunhofer.de; matthias.heesen@qiagen.com

M. Kienle • T. Weißgerber
Lehrstuhl für Ergonomie, Technische Universität München, Garching, Germany
e-mail: kienle@lfe.mw.tum.de; weissgerber@lfe.mw.tum.de

F. Flemisch
Institute of Industrial Engineering and Ergonomics, RWTH Aachen University, Aachen, Germany

Systemergonomie – Balanced Human Systems Integration, Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE, Wachtberg, Germany
e-mail: f.flemisch@iaw.rwth-aachen.de; frank.flemisch@fkie.fraunhofer.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_60

Abstract

With increasing technical possibilities in the area of assistance and automation, diverse challenges, risks, and chances arise in the design of assisted, partially and fully automated driving. One of the greatest challenges consists of integrating and offering a multitude of complex technical functions in such a way that the human driver intuitively understands them as a cohesive, cooperative system. A solution to this problem can be found in the H-Mode. It is inspired by the role model of horse and rider and offers an integrated haptic-multimodal user interface for all kinds of movement control. The H-Mode, as presented in this chapter, has been designed for ground vehicles and includes several comfort and security systems on three assistance and automation levels, which can be interchanged fluidly.

1 Introduction

With increasing technical possibilities in the area of assistance and automation, diverse challenges, risks, and chances arise in the design of assisted and partially and fully automated driving. One of the greatest challenges consists of integrating and offering a multitude of complex technical functions in such a way that the human driver intuitively understands them as a cohesive, cooperative system which can reliably, safely, and comfortably be operated. Thereby, the barriers between assistance and automation become increasingly blurred, and it becomes necessary to define complementing degrees of assistance and automation (Gasser et al. 2012). Therefore, it is sensible to focus more on the human involvement in the sense of cognitive compatibility and with regard to trust between human and automation or assistance (cf. Löper et al. 2008; Vanholme 2012 and ▶ Chap. 59, “Cooperative Guidance, Control, and Automation”) as well as on early considerations of the users in the development process (Flemisch et al. 2008b).

Cooperative guidance and control addresses these questions. It describes general degrees of freedom in the cooperation between the human and the automation as a generic concept, i.e., as different levels of vehicle guidance and control (cf. Fig. 11). The here-described H-Mode is an implementation of the concept of cooperative guidance and control.

The fundamental basis for the H-Mode is the H-metaphor, a design metaphor comparable to the desktop metaphor for PCs. It drafts the entire system behavior and the interaction between cooperative or partially as well as highly automated vehicles and the human. The H-metaphor as a basis for the H-Mode has been inspired by the biological archetype of rider-horse or driver-carriage.

The H-Mode emphasizes the haptic-multimodal interaction between driver and carriage or automation and human. With a modern perspective on assistance systems, the H-Mode can be viewed as integrating a system for keeping a safe distance, comparable to an ACC+, an active lane keeping assistance system (LKAS), a warning and interfering system when leaving the lane (Lane Departure

Mitigation/Prevention System), a collision mitigation system, and a traffic jam or highway assistant. It thereby enables the transition to autonomous driving. H-Mode also considers the sharing of dynamic control between human and technology as well as the question of takeover and reaction times that are appropriate in a specific situation.

2 From H-Metaphor to H-Mode

In cooperative vehicle guidance and control (Flemisch et al. 2014), the driving task can be executed by both parties together, i.e., the human driver and the cognitive automation. At the same time, it is also possible that different aspects of the driving task are distributed among the two parties. A central component of cooperative vehicle guidance and control is an automation that is capable of working together with the human in a cooperative way. ▶ Chapter 59, "Cooperative Guidance, Control, and Automation" shows in Fig. 1 an assistance and automation scale as a highly simplified model of the distribution of control between human and machine. It is hereby decisive that in addition to the extremes of manual and fully automated driving, also intermediate degrees may exist, such as partial or high automation, in which the human, as well as the automation, can affect vehicle guidance and control.

To sensibly support and disburden a human in the driving task, the technical system should act comprehensibly and understandably. The requirements in such a complex human-machine system exist not only with respect to technical assistance or automation, but also and especially with respect to the interaction between human and technical system. To support the intuitive comprehensibility of the cooperative automation by the human, a suitable design metaphor can be used – comparable to the desktop metaphor for PCs. Although highly automated guidance and control on the basis of machinable automation has only been introduced a relatively recent concept, there are well-known and established historic examples of a cooperative, shared guidance and control of two cognitively competent entities. The archetype of the H-Mode is the relation between human and riding or coach horse. A horse has powerful sensor systems, cognition, and actuating elements to move autonomously. A rider or carriage driver can influence the movements of the horse in variable degrees of autonomy by adjusting the reins, while the horse will autonomously avoid upcoming obstacles, follow the given path, or follow the rider's specifications more directly. This underlying idea has been exploited further

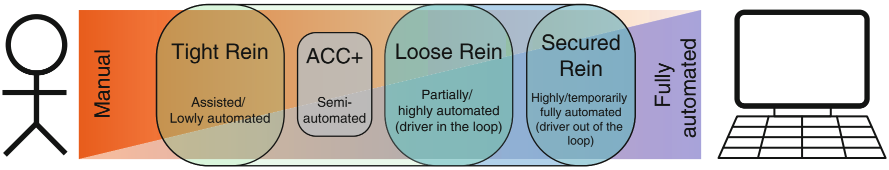


Fig. 1 Assistance and automation scale with H-Mode automation modes

over the course of several years. Research cooperations have been set up to investigate the design metaphor, which describes the fundamental roles and forms of interaction for partially and highly automated, cooperatively controlled vehicles and their users within the H(orse)-metaphor (Flemisch et al. 2003).

The concept of the H-Mode, which describes the haptic-multimodal interaction and execution of the driving task by the human and a highly automated vehicle, emerged from the H-metaphor. While the H-metaphor is a rather superordinate, metaphoric description of cooperative vehicle guidance and control, the H-Mode is the practical implementation of cooperative vehicle guidance and control as a haptic-multimodal interaction language for partially and highly automated vehicles.

3 Cooperative Guidance and Control with the H-Mode

One aspect of the H-metaphor and thereby of the H-Mode is the description of the different degrees of automation and the distribution of authority between driver and automation. Taking the example of the cooperation between horse and rider, it can easily be illustrated that the influence on the motion can be distributed in various ways, depending on the situation.

The H-Mode employs between two and three different degrees of assistance and automation. In the stage “Tight Rein,” the human steers the vehicle to a large extent herself receiving recommendations for action from the automation with respect to speed adjustments or positioning in the lane. In this assisted mode, the driver takes over a high degree of direct control, i.e., the driver’s lateral and longitudinal driving commands are transmitted immediately to the vehicle. The task of the automation in this mode is to support the human accordingly with the help of haptic cues. Among others, these can be physical forces or torques that act on the steering wheel for centering the vehicle in the driving lane.

In the partially automated “Loose Rein,” the automation takes over most parts of the lateral and longitudinal control on the trajectory and stabilization level. Nevertheless, the human remains in the loop in a meaningful way, e.g., by influencing the maneuver to be performed as well as ancillary conditions, such as speed. For this purpose, systems are used that are comparable to an active lane keeping assistant and an adaptive cruise control (ACC) but, as will be described subsequently, can have much broader functionalities. In the “Loose Rein,” this driving action will be transmitted immediately to the vehicle but is noticeable on haptic actuators, such as active steering wheel, foot throttle, and/or sidestick.

The “Secured Rein” describes highly or temporarily fully automated driving, comparable to the mode a carriage driver chooses when putting down the reins on safe roads. Transferred to cars, this could happen on suitable parts of the highway. Hereby, the human can exit the control loop for a predefined period of time, while the vehicle assumes all other aspects of the driving task autonomously. For being able to use such a mode, sufficient communication between vehicle, automation, and environment needs to be in place. The reason for this is that it can take some time for a human to be attentively integrated into the execution of the driving task in unforeseen situations. One possibility of ensuring a reliable, temporarily fully automated ride could be realized on special, certified streets, so-called secured lanes. Such a concept has been introduced by the "eLane" in the project CityMobil (Toffetti et al. 2009).

The degrees can be changed fluidly and/or discretely upon pushing a button. Figure 1 shows the modes of automation in the H-Mode on the assistance and automation spectrum. In the following, the different modes "Tight Rein," "Loose Rein," and "Secured Rein" will be exemplarily explained with the help of prototypical driving situations. Hereby, it is important to keep in mind that the H-Mode has been developed in direct interaction with drivers. The goal was to create a system that is highly intuitive without the use of language – in the end, it has also proven to be successful in a number of tests. The current chapter faces the challenge of only having the possibility of describing the haptic-multimodal interrelation in the H-Mode in words, which could be interpreted by the reader rather in a logical context. Logic and intuitivism are opposing concepts. The H-Mode successfully bridges logic, which can be implemented in a software, and intuitivism, which can be experienced by a driver – much of what is cumbersomely described in words and what needs to be reconstructed in the head of the reader comes naturally and easily when experiencing the H-Mode in practice.

3.1 Exemplary Use Cases for H-Mode

In "Tight Rein" (assisted/lowly automated) the driver guides and controls the vehicle and is assisted by the automation (Fig. 2). This assistance is a soft lane centering (comparable with a lane keeping assistance system, LKAS) as well as a soft velocity and distance control (comparable to an adaptive cruise control, ACC or ACC+, but with weaker intervening behavior). In this context, "soft" means that the driver is still accelerating, braking, and steering, while the automation uses haptic feedback to communicate with the driver.

An everyday traffic situation occurs when approaching another vehicle from behind (Fig. 3). The ego-vehicle has a larger velocity than the vehicle in front. In this situation, an adaptation of the driving behavior concerning velocity or choice of lane becomes necessary.

If the lane to the left is free and a lane change is allowed, in Tight Rein, the automation proposes to change the lane, e.g., with a weak impulse on the steering wheel ("tick"), and displays lane change trajectories (Fig. 3, Phase 1). If the driver

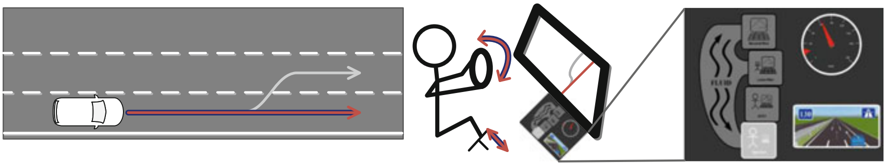


Fig. 2 Driving in "Tight Rein"/lowly automated

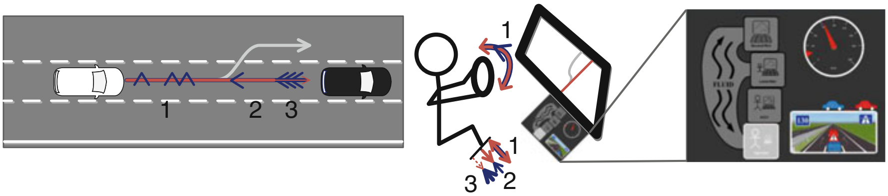


Fig. 3 Approaching another vehicle in “Tight Rein”

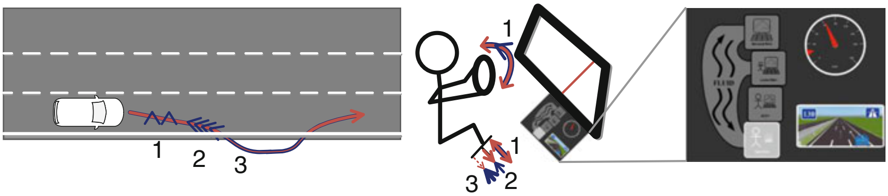


Fig. 4 Virtual Gravel Trap/Lane Departure Prevention/Mitigation System in “Tight Rein”/lowly automated

does not change the lane in time, the automation starts reducing the vehicle’s velocity, e.g., by using counterforces to hinder the accelerator (Fig. 3, Phase 2). If the driver still does not react and a collision is otherwise inevitable, the automation interferes and starts braking with a warning signal, comparable to an emergency brake (Fig. 3, Phase 3).

Another critical situation is unintended lane departure (Fig. 4). If the ego-vehicle risks to depart from the lane with its current trajectory, the automation will start indicating a course correction, e.g., via a short impulse on the steering wheel (“tick” Fig. 4, Phase 1). If the vehicle remains on its course and is close to departing from its current lane, the automation will try to brake (Fig. 4, Phase 2). Before actually departing from the current lane, the automation will interfere and steer back to the center of the lane (Fig. 4, Phase 3). This might also include a short decoupling of the driver. After the situation is resolved, the driver regains control.

The driver can switch from “Tight Rein” to partially/highly automated “Loose Rein” by employing several behavioral options. She can either push the “Loose Rein” button on the display or simply loosen the grasp of the steering wheel or harmonize the activities toward the automation’s activities in fluid operation mode (Fig. 5). The automation detects the driver’s withdrawal and fluidly takes over control, whereas the driver is always able to cancel this control transition. It is still an open issue whether this fluid transition possibility is activated at the beginning of the driving sequence and can be deactivated by the driver or if the driver needs to explicitly activate this mode (e.g., by pressing the respective fluid transition button, see Fig. 5).

In “Loose Rein,” the automation takes over the major part of the driving task, but the driver is still involved in the driving tasks. In this automation mode, the driver

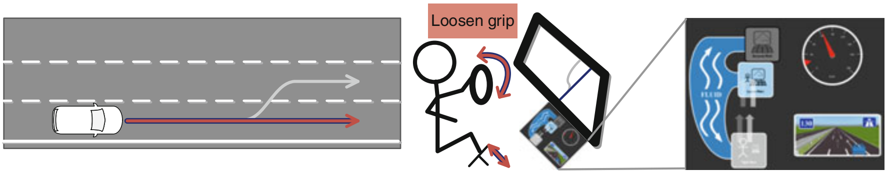


Fig. 5 Transition “Tight Rein”to “Loose Rein”

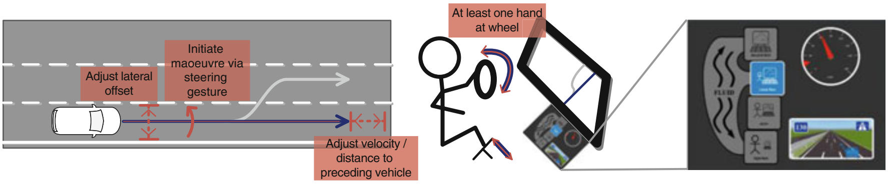


Fig. 6 Driving in “Loose Rein”

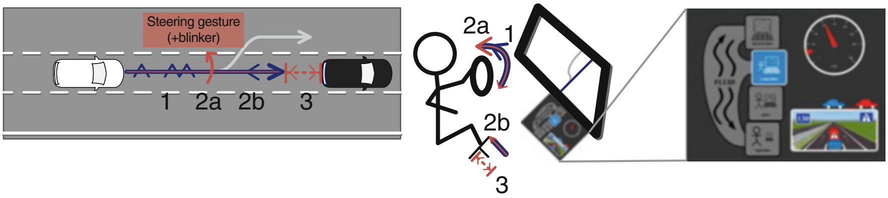


Fig. 7 Approaching a vehicle in “Loose Rein”

induces at most weak additional torques on the steering wheel. She can adapt the velocity, the distance to the vehicle in front, and the lateral offset by weak inputs via steering wheel and accelerator, she can also initiate maneuvers, e.g., a lane change, with a gesture to the respective input device (see Fig. 6).

When approaching a vehicle on a highway in “Loose Rein,” an overtaking maneuver is proposed to the driver if the lane to the left is free and a lane change is allowed. This can be indicated by a visual representation of a lane change trajectory and a decent impulse from the steering wheel (Fig. 7, Phase 1). If the driver performs a steering gesture in the respective direction (maneuver gesture, Fig. 7, Phase 2a), the automation will initiate the corresponding overtaking maneuver. This maneuver selection can be made safer by a necessary activation of the direction indicator. If the driver does not initiate a lane change, the automation will reduce the current velocity (Fig. 7, Phase 2b) and try to follow the vehicle in front with a safe distance, which can be adapted by the driver (Fig. 7, Phase 3). This mode is comparable to an adaptive cruise control (ACC) combined with a strong lane keeping assistance system (LKAS).

A fork-shaped trajectory is another use case for cooperative guidance and control. The automation plans a route and a respective driving trajectory on the

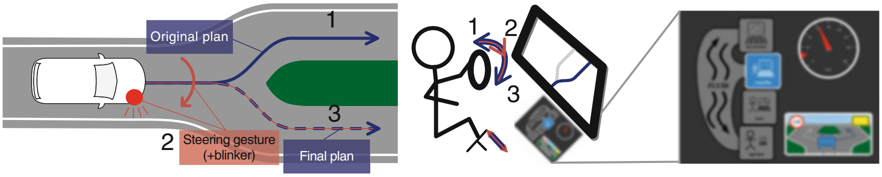


Fig. 8 Driving through a fork in "Loose Rein"

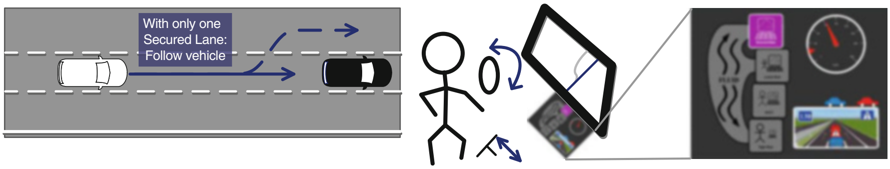


Fig. 9 Driving in "Secured Rein"

basis of the navigation data (Fig. 8, Phase 1). In this example, the planned trajectory corresponds to the left path and will be followed by the automation if the driver does not prefer an alternative solution. If the driver interacts with the automation, for example, by steering to the right (steering gesture) or by indicating another direction ("blink"; Fig. 8, Phase 2), the automation understands the driver's intention and supports the maneuver to the right path (Fig. 8, Phase 3).

In certain situations, the automation offers the Secured Rein mode, which can be activated by the driver during the "Loose Rein" by taking the hands off the steering wheel when in fluid operation mode (Fig. 9). In Secured Rein, the automation drives completely autonomously, and the driver resorts to other activities. The idea is that in driving environments with special technical features, a temporarily fully automated mode on a so-called Secured Lane can be provided. In this case, the automation takes over full control, including the choice of the driving maneuver as well as the control of steering and acceleration.

If the ego-vehicle approaches another vehicle (Fig. 9), the automation adjusts the velocity to the velocity of the vehicle in front. If another parallel Secured Lane is available, the automation can initiate a lane change autonomously.

In "Secured Rein," the driver can switch to the partially/highly automated "Loose Rein" by either pressing the "Loose Rein" button on the display (cf. Figs. 2, 3, 4, 5, 6, 7, 8, 9 and 10) or by grasping the steering wheel with at least one hand in fluid operation mode Fig. 10, left). The automation understands this active grasping and fluidly transfers more control to the driver, whereas the driver is always able to cancel this control transition by taking the hands off the steering wheel again. Afterward, the driver can switch from "Loose Rein" to "Tight Rein." If the driver grasps the steering wheel with a tight grip or with both hands, the automation understands this and again transfers fluidly most of the control to the driver (Fig. 10, right).

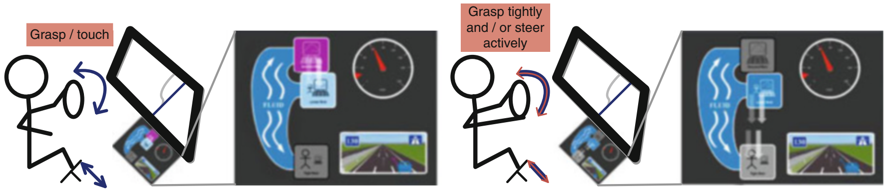


Fig. 10 Transitions: *left,* “Secured Rein” to “Loose Rein”; *right,* “Loose Rein” to “Tight Rein”

4 System Architecture and Mode of Operation

As an application of cooperative vehicle guidance and control, the H-Mode is based on the fundamental findings, principles, and mechanisms of the cooperative guidance and control of motion, as described in ▶ Chap. 59, “Cooperative Guidance, Control, and Automation”. For implementing the main features of the H-Mode, a combination of an adaptive cruise control system (ACC) and a lane keeping assistant system (LKAS) is already a good starting point. In this way, an integrated longitudinal and lateral automation can already be implemented, which – when extended by a suitable interaction concept – can result in the first step of the H-Mode. By using a cooperative system architecture, the potential of this concept can be exploited and implemented much more efficiently.

As introduced in the chapter on ▶ Chap. 59, “Cooperative Guidance, Control, and Automation,” such an architecture rests upon the two partners driver and automation, which are integrated into the entire control loop of the vehicle guidance and control according to predefined rules and procedures. In the preceding section, some exemplary use cases for this joint control have been depicted and the related aspects of human-machine interaction highlighted. In the following, the components of the H-Mode interaction and automation will be explained in more detail. Figure 11 shows cooperative vehicle guidance and control in the H-Mode as a schematic system chart on the basis of the generic cooperative control loop from ▶ Chap. 59, “Cooperative Guidance, Control, and Automation”.

4.1 Cognitive Automation in the H-Mode

The role model of the interaction of horse and rider in the metaphor points out that the partners in cooperative guidance and control are autonomous entities. The human just as well as her technical partner in the cooperation dispose of their own perception, cognition, and own means of realizing the driving task. Nevertheless, they both move together and complement one another’s skills. For implementing cooperative vehicle guidance and control in the H-Mode, a capable automation is the prerequisite. In the H-Mode, the automation for the cooperative vehicle guidance and control is a cognitive one, i.e., it is constructed in a way that is

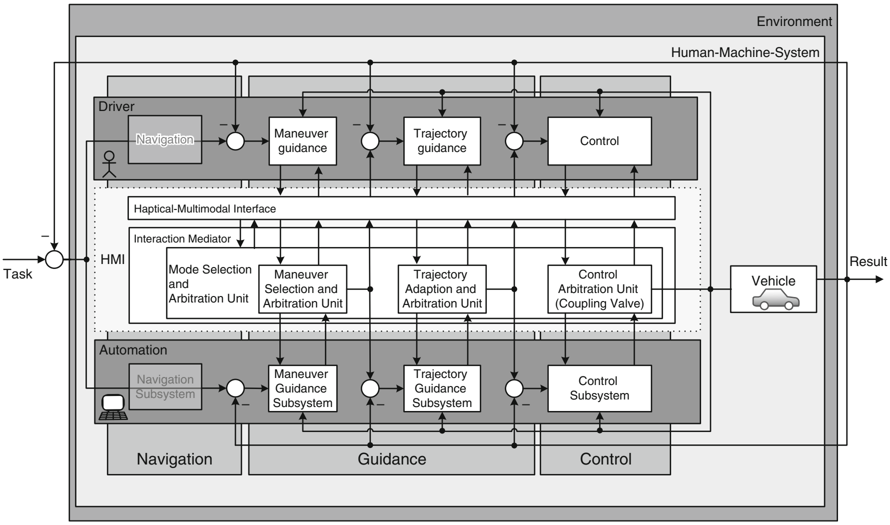


Fig. 11 Cooperative guidance and control with the H-Mode

comparable and compatible to a biological cognition. By doing so, one can achieve that the automation is designed to be internally and externally compatible as well as cooperative with the human driver (Löper et al. 2008; Flemisch et al. 2012a) . A detailed description of internal and external compatibility can be found in ▶ Chap. 59, "Cooperative Guidance, Control, and Automation" as well as in Flemisch et al. (2008b, 2012a), Bubb (1993).

Cognitive automation in the H-Mode is conceptualized on the basis of models of human cognition in vehicle guidance and control (Donges 1982; Löper et al. 2008). This can be achieved using a multilayered approach which considers several planning and execution horizons. For this purpose, one can resort to established approaches for human cognition in the context of automated technical systems (e.g., Rasmussen 1983; Parasuraman et al. 2000). The H-Mode hereby assumes a four-layered model of vehicle guidance and control with levels for navigation, maneuver planning and execution, trajectory planning, and control. Hereby, the three-layered model by Donges (1982) is extended in the sense that the guidance level is subdivided into two levels, namely, the maneuver and the trajectory level (Löper et al. 2008). Maneuver as a term of high semantic content describes the relation of temporally and spatially interconnected processes. A typical example for a maneuver in the H-Mode is "lane change to the left." Trajectories (e.g., Fig. 6) as planned tracks defined in spatial and temporal terms illustrate the practical implementation of corresponding maneuvers, which are adjusted to the environment (Altendorf and Flemisch 2014). This four-layered model of vehicle guidance and control serves as a basis for the cognitive automation employed in the H-Mode (cf. Fig. 11).

4.2 Interaction Mediation and Arbitration

In cooperative guidance and control, exemplified by the horse-and-rider metaphor, authority, responsibility, and control are distributed between the partners. In order to fulfill the requirements, an effective and compatible interaction between human and automation must be ensured, which in H-Mode is realized by the “Interaction Mediator.” Depending on the current situation, the Interaction Mediator designs and mediates interaction between human and automation.

The major tasks of the interaction mediation are the arbitration of conflicts as well as a continuous and unambiguous communication of intentions and states. An interaction mediation consists of several modules, e.g., a module for the selection and arbitration of the current level of automation (mode), where the distribution of responsibility and control depending on currently available authority, resources, and abilities is arbitrated. Conditional on the arbitrated level of automation, the interaction is customized for the specific planning level of vehicle guidance and individually designed in order to enhance the inner compatibility between automation and human (Baltzer et al. 2014) .

4.2.1 Arbitration of Conflicts

In the same manner as conflicts emerge during the interaction between two humans or, as suggested by the H-Metaphor, between a human and an animal, they are expected to emerge between human and automation during cooperative guidance and control of vehicles. These conflicts do not only emerge during the execution of actions but already during the planning phases on the different vehicle guidance levels. In order to prevent incapacity, conflicts need to be arbitrated as early as possible. “Human – machine arbitration includes a structured negotiation between a human and an automation with the intention to reach a common unambiguous decision on how to act in due course of time (Baltzer et al., 2014; Kelsch et al., 2006). During the arbitration process, the human and the automation are continuously coupled (haptic-multimodal). Also in uniquely defined exceptional cases, a decoupling of the respective partner can be useful, e.g., a decoupling of the human during a collision hazard reliably detected by the automation (Fig. 3) or, respectively, a decoupling of the automation during a malfunction reliably detected by the human.

In order to negotiate conflicts fast and effectively, the concept of action supporting and inhibiting interaction components are deployed. The basis of supporting and inhibiting interactions are tension fields that were concretized into the concept of “action tension” (Kelsch et al. 2012). Action tension is the directed motivation to a specific action (Kelsch et al. 2012).

In H-Mode 2D, the negotiations of different intentions of the human and the automation are conducted within the Interaction Mediator (Baltzer et al., 2014). These negotiations run multimodaly, both for the dynamic distribution of control between human and automation and the process of the driving task on different planning levels of vehicle guidance. Therefore, the Interaction Mediator is subdivided into different modules: On the highest level, the negotiation of control distribution is processed within the “Mode Selection and Arbitration Unit” (extension of the MSU presented in Baltzer et al. (2014)). On a lower level, the respective activities are processed in separate arbitration units for the different vehicle guidance and control levels: the maneuver to be performed in the “Manoeuvre Selection and Arbitration Unit,” the adaption of available driving trajectories in the “Trajectory Adaption and Arbitration Unit,” and the performance of control in the “Control Arbitration Unit” (Coupling Valve) (Baltzer et al., 2014).

4.2.2 Communication of Intentions and System States

Besides the arbitration of negotiation conflicts, the interaction mediator performs the task to communicate intentions and states explicitly and continuously between human and automation. By doing so, it needs to be ensured that driver and automation are constantly aware of which instance takes over which task, whether sufficient resources are available for accomplishing this task, and whether and with what time horizon control is really executed, in order to avoid control deficits or control surpluses.

Furthermore, both human and machine should know which intentions the other partner is currently pursuing and which action will soon be executed, e.g., which trajectory should be followed at a junction (Fig. 8) or whether an interaction pattern is activated because of a specific critical situation, such as preventing departures from the lane (virtual gravel trap, Fig. 4) or preventing a collision (Fig. 7).

An important means for interaction is the haptic interaction resource, which is mainly realized by the interaction between driver and automation via active interfaces in the H-Mode. Thereby, from an ergonomic point of view it can be sensible to implement the two-dimensional driving task with an interaction concept with two degrees of freedom, such as a sidestick. On the other hand, the H-Mode can also be sensibly transferred to an active steering wheel and an active foot throttle (Kienle et al. 2013). Whether stick or steering wheel/foot throttle, the haptic channel facilitates a continuous and directed interaction, which simplifies arbitration between driver and automation. Haptic coupling reduces reaction times (Brandt et al. 2007; Suzukia and Jansson 2003) and can increase the situational awareness (Flemisch et al. 2003; Abbink et al. 2008). Another important component in the communication of states is the visual representation of trajectories. Visualized trajectories enable the human to recognize and influence upcoming driving activities of the automation, because direct and stable feedback on initiated adaptations as well as initiated activities can be transmitted (Baltzer et al., 2014).

4.2.3 Transition Between Levels of Automation

Another focus of studies concerning the H-Mode is on the dynamically balanced transfer of distribution of control between driver and vehicle automation – so-called transitions (cf. Fig. 12). The H-Mode encompasses up to three degrees of automation between which one can switch dynamically during the ride (s. Fig. 1). One alternative of a dynamically balanced transition is the fluid transition, which has been described in one of the preceding sections. Hereby, the shift in the degrees of

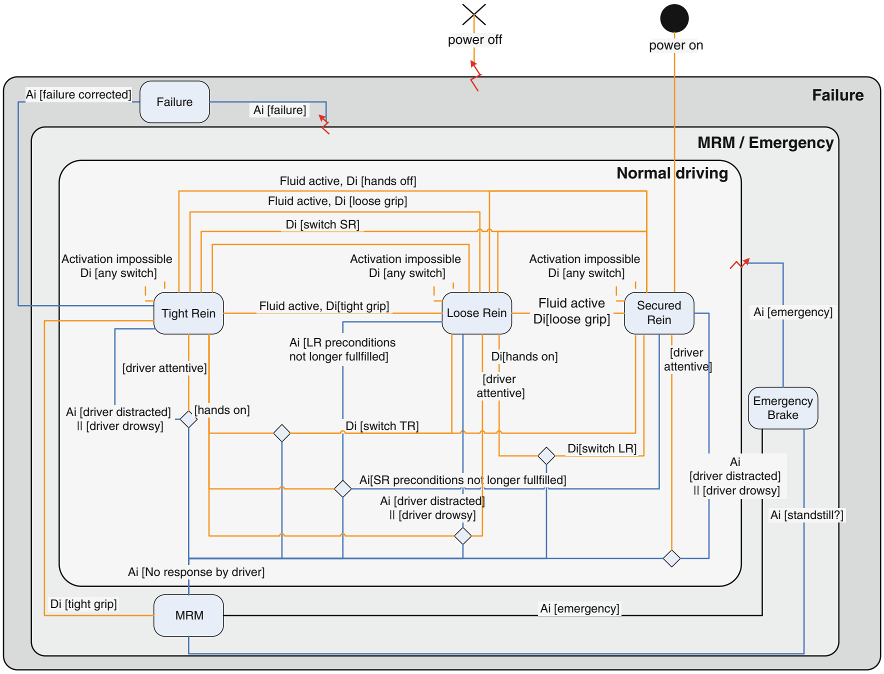


Fig. 12 State machine of the Mode Selection and Arbitration Unit according to (Baltzer et al., 2014)

automation does not take place by explicitly choosing a certain degree of automation, but happens implicitly by activity or involvement of the driver in the driving task, e.g., by grasping the haptic interface more loosely or more firmly. The transition between both states of control distribution is hereby not abrupt, but fluid. As soon as the human takes her hands completely off the haptic interface, this fluid form of interaction initiates the transition to the “Secured Rein.” When the human grasps the haptic interface lightly (detection via capacitive sensors), a transition to the “Loose Rein” is initiated, and when grasping the haptic interface distinctly (detection via force-sensitive resistors), a transition to the “Tight Rein” is initiated. Because of the conflict that in “Tight Rein” a very precise steering is needed from the human and that this requires the highest grip forces, the grip forces necessary for the transition have been defined as hysteresis, such that the grasp after a transition, e.g., after the Tight Rein, can be reduced to a more comfortable level. Besides these normal transition possibilities, so-called emergency transitions are implemented in the current prototype. In case the human exerts very high forces on the haptic interfaces, this will be interpreted in such a way that the human wants to resume control in both interaction forms, the fluid one and the one using push buttons. The transition happens immediately, in order not to lose time when facing a threat which has remained undetected by the sensor systems of the automation.

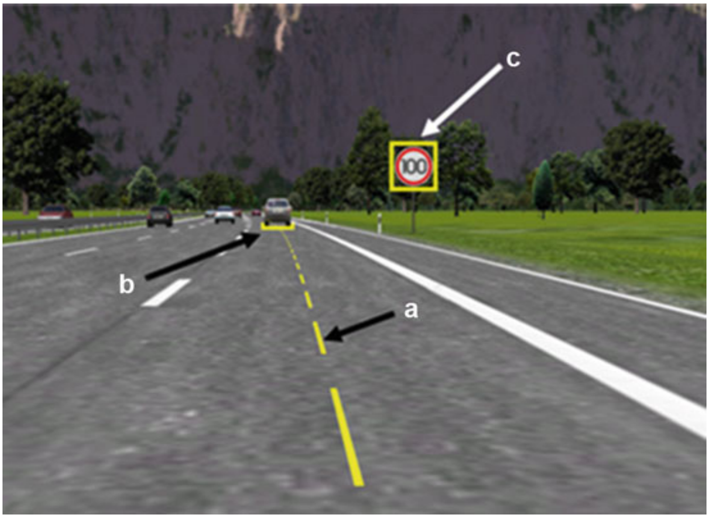


Fig. 13 Display symbols: (**a**) trajectory, (**b**) bracket, and (**c**) frame in the head-up display

4.3 Interaction Modalities in Cooperative Driving

An important component of the interaction concept for automated systems is the feedback to the driver. Preferably, this should be carried out within 200 ms and with a multimodal design (Bubb 1993; Bengler et al. 2012). Thereby, safe and fast processing is decisive; the information from different channels must not be contradictory and must not puzzle the driver but complement each other. This means that the haptic channel should inform the driver about which action is preferred, while the visual channel explains how this judgment has been reached. It is strongly believed that this combination has a positive influence on acceptance and system comprehension (Lange 2008). While the haptic feedback is supposed to especially support a quick reaction from the driver (Schieben et al. 2008), information on the visual channel can be more directed and more extensive. Furthermore, using the diversity of visual representations, the numerous functions of the automation can be indicated more explicitly. Thereby, the driver can permanently retrieve the state and the intention of the system. At the same time, the driver needs to be informed about the system boundaries to be able to identify malfunctioning in time (also cf. Vicente and Rasmussen 1992). Figure 13 shows a corresponding design: The vehicle driving ahead on the same lane is highlighted by a yellow bracket. The identified lane is marked by a trajectory placed in the middle. Identified traffic signs are framed analogously. Every display symbol visualizes a state of a subsystem in the entire automation. With the use of this display a driver can recognize erroneous or missing identifications early in time and can punctually initiate countermeasures.

5 Case Studies and Study Results

Overall, the H-Mode has been well accepted in simulator studies. Participants of an experimental study ($n = 20$), which has been conducted in 2013, evaluated the H-Mode on a seven-point scale with the second highest grade as “pretty good.” Also, usefulness, perceived cooperation between driver and automation, ease of driving, as well as perceived security were considered positively. Especially driving in the partially/highly automated domain of the “Loose Rein” is perceived as “pretty comfortable” by users. Also the highly/temporarily full automated mode “Secured Rein” is mainly well accepted (Fig. 14, Meier et al. 2013). In another study, the H-Mode versions with steering wheel as well as the alternative with active sidestick have been rated highly (Fig. 15, Flemisch et al. 2012b).

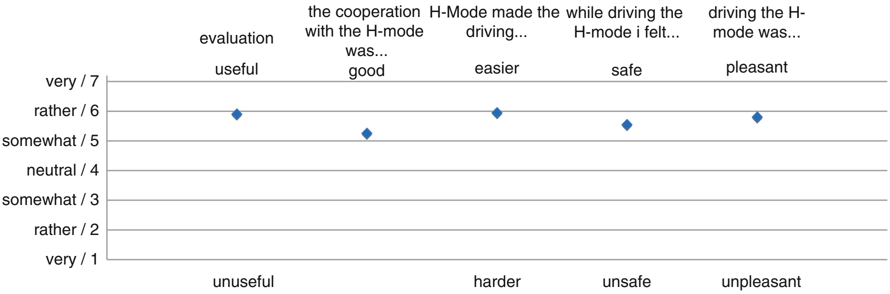


Fig. 14 Evaluation of the H-Mode, usability study 2013 (Meier et al. 2013)

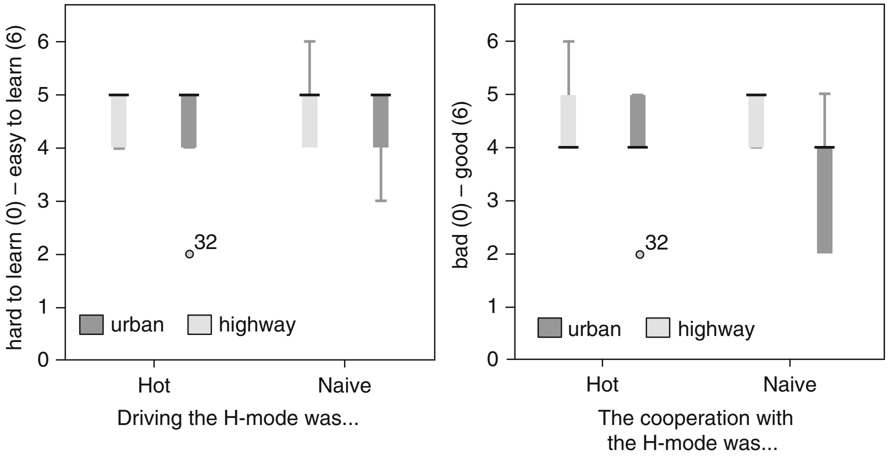


Fig. 15 Evaluation of the H-Mode, usability study 2011 (Flemisch et al. 2012b)

Beyond this, a further study could prove that the matching of haptic and visual feedback gives the driver the opportunity for fast-correcting actions in the case of system malfunctions. When driving with contact analogue displays, test persons could recognize an erroneous identification of the adjacent lane earlier and were therefore able to intervene more timely. Thereby, the use of a contact analogue display produced a much lower maximal lateral offset during the ride, compared to a situation without any visual feedback (Weißgerber et al. 2012; Damböck et al. 2012).

In addition to theoretical considerations of the subconcepts of the H-Mode, “Tight Rein”, “Loose Rein” as well as the interaction for specific maneuvers have been experimentally investigated. The focus was on the evaluation of the effects of varying degrees of driver involvement in “Tight Rein,” “Loose Rein,” and full automation on the controllability of automation breakdowns and malfunctions as well as the effect on the distribution of visual attention resources. The essential finding was that a well-calibrated, i.e., not too highly automated, “Loose Rein” can lead to an unburdening of the driver with freeing up visual resources for executing subsidiary tasks, while retaining an acceptable level of controllability in the case of an automation breakdown. This constitutes a fundamental advantage in comparison to full automation without driver involvement (Flemisch et al. 2008b; Heesen et al. 2009). Also, the interaction concepts derived from the H-metaphor can help to improve controllability for transitions at the system boundaries (Heesen et al. 2011; Schwarzmaier 2011). Another emphasis was put on the investigation of the negotiation between driver and automation (arbitration) (Griesche et al. 2012; Kelsch et al. 2013). This aspect has been examined at junctions where the automation and the driver had chosen different routes. In the investigated configuration, establishing the driver’s preference was possible without further ado in both “Tight Rein” and “Loose Rein,” and the comprehension of the intentions of the automation was high. Further studies have been conducted in the field of haptic-multimodal interaction with scenarios for approaching a vehicle in front, for a lane change, for emergency braking, and for avoiding obstacles. Thereby, haptic lane change recommendations, haptic cues on an active foot throttle, and partially/temporarily fully decoupling (Flemisch et al. 2010a) the driver from longitudinal and/or lateral control in emergency, evasion, and braking situations have been successfully examined and proven to be effective (Heesen et al. 2010; Kelsch et al. 2009).

In a comprehensive usability study, fluid transition has been investigated in a realization as active steering wheel as well as active sidestick in many different driving situations. The goal was to find out how well test persons can drive with a fluid transition between two degrees of automation, whether assisted driving is easy to learn, and whether implicitly executed transitions were disturbing or helpful in carrying out the driving task. The test persons were driving on a round course in a dynamic simulator, which has been composed of several driving situations. Several situations on a highway, consisting of driving on straight routes, cornering routes in varying turn radii, lane change, and braking scenarios have been mapped. Furthermore, test persons were driving in a city scenario with turning situations, junctions, and narrow turns. For measuring training and learning effects, the experiments began with a “Naive Run,” a ride without any prior knowledge about the H-Mode system. This was followed by a so-called Hot Run, a second ride with respective system experience. Among others, this showed that driving in the H-Mode with fluid transition, irrespective of the nature of the actuator, often already in the Naive Run, at the latest however in the Hot Run, was perceived as easy to learn by drivers and the cooperation of driver and automation was perceived as high.

6 Conclusion and Outlook

H-Mode, as an instance of cooperative guidance and control, is inspired by the role model of horse and rider. It offers an integrated haptic-multimodal user interface for all kinds of movement control. The H-Mode 2D, as presented in this chapter, has been designed for ground vehicles and includes several comfort and security systems on three assistance and automation levels, which can be interchanged fluidly.

A fourth automation level, semiautomated/highly assisted/ACC+, can be added, if required for migration purposes (Fig. 1). The H-Mode can also be applied when temporarily fully automated driving is not available. In situations where the whole transportation system is in a safe state, partial automation, such as the Loose Rein, is only a transition stage to higher automated driving modes. It is also thinkable that the societal consensus develops in such a way that conventional manual driving will not be offered any longer because of safety reasons, e.g., in certain areas or under certain environmental conditions. This might sound futuristic from today’s perspective, but it is already reality in aviation, where, e.g., on sufficiently equipped airports only highly automated landings with ILS (instrument landing system) are allowed in the case of bad sight.

An advantageous combination of the H-Mode with an attention monitor, as has been shown in the HAVEIt project (Hoeger et al. 2011), can also limit a potential misuse in the partially or highly automated mode (“Loose Rein”) (Flemisch et al. 2010b). H-Mode in the form illustrated in this chapter can be implemented with a conventional steering wheel and an active foot throttle but also with an active sidestick, smaller fingersticks, or active steering wheels with a combined foot throttle, as has been successfully examined (e.g., Flemisch et al. 2012b).

The insights from the H-Mode and cooperative automation in vehicle guidance and control can also be transferred to other domains in which machine and human cognitive abilities work together cooperatively. H-Mode has already been extended to three-dimensional movement processes toward the H-Mode 3D, which allows its application in the domain of aviation (Goodrich et al. 2006).

With increasing interconnection of human, vehicle, and infrastructure, a multitude of challenges and chances emerge for the application of cognitive and cooperative human-machine systems. Cooperative vehicle guidance and control of a single vehicle offers the basis for a systemic extension to a larger cooperation network.

References

Abbink DA, Boer ER, Mulder M (2008) Motivation for continuous haptic gas pedal feedback to support car following. In: Intelligent vehicles symposium, Institute of Electrical and Electronics Engineers (IEEE), Piscataway, pp 283–290

Altendorf E, Flemisch F (2014) Prediction of driving behavior in cooperative guidance and control: a first game-theoretic approach. CogSys, Madgeburg

Baltzer M, Flemisch F, Altendorf E. Meier S (2014) Mediating the interaction between human and automation during the arbitration processes in cooperative guidance and control of highly automated vehicles. In: Ahram T, Karwowski W, Marek T (eds) Proceedings of the 5th international conference on applied human factors and ergonomics AHFE 2014, Kraków, 19–23 July 2014

Bengler K, Zimmermann M, Bortot D, Kienle M, Damböck D (2012) Interaction principles for cooperative human-machine systems. Inf Technol 54(4):157–163

Brandt T, Sattel T, Böhm M (2007) Combining haptic human-machine interaction with predictive path planning for lane-keeping and collision avoidance systems. In: Proceedings of the 2007 intelligent vehicles symposium, Institute of Electrical and Electronics Engineers (IEEE), pp 582–587

Bubb H (1993) Systemergonomische Gestaltung (System-ergonomic design). In: Schmidtke H (ed) Ergonomie. Carl Hanser, München

Damböck D, Weißgerber T, Kienle M, Bengler K (2012) Evaluation of a contact analog head-up display for highly automated driving. In: 4th international conference on applied human factors and ergonomics, San Francisco, pp 6011–6020

Donges E (1982) Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen (Aspects of active safety when driving a passenger vehicle). Automob Ind 27:183–190

Flemisch FO, Adams CA, Conway SR, Goodrich KH, Palmer MT, Schutte MC (2003) The H-metaphor as a guideline for vehicle automation and interaction. NASA Langley, Report no. NASA/TM-2003-212672

Flemisch FO, Kelsch J, Löper C, Schieben A, Schindler J (2008a) Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In: de Waard D, Flemisch FO, Lorenz B, Oberheid H, Brookhuis KA (eds) Human factors for assistance and automation. Shaker, Maastricht, pp 1–16

Flemisch F, Kelsch J, Löper C, Schieben A, Schindler J, Heesen M (2008b) Cooperative control and active interfaces for vehicle assitsance and automation. FISITA World Automotive Congress, München

Flemisch F, Nashashibi F, Glaser S, Rauch N, Temme T, Resende P, Vanholme B, Schieben A, Löper C, Thomaidis G, Kaussner A (2010a) Towards a highly automated driving: intermediate report on the HAVEIt-joint system. Transport Research Arena, Brussels

Flemisch F, Heesen M, Kelsch J, Schindler J, Preusche C, Dittrich J (2010b) Shared and cooperative movement control of intelligent technical systems: sketch of the design space of haptic-multimodal coupling between operator, co-automation, base system and environment. In: Proceedings of 11th IFAC/IFIP/IFORS/IEA symposium on analysis, design, and evaluation of human-machine systems, Valenciennes

Flemisch F, Meier S, Neuhofer J, Baltzer M, Altendorf E, Ozyurt E (2012a) Kognitive und kooperative Systeme in der Fahrzeugführung: Selektiver Ruckblick über die letzten Dekaden und Spekulation über die Zukunft (Cognitive and cooperative systems in vehicle guidance: review of the last decades and a glimpse into the future). In: Workshop Kognitive Systeme, 2. Interdisziplinarer Workshop Kognitive Systeme: Mensch, Teams, Systeme und Automaten

Flemisch F, Meier S, Baltzer M, Altendorf E, Heesen M, Griesche S, Weißgerber T, Kienle M, Damböck D (2012b) Fortschrittliches Anzeige- und Interaktionskonzept für die kooperative Führung hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse mit H-Mode 2D 1.0 (An advanced display structure and interaction concept for cooperative guidance of highly automated vehicles: selected results of H-mode 2D 1.0) 54. Fachausschusssitzung

Anthropotechnik: Fortschrittliche Anzeigesysteme für die Fahrzeug- und Prozessführung, Koblenz

Flemisch, Frank Ole, et al. (2014) Towards cooperative guidance and control of highly automated vehicles: HMode and Conduct-by-Wire. Ergonomics 57(3): 343–360

Gasser TM, Arzt C, Ayoubi M, Bartels A, Bürkle L, Eier J, Flemisch F, Häcker D, Hesse T, Huber W, Lotz C, Maurer M, Ruth-Schumacher S, Schwarz J, Vogt W (2012) Rechtsfolgen zunehmender Fahrzeugautomatisierung – Gemeinsamer Schlussbericht der Projektgruppe Bundesanstalt für Straßenwesen (BASt)

Goodrich K, Flemisch F, Schutte P, Williams R (2006) A design and interaction concept for aircraft with variable autonomy: application of the H-mode; digital avionics systems conference

Griesche S, Kelsch J, Heesen M, Martirosjan A (2012) Adaptive Automation als ein Mittel der Arbitrierung zwischen Fahrer und Fahrzeugautomation (Adaptive automation as an approach of arbitration between driver and automation) AAET – Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel, Braunschweig, 8–9 Feb 2012

Heesen M, Schieben A, Flemisch F (2009) Unterschiedliche Automatisierungsgrade im Kraftfahrzeug: Auswirkungen auf die visuelle Aufmerksamkeit und die Kontrollübernahmefähigkeit (Different levels of automation in motor vehicles: impacts on visual attention and on ability to take control) 51. Tagung experimentell arbeitender Psychologen, Jena, 29 Mär–01 Apr

Heesen M, Kelsch J, Löper C, Flemisch F (2010) Haptisch-multimodale Interaktion für hochautomatisierte, cooperative guidance and control bei Fahrstreifenwechsel-, Brems- und Ausweichmanövern (Haptic-multimodal interaction for highly automated cooperative guidance and control during lane change, braking and evasive maneuvers) 11. Braunschweiger Symposium Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel (AAET), Braunschweig, 10–11 Feb 2010

Heesen M, Beller J, Griesche S, Flemisch F (2011) Shake it!: Intuitive haptische Interaktion zur Übergabe der Kontrolle von einer Automation zum Fahrer (Shake it!: intuitive haptic Interaction during control transitions.) In: Beiträge zur 53. Tagung experimentell arbeitender Psychologen, Seite 67. Pabst. Tagung experimentell arbeitender Psychologen, Halle, ISBN 978-3-89967-698-3

Hoeger R, Zeng H, Hoess A, Kranz T, Boverie S, Strauss M, Jakobsson E, Beutner A, Bartels A, To TB, Stratil H, Fürstenberg K, Ahlers F, Frey E, Schieben A, Mosebach H, Flemisch FO, Dufaux A, Manetti D, Amditis A, Mantzouranis I, Lepke H, Szalay Z, Szabo B, Luithardt P, Gutknecht M, Schömig N, Kaussner A, Nashashibi F, Resende P, Vanholme B, Glaser S, Allemann P, Seglö F, Nilsson A (2011) Deliverable D61.1 final report. HAVEit: Highly automated vehicles for intelligent transport

Kelsch J, Heesen M, Löper C, Flemisch F (2009) Balancierte Gestaltung kooperativer multimodaler Bedienkonzepte für Fahrerassistenz und Automation: H-Mode beim Annähern, Notbremsen, Ausweichen (Balanced design of cooperative multimodal HIS concepts for driver assistance and automation: H-mode during approaching, emergency brakes and avoiding) 8. Berliner Werkstatt MMS, Berlin, 7–9 Okt 2009

Kelsch J, Heesen M, Hesse T, Baumann M (2012) Using human-compatible reference values in design of cooperative dynamic human-machine systems. In: Contribution to 30th European annual conference on human decision-making and manual control (EAM 2012), Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut of Transportation Systems, Braunschweig

Kelsch J, Temme G, Schindler J (2013) Arbitration based framework for design of holistic multimodal human-machine interaction. In: Beiträge zur AAET 2013, AAET 2013, ISBN 9783937655291, Seiten 326–346, 6–7 Feb 2013

Kienle M, Damböck D, Bubb H, Bengler K (2013) The ergonomic value of a bidirectional haptic interface when driving a highly automated vehicle. Cogn Tech Work 15(4):475–482

Lange C (2008) Wirkung von Fahrerassistenz auf der Führungsebene in Abhängigkeit der Modalität und des Automatisierungsgrades (Dissertation) (Influence of driver assistance depending on the interaction-modality and the level of automation), Technische Universität München, Garching

Löper C, Kelsch J, Flemisch FO (2008) Kooperative, manöverbasierte Automation und Arbitrierung als Bausteine für hochautomatisiertes Fahren (Cooperative, maneuver based automation and arbitration as pices for highly automted driving). In: Gesamtzentrum für Verkehr Braunschweig (ed) Automatisierungs-, Assistenzsysteme und eingebettete Systeme für Transportmittel. GZVB, Braunschweig, pp 215–237

Meier S, Altendorf E, Baltzer M, Flemisch F (2013) Partizipative Interaktions- und Automationsgestaltung teil- bis hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse explorativer Nutzerstudien zu H-Mode 1.1 (Participatory design of interaction and automation of partially and highly automated vehicles: selected results of explorative user studies on H-mode 1.1). In: Brandenburg E, Doria L, Gross A, Günzler T, Smieszek H (eds) Proceedings of the 10. Berliner Werkstatt Mensch-Maschine-Systeme: Grundlagen und Anwendungen der Mensch-Maschine-Interaktion, Zentrum Mensch-Maschine-Systeme, Berlin, pp 461–468

Parasuraman R, Sheridan TB, Wickens CD (2000) A model for types and levels of human interaction with automation. IEEE Trans Syst Man Cybern Syst Hum 30:286–297

Rasmussen J (1983) Skills, rules and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans Syst Man Cybern 3:257–266

Schieben A, Damböck D, Kelsch J, Rausch H, Flemisch F (2008) Haptisches Feedback im Spektrum von Fahrerassistenz und Automation (Tactile Feedback in the spectrum of Driver Assistance and Automation) 3. Tagung “Aktive Sicherheit durch Fahrerassistenz”, Garching

Schwarzmaier M (2011) Unterschiede haptischer und visueller Hinweisreize an Systemgrenzen (Differences between haptic and visual signals at system boundaries) Bachelorarbeit, TU Braunschweig

Suzukia K, Jansson H (2003) An analysis of driver’s steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. JSAE Rev 24(1):65–70

Toffetti A, Wilschut ES, Martens MH, Schieben A, Rambaldini A, Flemisch F (2009) CityMobil: human factor issues regarding highly automated vehicles on eLane. Transportation Research Record: Journal of the Transportation Research Board, No. 2110, Transportation Research Board of the National Academies, Washington, DC, pp 1–8

Vanholme B (2012) Highly automated driving on highways based on legal safety. Dissertation, University of Evry-Val-d’Essonne

Vicente KJ, Rasmussen J (1992) Ecological interface design: theoretical foundations. IEEE Trans Syst Man Cybern 22(4):589–606

Weißgerber T, Damböck D, Kienle M, Bengler K (2012) Erprobung einer kontaktanalogen Anzeige für Fahrerassistenzsysteme beim hochautomatisierten Fahren (Testing a contact analogue display when driving highly automated). In: TÜV SÜD Akademie GmbH (ed) 5. Tagung Fahrerassistenz. Schwerpunkt Vernetzung, München

Autonomous Driving

62

Richard Matthaei, Andreas Reschka, Jens Rieken, Frank Dierkes,
Simon Ulbrich, Thomas Winkle, and Markus Maurer

Contents

R. Matthaei (✉) • A. Reschka • J. Rieken • F. Dierkes • S. Ulbrich • M. Maurer
Institut für Regelungstechnik, Technische Universität Braunschweig, Braunschweig, Germany
e-mail: matthaei@ifr.ing.tu-bs.de; reschka@ifr.ing.tu-bs.de; rieken@ifr.ing.tu-bs.de; dierkes@ifr.ing.tu-bs.de; ulbrich@ifr.ing.tu-bs.de; maurer@ifr.ing.tu-bs.de

T. Winkle
Daimler and Benz Foundation, Ladenburg, Germany
e-mail: winkle@carforensic.com

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_61

Abstract

In this chapter, a survey of the current state of research on autonomous driving is given and is set in the context of the requirements of an autonomous vehicle following the vision of an automated taxi. The overview is based on (scientific) publications and self-reports of the developing teams. Aspects of interest for this summary are approaches on environmental perception, self-perception, mission accomplishment, localization, cooperation, map usage, and functional safety.

Typically, emphasis is given to reliance on global satellite systems (e.g., GPS) and map data. Only a few approaches focus on environmental perception and scene understanding. Even though impressive demonstrations of autonomous driving have been presented in recent decades, this overview concludes that many aspects still remain only partially solved or even unsolved, especially when driving autonomously in public road traffic.

1 Introduction

1.1 Motivation

At present everyone is talking about the vision of "autonomous driving." The media report on successes derived from research, with numerous promises regarding product launches taking place in the near future; automotive companies are competing in their race to develop new technologies, and software companies are competing with the vehicle manufacturers. Subsequently, the hope for accident-free driving in society is born. We would then at last be able to lean back and relax – even during the course of our journey – and enjoy the trip, make phone calls to our hearts' content, surf the Internet, or do some preparatory work in advance instead of grumbling about the traffic jam. The elderly and the sick would then be able to enjoy enhanced, independent mobility in the long term, and autonomous vehicles would also contribute towards a more efficient use of the raw materials. It would be feasible for car-sharing offers to have their vehicles driven autonomously to the customers (cf. Laurgeau 2012) or have these vehicles powered independently. By taking the flow of traffic and the entire stretch of the journey into consideration, the actual trip itself could also be optimized in terms of energy.

Is all this still a vision that lies in the distant future? Will there ever be "autonomous road vehicles"? Or are they going to be launched on the market very soon? What in fact does "autonomous driving" actually mean? What are the technical challenges that have to be solved?

The projections regarding a market launch vary; the options are all included, with forecasts from 10 to 20 years to never at all. This is also due to the fact that we have no standardized understanding of the range of functions of an "autonomous vehicle." It will not be possible to answer any far-reaching questions in this chapter. Nevertheless, the attempt is made to explain the term "autonomous driving" more clearly by providing a short historical introduction and functional definition. Build up on this, the results of an intensive investigation on research groups that concentrate on autonomous driving in the broader sense are assessed. Our focus is thereby restricted to the technical aspects that will have only a small influence on whether autonomous vehicles will be involved in the public road traffic of the future. The following open questions are not dealt with in depth in this article: legal ambiguities ("Under what conditions can a registration take place?," "Who is liable in the event of damage?"), acceptance in society ("Will people trust machines that have the potential of inflicting deadly injuries on them?"), safeguarding ("How can we ensure that the autonomous vehicle will safely master every possible situation, i.e., perceive the situation and evaluate it?") (Maurer 2013), or service and maintenance during everyday operation ("Who will be carrying out regular checks on roadworthiness and operational safety in the future, as even today, hardly anyone carries out a visual check of the safety-relevant components of his vehicle before commencing a journey?").

1.2 Selected Historic Approaches towards Autonomous Driving

The development of autonomous vehicle technology began in the 1950s according to Fenton (1970). The first ideas of how to automate highway driving were developed in the General Motors Research Lab. Due to obvious limitations in computation and image processing in this era, a combination of vehicle sensors and infrastructure measures promised the best results. For example, an approach to integrate magnets into the road surface was developed and its feasibility was demonstrated (Fenton 1970).

According to Dickmanns (2015), the first approach for vision-based high-speed driving in road vehicles was developed at the Universität der Bundeswehr in München for the vehicle VaMoRs. In 1986, VaMoRs was able to drive on a closed test track and in 1987 on a closed Autobahn up to 90 km/h (Zapp 1988; Dickmanns 2007). In the famous PROMETHEUS project (PROgraMme for a European Traffic of Highest Efficiency and Unprecedented Safety, 1987–1994), many approaches of participants from European partners were developed. Within this project, research was continued at the Universität der Bundeswehr, this time in cooperation with Daimler-Benz AG. In the project's context, the vehicle VaMoRs-P (VaMP) of Universität der Bundeswehr in München demonstrated automated driving in France on public highways (Dickmanns et al. 1994, Maurer 2000). Similar results were shown with the companion vehicles VITA and VITA II by the Daimler-Benz AG (Ulmer 1992, 1994). A special event was an automated long-distance run from Munich to Odense and back with VaMP in 1995. The vehicle automatically drove 1678 km of the 1758 km while the system was switched on. In this demonstration, the lateral and longitudinal control was automated, and speeds up to 180 km/h were reached. Additionally, lane changes were triggered by the safety driver and executed automatically by the technical system (Maurer 2000).

In the 1980s, Japanese groups did research on image recognition, algorithms for lane marking detection, and object detection. Based on this acquired data, vehicles were automated (Tsugawa 1993, 1994). The demonstrations included driving functions like adaptive cruise control (ACC) and lane-keeping support at relatively low vehicle speeds. Alongside this, vehicle automation based on infrastructure improvements was investigated further to reduce complexity for onboard technology by combining vehicle and infrastructure technology (Hitchcock 1995; Zhang et al. 1990).

Within the project California Partners for Advanced Transit and Highways (PATH) (Shladover 2007) and at the Carnegie Mellon University in Pittsburgh (Thorpe et al. 1997), automated vehicles were developed and pioneering achievements shown with prototypes starting from the 1980s as well. A major milestone was the demonstration drive "No hands across America" in 1995. The experimental vehicle NavLab 5 crossed the USA mostly automatically. The developed system controlled the vehicle laterally in 4500 out of 4587 km driven on highways (Pomerleau and Jochem 1996). A safety driver was responsible for the longitudinal control of the vehicle.

The research vehicle ARGO from the VisLab institute of the Università degli Studi di Parma was used for long-distance drives in 1998 (Broggi et al. 1999). In a few days, 1860 km was driven automatically. The functionality was similar to VaMP, including longitudinal and lateral vehicle control and lane changes triggered by a safety driver and executed by the system (Broggi et al. 1999).

In all of these projects, the system was monitored by a human driver. The operational environment was mostly limited to highways. Referring to currently discussed categories of vehicle automation in Gasser et al. (2012) and SAE (2014), none of the experimental vehicles fulfill all requirements for high or full automation.

From 2004 to 2007 the research activities were pushed by several challenges started by the Defense Advanced Research Projects Agency (DARPA) in the USA. DARPA set the goal of developing driverless, automated vehicles for military purpose in the early 2000s. To reach this goal, DARPA started the first DARPA Grand Challenge in 2004. The task was to drive unmanned through a desert in Nevada. Because of a rather short development time, the results were not satisfactory: None of the experimental vehicles made it to the finish line (Thrun et al. 2006). In the following year, DARPA doubled the trophy and with a development time increased by 1 year, the participating vehicles were improved remarkably. Several teams developed vehicles which reached the finish line after driving 229 km through the desert unmanned. The winner of this Grand Challenge was Stanley from Stanford University. It finished first after about 7 h (Thrun et al. 2006). All vehicles in this competition were unmanned, but followed by a supporting vehicle which could stop them with a remote control. For a more detailed insight into the developed technology, Singh (2006) contains the approaches of the teams in the challenge.

The Grand Challenge 2005 was a great success, and consequently DARPA announced the 2007 DARPA Urban Challenge. Instead of driving off-road through the desert, autonomous driving technology was taken into an urban-like environment with streets, buildings, and cars driven by stunt drivers. This scenario challenged the teams with a much higher complexity of the operation environment. The teams had to solve missions which included following traffic rules, solving four-way-stop situations, and driving on a free navigation area. Some of the autonomous vehicles accomplished their mission and finished after several hours of autonomous urban driving. The result was a boost for autonomous driving technology, and follow-up projects are still developing new approaches in this field.

The Tartan Racing Team from Carnegie Mellon University won this challenge. The Stanford Racing Team finished second and the Team Victor Tango finished third. The scientific results of all teams in the final event were published in Singh (2008a, b, c). Again, all vehicles were unmanned and could be stopped by remote control from a supporting vehicle. This vehicle was following the autonomous vehicle, e.g., as presented by the CarOLO Team from the Technische Universität Braunschweig (Wille 2012). The experimental vehicle Caroline took part in the final event. This final confronted the vehicles with some unforeseen situations where human interference was necessary. These necessary human interventions showed that Caroline and the other autonomous vehicles were not yet ready for public traffic (Wille 2012).

In the subsequent years, many new projects have been started, and the vehicles from the DARPA Urban Challenge teams were developed further. Besides research institutions, vehicle manufacturers, suppliers, and other companies started the development of autonomous vehicles. Some of the most recent projects are part of this chapter. A more detailed look into technology is aggregated in the *Handbook of Intelligent Vehicles* (Eskandarian 2012).

1.3 Requirements for Autonomous Driving in Public Road Traffic

In this chapter an autonomous vehicle is understood (similarly to the definitions in Wachenfeld et al. (2015)) as a vehicle which is able to move “freely,” because it is not limited to rails, power supply lines, or a bus bar. Thus, the vehicle can be used in a more flexible way than a rail-mounted vehicle, for example, but has a limited amount of energy. This limited amount of energy, the limited installation space, and the appliance in the direct environment of human beings and animals lead to further requirements concerning the health effects of the applied technologies. Therefore, the amount of available technical solutions to this challenging task is further reduced.

The autonomous vehicle is operated by the human being on a very *intuitive* level, similar to the well-known interfaces of today’s satellite navigation systems, and is consequently provided at the most *abstract* level from the system’s point of view. This means, neglecting a service mode, the vehicle is only instructed by a mission.

Typically, a mission for an on-road vehicle consists of a transportation task. People, goods, or just the vehicle itself might be transported. In future systems, also surveillance and other tasks might be relevant for autonomous vehicles. In the case of the transportation of human beings, the mission must be adaptable to the current needs of the passengers at any time. Such an adaptation might be caused by triggering an emergency stop (see Wachenfeld et al. 2015) or by adding a stopover at a restaurant, the next bathroom, or a hospital. The overall functionality of an autonomous vehicle extends the definition of fully automated driving given by Gasser et al. (2012).

A special aspect of "autonomy" in this context is the self-motivated adaption of the current mission. Such an adaption of the mission might be to drive to a repair shop in the case of a self-diagnosed and not self-reparable defect or to stop at a gas station in the case of an empty tank or battery. Further elements defining the autonomy of the system are mechanisms of starting a "self-healing" process, e.g., restarting components (see Ghosh et al. 2007).

The appliance to the public road traffic increases the demands on an autonomous vehicle (in this case also automated vehicle) concerning both the environmental perception and the driving behavior. The urban environment in particular puts high demands on the environmental perception. It is necessary that the vehicle robustly detects and classifies the stationary elements (e.g., road course, signs, traffic lights) and the movable elements (e.g., traffic participants, human beings, animals). It is mandatory due to consistency reasons that human beings and technical systems use the same optical features for orientation as they share the same road environment (see, e.g., Bar Hillel et al. 2012; Huang et al. 2009).

In this case of mixed traffic (consisting of automated and manually driven vehicles), the locally defined road traffic regulations are of special interest. They define a minimal amount of environmental elements (signs, road markings, traffic participants, etc.) which have to be perceived and considered. Additionally, the regulations specify the behavior in defined situations (Wachenfeld et al. 2015). The basic components of the road traffic regulations are the mutual considerateness, a clear behavior pattern, as well as communication and cooperation.

In addition to these pure functional requirements, it is mandatory within the meaning of responsible acting that automated and autonomous vehicles do not constitute any danger to their environment. Therefore, the vehicle needs to be aware of its skills and abilities and has to act according to its current state. So the estimation of the skills and abilities including the surveillance of hard- and software is another mandatory requirement (onboard diagnostics). Moreover, the vehicle has to be resistant against misuse and manipulation.

In the euphoria of an increasing automation of the vehicles, the initial motivation should never be forgotten: The focus remains on human beings and their needs for individual mobility. An autonomous vehicle, which drives collision-free and even according to local road traffic regulations and moral and societal rules, but whose passengers do not trust the technical system and cannot enjoy a comfortable drive, would probably not be accepted by society.

1.4 Relevant Research Projects

This chapter focuses on civil and fully automated on-road motor vehicles, whose concepts are sufficiently published or whose detailed information was given directly to the authors by the relevant research team. Due to different requirements and thus different resulting technical solutions in civil or military applications (for military purposes, e.g., robustness against attacks or off-road navigation is also relevant), this chapter mainly examines civil systems. Additionally, only projects with the objective of realizing a highly or fully automated driving prototype according to Gasser et al. (2012) and SAE (2014) are part of the following discussion. Hence, research projects in the field of assisted or conditionally automated driving, whose objective always considers the driver as a surveillant and possible fallback solution, are explicitly not taken into account.

Nevertheless, even current prototypes for highly or fully automated driving currently reach only conditional automation in public road traffic according to our appreciation (see also Sect. 2.6).

Furthermore, only projects developing explicitly road vehicles, e.g., cars, commercial road vehicles, buses, or motor bikes, are considered in this chapter. This means the huge number of mobile robotic platforms, humanoid robots, or unmanned airplanes are not considered either. Projects, which are only announced or published by general media reports, cannot be discussed and compared to other projects due to missing scientific information. With the goal of giving the same chance to all teams to include their research results in this chapter, the authors developed a questionnaire (see Sect. 4). This questionnaire was sent to many research teams and institutions, automotive manufacturers, and other companies who are known by media reports and conferences.

The authors would herewith like to sincerely thank participants for the feedback.

Based on the aforementioned criteria, as well as the feedback from the questionnaire, the following list of projects, which are considered in this chapter, emerges:

- Special research project 28 of the DFG (Deutschen Forschungsgemeinschaft)
  - Karlsruher Institut für Technologie – experimental vehicle AnnieWAY
  - Universität der Bundeswehr – experimental vehicle MuCAR-3
  - Technische Universität München – experimental vehicle MUCCI
- BMW AG – Project Connected Drive
- VisLab Institut der Università degli Studi di Parma – experimental vehicle BRAiVE
- Carnegie Mellon University – experimental vehicle BOSS
- Stanford University – experimental vehicle Junior 3
- Daimler AG – Automated Drive on the Bertha-Benz-Route ("Bertha-Benz-Drive")
- Technische Universität Braunschweig – Project Stadtpilot with the experimental vehicle Leonie

1.5 Focus on Aspects of Autonomous Driving

Based on our functional requirements for an "autonomous vehicle" in Sect. 1.3, we identified the following major aspects (see also Fig. 1) which have to be handled by an autonomous vehicle and which depict a relevant differentiation factor among the research projects:

1. Onboard environmental and self-perception: Among other requirements, the autonomous vehicle has to perceive and interpret its local environment as

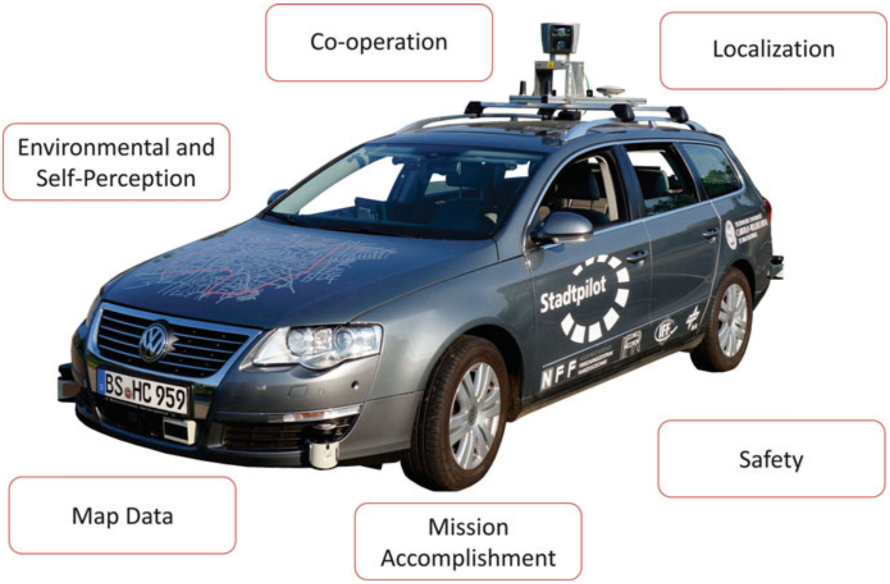


Fig. 1 Aspects of an autonomous vehicle using the example of the test vehicle Leonie of the research project Stadtpilot at Technische Universität Braunschweig

completely as required. This includes the detection of road markings and traffic signs, the detection of their position relative to the vehicle and their meaning, the detection of additional lane borders like curbs or grass, etc, the detection of raised targets, pot dots and holes, traffic lights, other traffic participants (pedestrians, bicycles, motor bikes, commercial vehicles and buses, trams, ambulances, animals, etc.), weather conditions, as well as the recognition of the own vehicle state (fuel level, tire pressure, wheel ticks, etc.).

2. Mission accomplishment: The mission must be accomplished by the vehicle, starting with the route planning down to the control of the vehicle's actuators. In this chapter, we also include the situation assessment of this aspect.
3. Localization: The vehicle needs to know its absolute or map-relative global pose.
4. Usage of map data: Without map data, route planning becomes challenging. But the level of map data usage also indirectly provides information about the capabilities of the localization and perception approaches. In many cases, all non-perceivable environmental features are mapped manually and then provided as map data to the vehicle's guiding system. The big disadvantage is that these data are very soon outdated. Therefore, how much the different approaches trust in and rely on map data is analyzed.
5. Cooperation: The vehicle has to be integrated into the traffic flow and thus needs to cooperate with other traffic participants, no matter whether they are humans, driven by humans, or driven automatically. Otherwise, the automated vehicle will be a foreign object in mixed traffic.

6. Functional safety: Without assuming the human driver as a fall-back solution, it must be ensured that the vehicle does not constitute any danger to its environment.

The basis of our survey on different projects is the functional system architecture, which has been developed within the research project Stadtpilot at the Technische Universität Braunschweig (Matthaei and Maurer 2015; Matthaei 2015) and which is shown in Fig. 2. The functional system architecture is designed as a modular building block system covering multiple ways of designing an autonomous vehicle. It includes all aforementioned aspects of an autonomous vehicle and especially combines localization-driven and perception-driven approaches in one single system description.

Each of the following sections focuses on a certain column, except for the sections concerning safety and cooperation, which describe interdisciplinary aspects. This survey does not consider the many publications in the field of driver assistance systems, which are not published within the context of integration into an overall system. On the one hand, their consideration would exceed the scope of this chapter; on the other hand, it would falsify the real state of research in the field of autonomous driving. A big challenge of autonomous driving is to master the system complexity. While driver-assistant systems usually only cover smaller subtasks, an autonomous vehicle needs to cover the whole functionality which is usually managed by the human driver. This consequently raises the complexity of the entire system. For example, it might happen that unforeseen incompatibilities occur or the limited resources are exceeded while integrating various single solutions into an overall system. This can even exclude certain single solutions from the integration into an overall system.

As already stated in Sect. 1.4, the following state of research is only based on self-disclosed information of the research teams in the form of scientific publications or a response to our questionnaire (see Sect. 4). Due to missing metrics, it is demanding to benchmark the quantitative capabilities of the different systems, e.g., the perception systems. Thus, the comparison focuses mainly on qualitative differentiators.

2 State of Research

2.1 Perception

The most challenging aspect of autonomous driving seems to be the computer-based environmental perception (see Bar Hillel et al. 2012). This is approved indirectly by the often applied strategy of an extensive map usage as presented in, e.g., Wille (2012) or Ziegler et al. (2014), to be able to provide at least any automated driving function without the interference of a human driver.

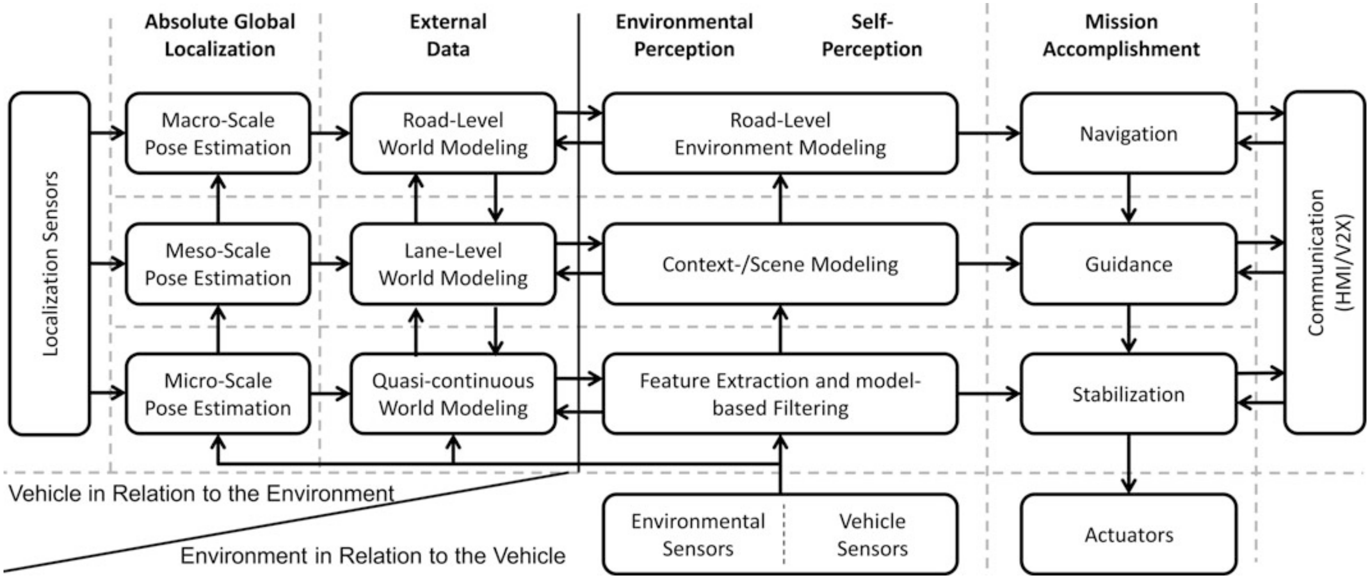


Fig. 2 Functional system architecture for an autonomous vehicle as a building block system according to Matthaei and Maurer (2015) and Matthaei (2015)

Activities of the Universität der Bundeswehr München, and especially the concepts in Broggi et al. (2013), are mainly driven by online environmental perception. During the DARPA Urban Challenge, some teams were using concepts dominated by online environmental perception as well (e.g., Leonard et al. 2008). However, most of the competitive teams relied on a priori map data, at least for the stationary environment (see Bar Hillel et al. 2012).

2.1.1 Perception of the Stationary Environment

In this chapter, the term "stationary environment" describes the pathway of lane markings in the direct detection range of the vehicle (e.g., at inner-city intersections), drivable areas, as well as static obstacles. Positions and implications of traffic signs, types of lane markings, and positions and phase of traffic lights, road curbs, entries and exits of tunnels, bridges, etc., are also part of this definition. This incomplete list already shows how heterogeneous the expected environment might be.

Researchers at the Universität der Bundeswehr München developed concepts for the detection of road courses without the usage of a priori map data (e.g., Manz 2013; Müller et al. 2011). Manz shows in Manz (2013) concepts which are capable of detecting and modeling intersections and splitting roads from this data. These activities were mainly focused on unpaved areas. The height profile of the environment is also modeled (see Manz et al. 2011). The algorithms were designed to perform even under adverse conditions, e.g., partial occlusion or a limited range of view, caused by rain or snow (Manz 2013, pp. 173ff). Based on the feedback on our questionnaire, traffic lights and signs cannot be detected at the current stage of development.

The project BRAiVE has also developed a wide range of environmental perception systems. As depicted in Broggi et al. (2013), the vehicle is able to detect and classify the complete set of the Italian traffic signs within 100 ms. Lane markings can be perceived by a mono and a stereo camera. The system can distinguish between white and yellow as well as dashed and solid lane markings. A height profile can be generated, similar to the approaches of the Universität der Bundeswehr München. One unique characteristic is the detection of the domain (highway, rural road, and city) based on the detected features. These concepts seem to be able to detect tunnel entries. Additionally, a parking lot detection was developed for certain scenarios (Broggi et al. 2013).

The research team of BMW Group Research and Technology published an algorithm in Homm et al. (2011), which is able to correctly detect 100 % of the lane markings on a designated test site in Munich. This algorithm is based on a fusion of laser and camera data. Merely the number of available lanes was extracted from regular navigation system map data.

The authors in Levinson (2011) and Levinson et al. (2011) have shown concepts for the detection of traffic lights and signs based on the data of a laser scanner.

The stationary environmental perception is also addressed in the Stadtpilot project (Matthaei et al. 2014b). However, stationary data, especially the position of lane boundaries, is currently still provided by a priori map data (Matthaei et al. 2015). An online adaption of the driving tube to additional stationary obstacles (e.g., a parked delivery van) is planned (see Wille 2012), but not yet shown in public traffic.

The perception of lane markings during the conditionally automated drive of a Mercedes-Benz S500 Intelligent on the Bertha-Benz-Route was solved with the support of map data (see Ziegler et al. 2014). Stereo cameras detected the stationary environment. Based on the available documents, it cannot be determined whether a detection of traffic signs was used.

2.1.2 Perception of the Movable Environment

Compared to the perception of the stationary environment, the perception algorithms and techniques for dynamic elements of the vehicles' environment are more sophisticated. Besides the perception of traffic participants, i.e., trams, buses, trucks, cars, bicycles, wheelchairs, and pedestrians, strictly speaking, also animals have to be perceived on time for inner-city and country roads.

The research vehicle BRAiVE is able to detect and classify preceding vehicles via computer vision. The algorithms are designed to detect symmetric patterns of a vehicle back view as well as backlights. Laser sensory was used to determine the exact distance to the target vehicle. The algorithms are also capable of tracking other vehicles inside winding roads. During the night, the headlights of forthcoming vehicles are perceived by computer vision. The vehicle can detect pedestrians inside the area of risk by using a fusion of computer vision and laser scanner data (Broggi et al. 2013).

The work of the Universität der Bundeswehr München published in Manz (2013) and Fries et al. (2013) provides algorithms for the detection and tracking of significant contour features (e.g., tires, car windows, vehicle silhouettes, vehicle lights) even under severe weather conditions. Based on the results presented in this work, the concepts are also capable of detecting crossing traffic and classify vehicles and pedestrians (Himmelsbach and Wünsche 2012).

The authors in Aeberhard et al. (2011) (BMW Group Research and Technology) have shown detection and tracking algorithms for vehicles on highways. The research vehicle was equipped with laser and RADAR systems covering 360° of the vehicle environment. It provides a large detection range in front region of the vehicle (Ardelt et al. 2012).

Team AnnieWAY has focused on highway scenarios during the Grand Cooperative Driving Challenge and used RADAR technology for vehicle tracking (Geiger et al. 2012). In the Stadtpilot project mainly laser scanners are used for the detection and tracking of other traffic participants (Ulbrich and Maurer 2013; Matthaei et al. 2015).

2.1.3 Self-Representation

In this chapter, the term "perception" includes the detection of the current state of the vehicle and the representation of its own performance capability. Many projects already use some form of ego motion estimation, used inside the time-based fusion of the vehicle environment and to enhance the global localization. The self-representation addressed in this paragraph goes far beyond these techniques and also includes the performance of all sensors, actuators, hardware and software components, and the vehicle itself, as briefly discussed in Maurer (2000), Siedersberger (2003), and Pellkofer (2003). In addition to the basic operability, the quality of self-representing data and its correctness are considered. This information gathered is necessary to evaluate possible vehicle (re)actions regarding their safe executions and degrade these actions if required. The project Stadtpilot at the Technische Universität Braunschweig uses sensor data to detect weather and road conditions and uses this data to influence the vehicle guidance systems (Reschka et al. 2012a). At this point, based on the publications available, it cannot be determined whether this kind of representation is considered in other projects.

2.1.4 Context Modeling

Based on the aforementioned modules of environmental perception, it is mandatory to link the respective results, in order to create a model of the local context around the automated vehicle. The author in Brown (1996) defines a context as a “combination of elements of the user’s environment which the computer knows about.” The addressed user has to be replaced by the automated vehicle in the scope of automated driving. Many approaches and concepts deal with the context modeling paradigms. In Chatila and Laumond (1985), Becker and Dürr (2005), and Strang and Linnhoff-Popien (2004), some of these concepts are extensively discussed. Several projects use an association of detected traffic participants to road lanes (e.g., Wille 2012; Geiger et al. 2012; Ziegler et al. 2014), although this step is not regarded as a part of the environmental perception.

In the project Stadtpilot a central context model is provided, modeling, among other elements, even the phases of traffic lights (Wille 2012). Based on the results from our survey, the research projects of the Universität der Bundeswehr München store detected roads, intersections, and dynamic objects inside a scene graph structure. Static obstacles are represented inside an occupancy grid.

2.1.5 Conclusion

As depicted in Bar Hillel et al. (2012) and Ziegler et al. (2014), the machine-based environmental perception has still a long way to go towards a complete perception of the vehicles’ environment. In Sect. 2.6, some dilemmas are presented, which could possibly lead to legal or ethical conflicts of decision. Nevertheless, nowadays systems are not capable of detecting these cases. The features required for this comprehensive environmental perception cannot be measured with current sensor systems (e.g., a laser system solely measures distance and reflectance of a target, but does not provide information about elasticity or mass). Algorithms required for the information extraction from sensor data are not designed or not able to perform in real time (e.g., sophisticated computer vision techniques). Vehicles do not know about the number of passengers in other vehicles and in most cases cannot distinguish between a child, an animal, and a trash bin when detecting an obstacle on the road. The availability of presented solutions does not yet meet the requirements of automatic driving vehicles. To provide at least a driving function, many research groups rely on map data, which is created manually. By doing this, required interpretation skills of the online perception can be avoided (see Sect. 2.2).

The concepts of the project BRAiVE (see Broggi et al. 2013) and the Universität der Bundeswehr München rely mainly on online perceived data and focus on their interpretation, which distinguishes them from other research groups.

2.2 Usage of Map Data

2.2.1 Definition of Terms

In this chapter, the term "map" is understood as an image acquired outside the vehicle of the stationary environmental features. Due to this limitation it becomes clear that verifying the quality and up-to-dateness of map data is difficult, because the map has usually been drawn up at an earlier time. Thus, due to the lack of continuance in the surveillance of the environment, it cannot be ensured that map data is up-to-date and all potential changes of the stationary environment are considered. In a certain sense, it therefore makes no difference if map data had been acquired 1 h, 1 week, or 1 year ago. According to the authors' understanding, this is the reason why autonomous systems must not rely on map data for stabilization purposes due to safety reasons, especially for collision avoidance and lateral control. The alleged stationary environment changes too quickly for the update rate of map data in the foreseeable future. In the case of navigation, map errors are annoying, because they might lead to a detour or do not lead to the desired destination, but they do not present a direct safety threat.

Today, map data is widely used in vehicles for navigation purposes and also for supporting the environmental perception, e.g., traffic sign recognition. Hence, it is comprehensible that all approaches on autonomous driving considered in this chapter also require map data. The usage of map data often exceeds the pure navigation tasks and can be assigned to the following three objectives:

1. Extension of the field of view (e.g., for navigation purposes)
2. Supporting the environmental perception and compensation of the sensors' limitations (e.g., by using position and type of road markings from map data)
3. Supporting the localization and compensation of the limits of GNSS (global navigation satellite system)-based localization (e.g., map-aided localization)

Map data in general differs, among others, in the type of stored features (see also classes of landmarks in Hock (1994) and Gregor (2002) or different abstraction levels in Matthaei (2015) and Matthaei and Maurer (2015)) as well as in their geometrical, semantical, and topological correctness and completeness. The geometry describes the position of the features, the semantic describes their meaning and class (e.g., posts, tree, building, traffic sign, etc.), and the topology describes the connections between the features, e.g., the road network.

2.2.2 Map Data in the Context of Autonomous Driving

During the DARPA challenges, the “route network definition file” (RNDF; see DARPA 2007) was introduced. The RNDF provides lane-level maps for areas with roads and a description of so-called zones which define the boundaries of unstructured areas as well as the positions of parking lots. The road description contains all types of information (geometrical, semantical, and topological): the course and the width of the lanes (geometrical), the connection of roads and lanes (topological), and the types of lane markings (semantical). However, the geometrical information in particular is just a rough representation of the environment containing only a sparse set of imprecise support points. That is the reason why some teams manually edited these maps (e.g., Bacha et al. 2008; Miller et al. 2008).

During the Cooperative Driving Challenge 2011, team AnnieWAY used highly accurate maps with lane courses. These maps were created by driving manually in the center of the right-hand lane and recording the GPS positions. Neighboring lanes then were added by a model-based estimation assuming the lanes to be parallel. These lanes were used for assigning traffic participants to certain lanes and thus for supporting the environmental perception (Geiger et al. 2012).

In the research project Stadtpilot, even more detailed and more accurate map data are applied (see p. 97 in Wille 2012). They contain topological information similar to the RNDF in the form of connected lanes (Nothdurft et al. 2011b). Geometrical information is manually extracted from highly accurate aerial images. The stored features are the courses of the lane borders. In this project, map data is used for supporting the environmental perception (see Wille 2012, p. 105) as well as to enhance the localization (see Wille 2012, p. 97).

The Bertha-Benz-Drive was also based on highly accurate maps which were “of prime importance” (Ziegler et al. 2014). According to Ziegler et al. (2014), three different map types were created: maps with 3D point landmarks, maps with an exact representation of road markings (lane boundaries, stop lines, and curbs) and rails, and maps with more abstract information on lane level. They had been used for all three aforementioned objectives (localization and support of the environmental perception within and outside the sensors’ field of view). The mapping process of the 3D point landmarks was done offline but fully automated with a stereo camera system. The projection of the 3D point landmarks to the ground plane provides the data basis of the other two map types. The lane-level map then is manually created by editing the road marking map and is then stored in the file format of OpenStreetMaps. It contains in addition to the course of the lane boundaries the lane topology, rights of way information, and traffic lights (Ziegler et al. 2014).

Map data with centimeter accuracy of the lane information is also used by the BMW Group Research and Technology while driving highly automated on the highway (Ardelt and Waldmann 2011). The exact objective of using map data is not mentioned in detail, but maps seem to be applied for both supporting the localization and supporting the environmental perception (see Fig. 3 in Ardelt and Waldmann 2011).

The VisLab Intercontinental Autonomous Challenge (VIAC) from Italy was performed without any map data, because map data was not available for the planned route in some regions. But the team intends to integrate map data for the following research activities. In Broggi et al. (2013), a navigation level (according to Broggi: long-term planning) is introduced. For this purpose, OpenStreetMaps are enriched with additional information such as the number of lanes, their width, and traffic lights (see Broggi et al. 2013, p. 1412). A support for the localization or environmental perception is not mentioned in the publications.

In the context of the research activities at the Universität der Bundeswehr in München, only imprecise road maps with an accuracy of about 10 m are used according to their own statements. They provide data for route planning and for initializing intersection hypotheses of the environmental perception. These maps are obtained from OpenStreetMaps, for example, without any further enrichment with details of the vehicles' environment (see also Müller et al. 2011).

2.2.3 Conclusion

Map data is currently very important for autonomous driving for many purposes: For commonly known route planning (e.g., Broggi et al. 2013 or Müller et al. 2011), as a replacement of the perception of lane markings (e.g., Wille 2012), or for supporting the GNSS-based localization (e.g., Ziegler et al. 2014 or Levinson 2011), all versions of integrating map data in the system can be found. None of these approaches discuss solutions for handling short-term changes of the environment. The challenge of ensuring the up-to-dateness of map data also remains unanswered. However, as long as the stabilization of the vehicle is based on map data, they are a safety-relevant data input.

Concerning the requirements of map data, those projects are leading, according to the authors' point of view, which do not completely rely on map data and which are also able to deal with imprecise map data. Our own experiences in the research project Stadtpilot demonstrate again and again how vulnerable map-based approaches to the smallest changes in the environment. It is very simple to change the position of stop lines, to change road markings from dashed to solid, to change the prescribed driving direction of certain lanes at intersections, and to add a new speed limit sign or to start roadwork. All these are small changes, which do not affect the infrastructure on a macroscopic level but would lead to a malfunction of an automated vehicle mainly relying on map data. This might also result in a behavior which is not allowed regarding road traffic regulations or even lead to endangering behavior.

Approaches to speed up the update of map data are already being pursued. Some years ago the project ActMap (Flament et al. 2005) researched online map updates from a central server to the vehicles. In the following project FeedMap (Visintainer and Darin 2008), additional approaches were examined to send data from different vehicles to a central server and thus increase the up-to-dateness of map data. This idea is also mentioned in Ziegler et al. (2014) and conceptually integrated into the system architecture in Matthaei and Maurer (2015)and Matthaei (2015).

In a further step it might be possible that autonomous vehicles themselves send perceived environmental features to a central server and thus update their maps themselves. This would be an example for collaboration. But this assumes that the vehicles are basically able to perceive and understand their environment completely and that they know exactly their position. Each vehicle might be the first one which reaches an unknown or a changed place and it has to react in a proper way. Based on this discussion, it is debatable whether approaches with a high trust in map data really show the way forward to "autonomous driving."

2.3 Cooperation

2.3.1 Definition of Terms

The term "cooperation" describes a certain form of social collaboration between at least two participating partners, aiming at an enhancement for ourselves as well as the other party in comparison to an egoistic approach (Spieß 2014).

In more technical terms, referring to this definition and based on (Stiller et al. 2013), cooperation is the approach to get a "better" solution for a given problem in terms of a to-be-defined optimization criterion. This criterion can have various characteristics, and its agreement is a key aspect of cooperation.

The criteria might be defined a priori by the systems' programmer (e.g., avoid collisions with other vehicles), or, in a more sophisticated way, might be dynamically defined between the traffic participants. The coordination of these criteria and their fulfillment require a high amount of negotiation between the participating parties and are the main topic when talking about cooperative aspects in public traffic, especially regarding contradicting objectives among some parties.

Therefore, one main prerequisite for cooperation is the ability to communicate with the other parties. This communication can take several forms. Technical approaches use V2X technologies for this purpose (▶ Chap. 27, "Vehicle-2-X"). The idea of communication between traffic participants did not arise with this technology, but can be found already in current traffic regulations. These dictate the presence of, e.g., turn indicators, brake lights, and a signal horn. Other very important ways of communication are gestures and the behavior of traffic participants. Human drivers will communicate their intentions by hand signs, and humans are quite good in deriving the intentions of other drivers from their behavior. One well-known situation requiring communication as well as cooperation is the merging process at on-ramps of highways, as depicted in Fig. 3.

The term cooperation is widely used in science and the context of (advanced) driver assistance systems, but has not been finally defined for autonomous driving. When using the term cooperation in a more general way, two levels of cooperation can be found. The first one includes the compliance with current traffic regulations required for driving in public domains. Collision avoidance and basic strategies of traffic flow control are the main contents, e.g., the right of way and merging procedures at highways. The communication of the driver's intentions is made by

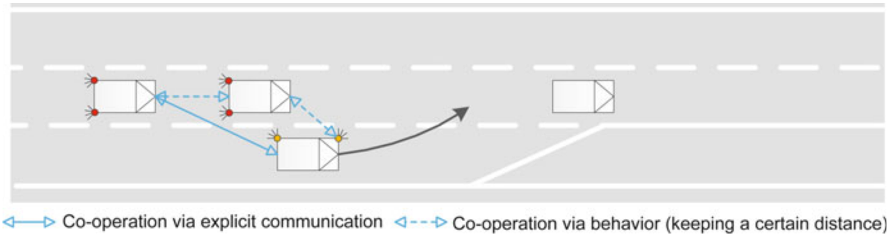


Fig. 3 Today's optical communication technologies (turn indicators) only allow an indirect communication, for example, at on-ramps. V2V communication would enable a direct communication of the driver's intention to the entire environment

the indicator lights or the driving behavior and is used as cooperation request to other participants.

The second level is defined by concepts for a more sophisticated optimization of traffic flow and guidance, e.g., leaving a larger gap to the preceding vehicle, so that other vehicles can merge into our lane more easily. Both levels of cooperation do not require the usage of V2X or other explicit communication technologies, but can basically also be realized using onboard sensors. The cooperation aspect is located in the guidance module of the functional system architecture in Fig. 2.

2.3.2 Cooperation in the Current Context of Autonomous Driving

Up to now, cooperative aspects of the second level are rarely established. First approaches were presented in the research project "Cooperative and Optimised Traffic Signalling in Urban Networks" (KOLINE) on the inner city ring of Braunschweig (Brunswick), Germany. The main goal was the optimization of traffic flow regarding fuel usage and noise pollution by implementing a cooperative optimization of the traffic flow (Saust et al. 2010). Sensors mounted on infrastructure elements and telemetry from test vehicles were used to determine the traffic flow speed and calculate an optimal approaching strategy to the next intersection. Additionally, the adaption of traffic light phases was investigated. As a result, the amount of required stop-and-go maneuvers could be reduced by approximately 20 %, and the fuel usage was reduced by around 5 % (Bley et al. 2011).

Other results were published as part of the "Grand Cooperative Driving Challenge" (GCDC) in 2011. As in the previously presented project, the main focus was the optimization of traffic density by using an automatic grouping of traffic participants (platooning) (Nunen 2012). Communication was based on a V2X platform. Nine teams participated in this challenge and had to prove their concepts. The team "AnnieWAY" won this challenge, showing that this scenario can technically be handled (Geiger et al. 2012).

2.3.3 Conclusion and Outlook

Although the usage of explicit communication patterns, e.g., V2X technology, gives the potential of many sophisticated uses, some challenges arise. Public traffic domains still allow for participants without this explicit communication technology, so these participants have to be regarded as well as fully equipped vehicles.

V2X technology, as well as every other communication technology nowadays, has a certain potential of – intended or unintended – misuse. As a result, data security and safety have to be ensured when bringing these technologies to the market. The presence of non-V2X participants leads to the conclusion that these technologies can only be an additional data source to onboard sensory. This was already one main experience in the GCDC, where a RADAR system was used as a plausibility check against the V2X messages from other participants (see Geiger et al. 2012, p. 8).

The fusion of sensor data, perceived by other traffic participants as well as infrastructure systems, seems to emerge as one of the next research topics in cooperative technologies. These topics are described by the term "cooperative perception." First ideas were published in the research initiative "Co-operative sensory and co-operative perception for preemptive safety in public traffic" (KoFAS) in 2013 (Goldhammer et al. 2012; Rauch et al. 2013). So far, these topics have not been addressed in the context of autonomous vehicles in terms of this chapter. Another extent of cooperative technologies could be the joint generation of map data, as discussed in Sect. 2.2.

2.4 Localization

Another key aspect of autonomous driving is the localization: Without a proper localization of the host vehicle, the usage of map data is not possible, and without knowing the relative motion of the vehicle between two points in time, the perception of the environment is at least more complicated. Additionally, the communication among the vehicles in the context of cooperation requires in most cases an exchange of the vehicles' positions for an assignment of the messages to locally perceived objects.

2.4.1 Lessons Learned from the DARPA Challenges

Based on the experiences made during the DARPA Challenges, one major finding referring to the localization is that the relative motion of the host vehicle should be strictly separated from the absolute localization (Moore et al. 2009). These two localization solutions differ in their optimization goal. The relative motion estimation (in Moore et al. 2009 called "local frame") describes a continuous sequence of positions starting at an arbitrary position with the objective to provide the incremental change in position as exact as possible. The long-term positioning may drift away.

On the contrary, the absolute localization (in Moore et al. 2009 called "global frame") determines the best estimate of a position at the current time step within a reference frame fixed in place. It thus does not drift, but the sequence of positions can contain discontinuities.

The following comparison of the research approaches concentrates on this global localization and mainly focuses on map-relative localization because in most cases the absolute global localization is just a first step towards more relevant map-relative localization.

2.4.2 Localization in the Context of Autonomous Driving

One of the latest publications on the subject of autonomous driving was made in the context of the Bertha-Benz-Drive. Their approach of a map-relative localization is based on a mono camera looking backwards (see, e.g., Knapp et al. 2009; Lategahn and Stiller 2012). The exact map-relative pose of the vehicle is determined by matching online perceived road markings with those road markings stored in the map. The detection of road markings in the image is pre-initialized based on map data. Hence, the detection of the road markings is processed with detailed a priori information (see also Sect. 2.1 and Ziegler et al. 2014). In addition to this approach, single features obtained from a mono camera are matched to a map with 3D point landmarks (see, e.g., Knapp et al. 2009). According to the researcher’s own information, the implemented approach is able to work without GPS (Ziegler et al. 2014).

In Levinson (2011), two fundamentally different approaches are published. On the one hand, a localization solution based on a highly accurate INS-DGPS platform was used during the DARPA Urban Challenge for a map-relative localization. Map errors and errors of the global localization were corrected with the aid of matching between map data and curbs obtained from laser scanners, as well as reflectance values of road markings. On the other hand, a previously recorded dense grid-based map of the ground plane containing reflectance values of a laser scanner is generated in an offline process, which is much more detailed than the lane-level representation of the RNDF. In a second run, the vehicle can now match the currently perceived environmental data to this detailed map and thus calculate the correct pose (position and orientation) similar to the aforementioned approach.

On the contrary, Leonie of the research project Stadtpilot drives based on an INS-DGPS system. However, matching approaches based on lane markings were also developed and trialled on a test circuit (Nothdurft 2011a; Wille 2012; Matthaei et al. 2014a), but are currently not used for an automated control of the vehicle.

The approach developed at the Universität der Bundeswehr in München follows a fundamentally different philosophy. Based on the lesson learned – “never trust GPS” (see Luettel et al. 2009) – they developed in the tradition of (Dickmanns et al. 1994) a system, which relies almost completely on the perception and which uses GPS and map data only as rough hints and for route planning. According to the team’s statements, accuracies of 10–20 m for the GPS-based localization are sufficient. The global pose estimation is also supported by a matching between environmental and map data, but on a higher abstraction level (Müller et al. 2011).

2.4.3 Conclusion

The determination of the global position of the host vehicle is necessary in most projects for integrating external data. This data covers in most cases map data but may also be V2X data. Obviously, even today's highly accurate localization systems are not sufficient for a reliable stabilization of the vehicle (Levinson 2011; Nilsson et al. 2012). For that reason, highly accurate and detailed map data is often used to compensate this lack of accuracy by matching environmental features to this map data. Some approaches don not even need an absolute global pose, but only require a map-relative global pose (e.g., Müller et al. 2011).

Some other projects follow the objective of becoming more independent of highly accurate absolute global positioning by increasing their trust in map data or in their environmental perception.

2.5 Mission Accomplishment

Behavior planning and control are integral parts of the driving task an autonomous vehicle must by definition be able to accomplish.

In Donges (1999) and ▶ Chap. 2, "Driver Behavior Models," the driving task is divided into three levels: navigation level, guidance level, and stabilization level. A similar hierarchy can be found in the architectures of many projects in the field of autonomous driving (Broggi et al. 2013; Urmson et al. 2008; Montemerlo et al. 2008; Kammel et al. 2008; Matthaei et al. 2015; Matthaei and Maurer 2015). As it provides a clear and hierarchical structure for mission accomplishment in automated driving, the following discussion will also be using this three-level model. The terms "navigation," "guidance," and "stabilization" are used accordingly.

Decisions made at the navigation level affect the whole mission. Thus, it is also called the strategic level. At the guidance level, tactical driving decisions are made, such as the selection of a particular driving maneuver. For these, the current driving situation is assessed and command variables for the underlying stabilization layer are derived. Because of the local planning horizon, it is also called the tactical level. Modules in the stabilization layer take care of the control of command inputs from the two superordinate levels. Figure 4 illustrates the different hierarchical levels and an additional human-machine interface for a passenger or system operator. This interface provides the possibility of entering or modifying mission goals (see Sect. 1.3).

All teams cited in this section address planning and control only for situations that are known and considered at the design time of the system. Handling unknown or unconsidered situations may be necessary for autonomous driving in its final stage, but is for most teams beyond the scope of their challenges to be currently addressed.

2.5.1 Navigation

A passenger mainly communicates with the automated vehicle on the navigation level. It is possible to enter the desired destination or certain waypoints. A common approach is to represent the map data as a directed graph. This way it is possible to solve the navigation task by using graph search algorithms.

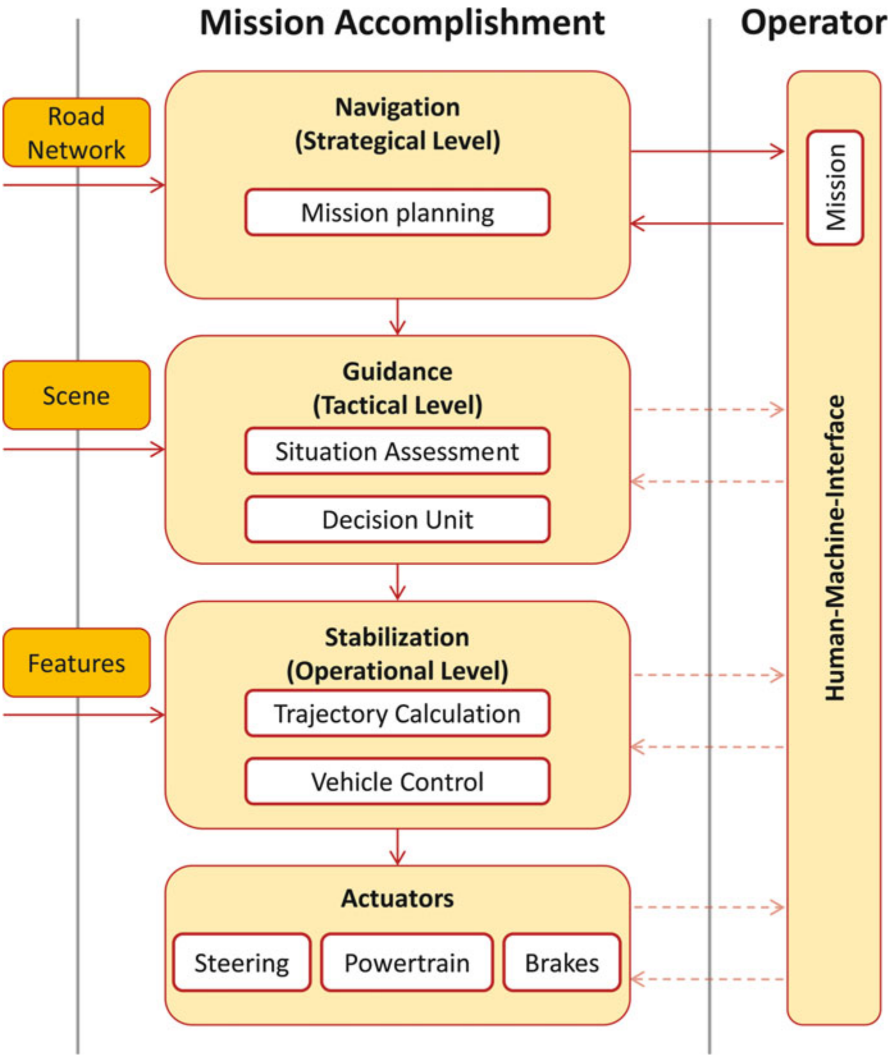


Fig. 4 Subdivision of the driving task similar to Donges (1999), Matthaei and Maurer (2015), and Matthaei (2015)

In the Urban Challenge a lane-level map of the road network was given (RNDF; cf. Sect. 2.2). Team AnnieWAY applied an A* algorithm to compute an optimal lane-level route with respect to the expected travel time (Kammel et al. 2008). In a preprocessing step, the waypoints of the given map were interpolated, using a spline-based geometric representation. Thus, the resulting route represents already a continuous path to the target position. A reactive layer, however, is still allowed to vary the path locally (cf. Sect. 2.5.3). If a road is recognized as blocked, the internal representation of the road network is updated and a new route is computed.

For Boss (Urmson et al. 2008) and Junior (Montemerlo et al. 2008), a different approach was used: Instead of planning an explicit route to the destination, for each edge of the graph that represents the road network, the remaining costs to reach the destination are computed. The decision which route is actually taken is shifted to the guidance level. Again, the costs depend on the expected travel time. On the guidance level then, these strategic costs are combined with the costs that arise from the current traffic situation, e.g., the costs of carrying out a lane-change maneuver. Analogously to AnnieWAY, if a road is recognized as blocked, the internal representation of the road network is updated and the costs are recomputed.

2.5.2 Guidance

On the guidance level, the autonomous car has to interpret the traffic situation with respect to its goals and the goals of other traffic participants. It has to generate alternative options of action, evaluate these options, and finally make a decision. Cooperative behavior is also part of this level, but it is already treated in Sect. 2.3.

A very common approach on this level, in the Urban Challenge and in subsequent projects, is to use state machines (Urmson et al. 2008; Montemerlo et al. 2008; Rauskolb et al. 2008). Such a state machine may have system states like the execution of lane change and overtaking maneuvers, the approaching of stop lines in and in front of intersections, decisions to cross a traffic light by its signal change from green to amber, or the execution of cooperative maneuvers, e.g., intentionally leaving a gap so that a vehicle on an on-ramp can merge in front of the automated vehicle. Team Carolo from Technische Universität Braunschweig (Rauskolb et al. 2008) used a hybrid approach of traditional, rule-based state machines to handle complex maneuvers like parking, U-turns, and crossing intersections and a behavioral DAMN arbitration model (Rosenblatt 1997) for regular driving along roads and for collision avoidance with obstacles.

Similar to the aforementioned approach, at the Technische Universität München (Goebl et al. 2008), a state machine combined with a fuzzy logic for situation assessment is employed. At the Universität der Bundeswehr in München, a hierarchical state machine with meta states like convoy driving, tentacle navigation (cp. Sect. 2.5.3), and U-turning is used on the tactical level (Luettel et al. 2011).

The approach taken by BMW in the ConnectedDrive project, which focuses on highly automated driving on controlled-access highways, differs from the approaches mentioned above in separating longitudinal and lateral control on the guidance level (Ardelt et al. 2012; Ardelt and Waldmann 2011). A hybrid, deterministic state machine is used to define the superordinate driving behavior, and a decision tree is used as a hierarchical decision-making process. The superordinate state is determined by traversing the decision tree, depending on the driving goal derived by the situation interpretation and the current feasibility of maneuvers.

Rule-compliant behavior at intersections is particularly demanding for autonomous road vehicles. In the Urban Challenge, the Stanford Racing Team used so-called critical zones that were encoded in the map data, to check if Junior has to give way to other vehicles at intersections without traffic lights (Montemerlo et al. 2008). The central element of the approach of the team from Carnegie Mellon

University is a module that determines the right of way by observing the order of arrival at an all-way stop. Moreover, it identifies gaps that are sufficient for crossing through or merging into moving traffic at an intersection (Urmson et al. 2008).

Behavior planning for crossing intersections with traffic lights is addressed in Saust et al. (2010, 2012). Via V2X communication, the remaining time of the current traffic light phase is transmitted to the automated vehicle and an energy-optimal approaching strategy is calculated, also considering possible traffic tailbacks.

From the authors' point of view, a central issue on the tactical level is the handling of perception and prediction uncertainties. At Carnegie Mellon University, an analytical model for predicting the driving behavior of other vehicles is used to evaluate tactical driving maneuvers (Wei et al. 2010). Here, the evaluation is limited to simulated data, and measurement uncertainties are not yet considered. However, an easy-to-comprehend driving behavior is achieved by the separation of a prediction and cost model. In Wei et al. (2011), the same team demonstrated the consideration of uncertainties in longitudinal planning for the task of single-lane automated driving under uncertainty using a Markov decision-making process.

At the Technische Universität Braunschweig, the consideration of perception and prediction uncertainties has been demonstrated in a first implementation of partially observable Markov decision processes for lane-change decision making in inner-city urban traffic (Ulbrich and Maurer 2013).

2.5.3 Stabilization

The stabilization level covers reactive path and trajectory planning and closed-loop control of the vehicle's actuators (steering, powertrain, brake). In the following, the focus is on trajectory generation in structured environments like roads. In the Urban Challenge, path and trajectory-planning methods for unstructured environments were also applied, e.g., when a vehicle had to navigate in a parking lot or when it had to handle road blockages (Urmson et al. 2008; Montemerlo et al. 2008; Kammel et al. 2008).

Team AnnieWAY, as most participants in the Urban Challenge, used a path-based approach for motion planning (Kammel et al. 2008). A reference path that follows a lane or crosses an intersection (and that may also include a lane change) is generated on the guidance level. Obstacles and other road users are not considered in the generation of the path. For this reason, the reactive layer tests, based on an occupancy grid that is generated by laser measurements, if the given path is collision-free. If it is not, a set of precomputed alternative paths that represent velocity-dependent motion primitives for collision avoidance is evaluated and the path with the highest utility is chosen. The alternative paths are also referred to as tentacles, as their use is similar to the tentacles of insects (Hundelshausen et al. 2009). The low-level control of the vehicle is split into longitudinal and lateral control. For the lateral control, i.e., to follow the selected path, a velocity-independent orbital tracking controller is used. The longitudinal control is responsible for following other vehicles at safe distances, stopping at certain positions, and keeping the speed limits.

The tentacle-based approach is also used in the test vehicle MuCAR-3 (Luettel et al. 2012). However, in this case no reference path is given, but a route consisting of distant global positions. The local navigation is done by using the deviation from the straight line between two waypoints as an input of the utility function of the tentacles. Another path-based reactive approach is presented in Broggi et al. (2013).

At higher traffic densities, as they are common in urban traffic, a trajectory-based motion planning is necessary (Werling et al. 2010). The trajectory-planning method presented in Werling et al. (2010) generates, similar to the tentacle approach discussed above, a set of trajectories with minimal jerk in lateral and in longitudinal direction. The candidate trajectories vary in their end times as well as in their end positions. The latter are described as a longitudinal position and a lateral offset with respect to a given reference path. Again, the given reference path typically follows a lane. In a second step, the trajectory with the highest utility is selected, e.g., based on the predicted motion of other traffic participants. If the internal representation of the environment is consistent over time (i.e., if the prediction of the environment matches with future measurements), this method realizes an optimal (open-loop) control and the generated trajectory is consistent over time. However, both processes, perception and prediction, inevitably involve uncertainty, and thus the overall process has to be considered a closed-loop control. The method was applied in AnnieWAY after the Urban Challenge and is also used in Junior 3 (Levinson et al. 2011). A method for generating reference paths with minimal curvature is used in the Stadtpilot project at the Technische Universität Braunschweig (Wille 2012).

With their participation at the Pikes Peak International Hill Climb race, a cooperation of Stanford University and the Electronics Research Lab of Volkswagen of America demonstrated that it is already possible for an automated vehicle to follow a precomputed path at the friction limits of the vehicle (Funke et al. 2012).

2.5.4 Conclusion

The most common approach for structuring the task of mission accomplishment is a hierarchical (three-level) architecture, similar to the three-level model used to structure the above discussion. Research focuses currently on tactical decision making, maneuver execution, and trajectory planning. On the tactical level, lane-change maneuvers and cooperative driving (see Sect. 2.3) are popular subjects of research. Further research is necessary, especially in the field of situation prediction and situation assessment and in general on coping with uncertain information and the unknown intentions of other road users. On the stabilization level, comfortable yet accurately targeted trajectory planning is the focus of research for many teams. Many teams use planning-ahead approaches and choose among a set of generated candidate paths or candidate trajectories.

2.6 Functional Safety

2.6.1 Requirements

From the authors' point of view, functional safety for unmanned road vehicles is one of the main challenges for the introduction into public traffic. It is not yet clear when an autonomous vehicle is safe enough, and thus a societal consensus has to be found. This includes the definition of an acceptable level of operational risk. This level defines whether an autonomous vehicle is in a safe or unsafe state. There are only few research activities which investigate this issue, e.g., the project "Villa Ladenburg" at the Daimler and Benz Foundation (Maurer 2013; Reschka 2015) and an chapter by Bryant W. Smith about legal implications (Smith 2014).

The operation of an autonomous vehicle has to be safe in normal operation as well as in unforeseen situations and in case of technical failure, misbehavior of others, and bad road or weather conditions. The vehicle control system has to maintain a safe state or transfer the vehicle into a safe state in any situation. This is to avoid hazards from the vehicle to other traffic participants and passengers inside the autonomous vehicle. A possible solution for a safe state is a full stop of the vehicle at a safe location (Isermann et al. 2002). A safe location is a place where the vehicle is no hazard to others, like an emergency lane on a highway, a wide curbside on rural roads, or a parking area. In urban traffic with low relative speeds between traffic participants, a stop on a driving lane is also imaginable, if no emergency routes are blocked.

The challenge of transferring a vehicle into a safe state without handing over control to a human driver is one of the main reasons for the necessity of a human driver monitoring advanced driver assistance systems. In an autonomous vehicle, no human driver is available, because the vehicle can be operated driverless and even unmanned. For highly and fully automated systems, this fallback solution is not available either. The technical system has to be designed to cope with this situation, and it can be developed for a safe and reliable operation by using functional redundancy, hardware redundancy, and software redundancy. For example, a lane change on a highway or a stop at roadside on a highway to maintain a safe state requires a functioning environmental perception, a decision process which includes a risk estimation of possible actions, and a reliable actuator control. Such a stop at the roadside is a requirement in the homologation process for autonomous vehicles in Nevada and can be preceded by one or more lane changes (NDMV 2012).

In dangerous situations an autonomous vehicle has to react in a way that avoids any harm to persons. If no humans are threatened, damage to property has to be also avoided. A solution to solve such situations could lead to a violation of applicable law, e.g., traffic rules. This violation seems acceptable as long as damage to persons and property can be avoided, e.g., crossing a solid line to avoid a collision. Further, it is imaginable that situations do not have a solution without damaging persons and/property and violating rules. In such dilemma and polylemma situations, a decision between several options is necessary. This decision has to be taken under judicial and ethical rules. In current research and development, the consideration of such rules is not known to the authors.

Thinking of dilemmas and polylemmas, further questions with a strong connection to the public acceptance of the technology arise: Has the safety of passengers a higher priority than safety of other traffic participants? How should an autonomous vehicle behave, if damage to persons is inevitable? May this ever happen or is an inevitable accident a result of an engineering/design issue concerning the speed limit chosen by the automated vehicle itself? Besides stopping the vehicle, more actions to reduce harm are imaginable, e.g., lower vehicle speeds, increased safety distances, and changes in maneuver planning. In advanced driver assistance systems, so-called action plans can be used to maintain a safe state or transfer the vehicle into a safe state. Hörwick and Siedersberger (2010a, b) propose to stop the vehicle in case of a technical failure in a traffic jam assist system, if the human driver does not react on a takeover request.

Additionally, actions to restore the system's performance are useful and desirable. In Ghosh et al. (2007) such self-healing methods are described. The awareness of dilemmas and polylemmas, decisions to avoid an unsafe state or to restore a safe state, and the reduction of accident consequences demand the knowledge of the vehicle's own performance capabilities. In combination with a scene and situation awareness and a prediction of both, possible actions can be identified. The best one has to be chosen and executed (Maurer 2013; Isermann 2006; Reschka et al. 2012b).

2.6.2 Functional Safety in Current Autonomous Driving Efforts

In this section, safety concepts of the experimental vehicles from projects covered in this chapter are investigated. In all projects a safety driver or at least a human monitoring the system (operator) is necessary in public traffic, because of the difficulties and challenges described above. The safety driver or the operator has to control the vehicle in dangerous situations. As a consequence, all autonomous driving efforts so far have to be categorized as partial automation according to Gasser et al. (2012) and SAE (2014). The safety drivers and operators are able to take over control immediately by using the control elements inside the vehicle or from outside of the vehicle.

In the project Stadtpilot at the Technische Universität Braunschweig, the safety driver overrules the technical system by his or her intervention. In case of technical failures, control is handed over to the safety driver. As a consequence, he or she has to monitor the system and the traffic continuously. On a closed test track functions are used, which reduce the system performance capabilities by monitoring critical system parameters. For example, the current accuracy of the localization in the world and in the map reduces the maximum possible speed of the vehicle. Additionally, road and weather conditions are monitored and cause the vehicle to drive more carefully (Wille 2012; Reschka et al. 2012a, b).

The experimental vehicle MuCAR-3, developed at the Universität der Bundeswehr in München, is able to monitor itself and to reduce its functional capabilities based on the monitoring data. This reduction could lead to emergency braking maneuvers. The safety driver can overrule the system as well. Data collected by monitoring heartbeats of hardware and software modules and checking measured and calculated values is used to trigger restarts of components.

Additionally, failing system parts can be reconfigured. Altogether, these self-healing functions lead to a safer and more reliable operation of the vehicle. If self-healing does not restore enough functional capabilities, an emergency braking maneuver is triggered (Goebl et al. 2008).

As already mentioned, the experimental vehicle BRAiVE demonstrated automated driving in public traffic in 2012 (Broggi et al. 2013). On some parts of the roads driven in this demonstration, no safety driver was present on the driver's seat. The system was monitored from a person sitting on the co-driver's seat. This person was able to intervene by using the gear level and an emergency stop button to stop the vehicle immediately. Although this demonstration was impressive, it seems that a fully driverless and, even more important, a fully unmonitored operation are not possible with BRAiVE. An external function to stop the vehicle was available as well. This e-Stop function uses a remote control to trigger an emergency stop (Bertozzi et al. 2013).

At the Carnegie Mellon University in Pittsburgh, the experimental vehicle BOSS was improved after the DARPA Urban Challenge in 2007. The implemented SAFER (safety for real-time systems) approach for software redundancy is able to compensate failing software components by switching to redundant components (Kim et al. 2013). This switch from one component to another is executed in real time. The SAFER approach has to be seen as an addition to other approaches for safety functions, because it does not cover any hardware errors. Using the approach on distributed hardware could further improve its advantages for safety reasons. Such a system is currently being developed in the Controlling Concurrent Change (CCC) project in Braunschweig (Project homepage: http://www.ccc-project.org).

After the DARPA Urban Challenge, the Stanford University and the Volkswagen Electronics Research Lab have developed the experimental vehicle Junior 3. As a safety measure, the vehicle has so-called silver switches which control the activation of the vehicle guidance system. If these switches are activated by a human safety driver, the control commands from the vehicle guidance system are forwarded to the actuator control units. If the safety driver overrules the system or the system deactivates itself for another reason, the switches stop the forwarding of control commands to the vehicle. In this fail-safe position, the safety driver controls the car. This concept is similar to the one applied in Leonie from the project Stadtpilot.

The monitoring of the system in Junior 3 is done by a health monitor. This system detects malfunctions of software components. In contrast to the SAFER approach, no redundancy is implemented, but self-healing functions like component restarts are triggered. With these safety measures, a partial automated operation in public traffic is possible. On closed test tracks, a driverless valet parking function is available. The vehicle can then be stopped with a remote control (Levinson et al. 2011; Stanek et al. 2010).

2.6.3 Conclusion

High safety requirements for driverless operation of automated vehicles demand a safety system which can maintain a safe state or transfer the vehicle into a safe state.

The projects discussed take different approaches to solving safety issues. Unfortunately, no safety concept exists, which does allow an unmanned operation in public traffic. All research vehicles described are therefore conditionally automated, because either a safety driver or an operator monitors the system and has to intervene in case of technical failures.

In the DARPA Urban Challenge 2007, the unmanned vehicles were operated on a closed test track. The reduced complexity of the traffic and the trained stunt drivers of not automated vehicles compensated the resulting risk. Additionally, a remote stop function was used to stop the vehicles in any dangerous situation.

More recent research projects challenge the difficulties in unmanned operation. The Villa Ladenburg project of the Daimler and Benz Foundation (Maurer 2013) in Germany covers social and ethical aspects of unmanned vehicles in public traffic. Google Inc. has presented a vehicle without control elements such as a brake pedal, a steering wheel, and an accelerator. In Germany an unmanned safeguard vehicle for controlled-access highways is currently being developed and will be used in public traffic in the next years.

These approaches will likely rely on the above described safety functions in research projects and maybe combine them to more powerful safety concepts, which allow an unmonitored operation of vehicles in public traffic.

3 Outlook and Challenges

Autonomous driving is indeed a fascinating topic, especially because it concerns each one of us, directly – whether as a driver or a pedestrian. However, the vision of the final stage of any driver assistance system being “to leave a vehicle to its own means” arouses ambivalent feelings in society – somewhere between curiosity combined with the urge to explore and skepticism, possibly even paired with an apprehensive prejudice against technology.

Until a short time ago, autonomous vehicles were primarily discussed in technical circles. The general public only became aware of the future visions of driverless vehicles in cinemas. Recently, however, the general interest and expectations have been aroused in the media by regular success stories and short-term promises to launch these driverless vehicles on the market. For the future, we still expect a lot of details to be solved to finally have vehicles that are fully automated and more resource efficient and accomplish missions more safely than humans do today.

At first glance, the status of today’s driver assistance, in serial production or during the research phase, gives rise to hope. A large number of new functions have been shown during the past few years and are also dealt with in this book. However, our research has shown that the way towards autonomous driving seems to be longer than is sometimes communicated at the moment. This is possibly due to the fact that human performance, particularly when supported by carefully developed driver assistance systems (Knapp et al. 2009), is frequently underestimated. Contrary to driver assistance systems that primarily have the aim of compensating gaps in human abilities, which have been identified on the basis of accident analyses (Chiellino et al. 2010; Buschardt et al. 2006; Winkle 2015a), or of automating routine driving situations under a person's supervision, autonomous systems must reach the abilities of an attentive human driver. It is not until then that autonomous systems will be able to go beyond the human abilities and lead to a further reduction in the number of accidents.

A step that is not to be underestimated is to secure a current assistance system in such a way that it functions in future as expected in an autonomous vehicle, without the driver's supervision (this means, among other things, error-free in every traffic situation). There is a high probability here of unpredictable constellations not being taken into consideration that might possibly cause the system to react inadequately or not to react at all.

Even the dependency of the autonomous vehicles on automated map updates has further consequences. The autonomous vehicle is no longer the uppermost instance in an environment, but part of a superordinate system. This, in turn, will have an effect on the concept of the vehicles. The authors have been unable to identify any research activity in this respect in the context of autonomous driving, until now.

Furthermore, the questions outlined at the beginning are not at all clear, either. For example, there has not been any strategy as yet for evaluating the perception and/or interpretation of a system on a semantic level. It is still frequently the case that the topic of redundancy is not acute at all, in an urban environment, for example, not even a nonredundant solution can be implemented, even when endeavoring to make use of every possible means. In contrast to the stabilization level, it is probably not possible, however, to resort here to redundancy concepts from other disciplines, such as aerospace or power plant technology.

Adaptations in the infrastructure are controversial, because they are extremely costly and even maintenance intensive where technical enhancements are involved. At present, the legal situation is being partially adapted for a trial operation, but, as yet, never without safety drivers. Thus, according to Gasser et al. (2012), all the public demonstrations are, per definition, partially automated, even though the targets set in the projects require highly automated, fully automated, or even autonomous driving.

This raises an essential point in today's public debate. As mentioned at the beginning, no consensus has yet been reached on the range of functions for autonomous driving. Furthermore, statements concerning the introduction of autonomous driving appear to be very optimistic in many cases. On the other hand, the introduction of partially automated systems is already under way.

The research project "Villa Ladenburg," of the Daimler and Benz Foundation, has initiated a multidisciplinary debate in society on an interdisciplinary approach towards a holistic development and risk acceptance. Numerous questions and aspects in the research and development process were identified there (Winkle 2015b). On gaining successful proof that full automation is superior for road safety in the long term, a completely new question could then be posed, namely, whether the error-prone human should be allowed to continue driving a vehicle independently (Maurer 2013).

Appendix: Questionnaire on "Autonomous Vehicles"

Project Organization and Objective

This section addresses general and organizational aspects of your project.

1. What is the name of the project?
2. With which universities and/or industrial partners do you cooperate in this project?
3. When was the project started/how long have you been working on the project now?
4. What is the objective of the project?
5. What are the assumptions, constraints, and restrictions in the project?
6. Was the system demonstrated on public roads? If so, which capabilities were publicly demonstrated in which domains?

Perception

This section addresses the perception and localization of your system as well as its environment representation.

1. Please describe briefly the perception architecture of the system.
2. What sensor technologies (and which devices) are used?
3. How are dynamic objects perceived and represented by the system? What kind of dynamic objects are perceived?
4. How are static objects and road boundaries perceived and represented by the system?
5. Is the system capable of perceiving the state of traffic lights? How is this done?
6. What kind of traffic signs are perceived by the system?
7. Does your system perceive pedestrians and cyclists? If so, under which conditions and by the use of which sensors and algorithms?
8. How are lanes perceived? Which conditions must be fulfilled by a lane?
9. What are the requirements for perceiving lateral traffic at intersections? Does the system identify the right of way?
10. Is the system capable of determining the topology of intersections from perceived data? How is this done?
11. Which intentions of other road users can be perceived by the system? Which conditions must be fulfilled therefore?
12. How are the overall and the current capabilities of the system represented and monitored? How do the current capabilities determine the behavior of the system?
13. How is the relative position of the vehicle with respect to the lane perceived? Which conditions (e.g., lane markings, geometric models) must be fulfilled therefore?

14. Is information from digital maps used? How accurate are these maps?
15. Is the localization in digital maps solely based on satellite-based positioning or are other sensors used additionally? If so, which additional sensors and algorithms are used for localization?
16. Is the system capable of communicating with other road users or the road infrastructure (i.e., Car2x technologies)?
17. Are perceived features (e.g., objects, lanes, road boundaries) combined into a generic environment model? If so, please give a brief description of this model.

Function

This section addresses functions and maneuvers implemented in your test vehicle.

1. Is the system capable of autonomously executing lane-change maneuvers without any support of the driver? How is this maneuver implemented? In which domains (highway, rural roads, and urban environment) can it be executed?
2. How does the system react to the state of traffic lights?
3. What kind of turn maneuvers are implemented? Is the system capable of executing a turning maneuver into moving traffic?
4. How does the system deal with intersections without traffic lights?
5. Is the system capable of merging into traffic on rural roads and/or highways? How is this maneuver implemented?
6. What kind of emergency situations are considered by the system (e.g., emergency braking of the vehicle in front, pedestrian crossing the street)? How does the system react to these situations?
7. What concepts are implemented for keeping the vehicle in the current lane?
8. Are there any implicit and/or explicit mechanisms for cooperating with other road users?
9. How is the mission planning implemented? Is it done online or is the mission plan precomputed?
10. How does the system react to interventions of the driver?
11. Are there any further, not yet addressed, capabilities or maneuvers implemented?
12. Is the system capable of traversing from one domain to another (e.g., exiting the highway onto an urban street)? Which transitions are possible?

Safety Concept

1. Please describe briefly the safety concept for driving on public roads.
2. How do you verify that the functions discussed above work as expected? What is the test procedure?

3. How does the system react to the loss of one or more components or capabilities? What is the degradation concept?

System Architecture

Please describe the architecture of the system (functional, hardware, software) and the main design criteria.

Something Is Missing?

If there are any further system characteristics or features worth mentioning, please note them here.

References

Aeberhard M, Paul S, Kaempchen N, Bertram T (2011) Object existence probability fusion using dempster-shafer theory in a high-level sensor data fusion architecture. In: Intelligent vehicles symposium, IEEE, pp 770–775

Ardelt M, Waldmann P (2011) Hybrides Steuerungs- und Regelungskonzept für das hochautomatisierte Fahren auf Autobahnen (Hybrid open- and closed-loop control concept for highly automated driving on highways). Automatisierungstechnik 59(12):738–750

Ardelt M, Coester C, Kaempchen N (2012) Highly automated driving on freeways in real traffic using a probabilistic framework. IEEE Trans Intell Transp Syst 13(4):1576–1585

Bacha A, Bauman C, Faruque R, Fleming M, Terwelp C, Reinholtz C, Hong D, Wicks A, Alberi T, Anderson D (2008) Odin: Team VictorTango's entry in the DARPA Urban Challenge. J Field Robot 25(8):467–492

Bar Hillel A, Lerner R, Levi D, Raz G (2012) Recent progress in road and lane detection: a survey. Mach Vis Appl 25(3):727–745

Becker C, Dürr F (2005) On location models for ubiquitous computing. Pers Ubiquit Comput 9:20–31, Springer

Bertozzi M, Broggi A, Coati A, Fedriga RI (2013) A 13,000 km intercontinental trip with driverless vehicles: the VIAC experiment. Intell Transp Syst Mag 5(1):28–41

Bley O, Kutzner R, Friedrich B, Saust F, Wille J M, Maurer M, Wolf F, Naumann S, Junge M, Langenberg J, Niebel W, Schüler T, Bogenberger K (2011) Kooperative Optimierung von Lichtsignalsteuerung und Fahrzeugführung (Cooperative and optimised traffic signalling in urban networks). In: AAET 2011 – Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel (Automation systems, driver assistance systems and embedded systems for transport technologies), pp 57–77

Broggi A, Bertozzi M, Fascioli A (1999) ARGO and the MilleMiglia in Automatico tour. IEEE Intell Syst Appl 14(1):55–64

Broggi A, Buzzoni M, Debattisti S, Grisleri P, Laghi MC, Medici P, Versari P (2013) Extensive tests of autonomous driving technologies. IEEE Trans Intell Transp Syst 14(3):1403–1415

Brown PJ (1996) The stick-e document: a framework for creating context-aware applications. Electron Publ Chichester 8:259–272

Buschardt B, Donner E, Graab B, Hörauf U, Winkle T (2006) Analyse von Verkehrsunfällen mit FAS Potenzialeinschätzung am Beispiel des FAS Lane Departure Warning (Analysis of traffic accidents with ADAS potential estimation on the example of the ADAS lane departure warning) In: Tagung Aktive Sicherheit 2006. Technische Universität München, Lehrstuhl für Fahrzeugtechnik, München

Chatila R, Laumond J (1985) Position referencing and consistent world modeling for mobile robots. Proc IEEE Int Conf Robot Autom 2:138–145

Chiellino U, Winkle T, Graab B, Ernstberger A, Donner E, Nerlich M (2010) Was können Fahrerassistenzsysteme im Unfallgeschehen leisten? (What can driver assistance systems do in the event of an accident?). Zeitschrift für die gesamte Wertschöpfungskette Automobilwirtschaft 3:131–137, TÜV Media GmbH, Köln

DARPA – Defense Advanced Research Projects Agency (2007) Urban challenge – route network definition file (RNDF) and mission data file (MDF) formats. http://archive.darpa.mil/grandchallenge/docs/RNDF_MDF_Formats_031407.pdf

Dickmanns ED, Behringer R, Hildebrandt T, Maurer M, Thomanek F, Schiehlen J (1994) The seeing passenger car 'VaMoRs-P'. In: Intelligent vehicles symposium, IEEE, pp 68–73

Dickmanns ED (2007) Dynamic vision for perception and control of motion. Springer

Dickmanns ED (2015) Personal communication

Donges E (1999) A conceptual framework for active safety in road traffic. Veh Syst Dyn 32(2–3):113–128

Eskandarian A (ed) (2012) Handbook of intelligent vehicles. Springer

Fenton R (1970) Automatic vehicle guidance and control – a state of the art survey. IEEE Trans Veh Technol 19(1):153–161

Flament M, Otto HU, Alksic M, Guarise A, Löwenau J, Beuk L, Meier J, Sabel H (2005) ActMAP final report/Ertico. Forschungsbericht, D1.2

Fries C, Luettel T, Wuensche HJ (2013) Combining model-and template-based vehicle tracking for autonomous convoy driving. In: Intelligent vehicles symposium, IEEE, pp 1022–1027

Funke J, Theodosis P, Hindiyeh R, Stanek G, Kritatakirana K, Gerdes C, Langer D, Hernandez M, Muller-Bessler B, Huhnke B (2012) Up to the limits: autonomous Audi TTS. In: Intelligent vehicles symposium, IEEE, pp 541–547

Gasser TM, Arzt C, Ayoubi M, Bartels A, Bürkle L, Eier J, Flemisch F, Häcker D, Hesse T, Huber W, Lotz C, Maurer M, Ruth-Schumacher S, Schwarz J, Vogt W (2012) Rechtsfolgen zunehmender Fahrzeugautomatisierung (Legal consequences of increasing levels of vehicle automation). Berichte der Bundesanstalt für Straßenwesen F83. Wirtschaftsverlag NW, Bergisch Gladbach

Geiger A, Lauer M, Moosmann F, Ranft B, Rapp H, Stiller C, Ziegler J (2012) Team AnnieWAY's entry to the grand cooperative driving challenge 2011. IEEE Trans Intell Transp Syst 13 (3):1008–1017

Ghosh D, Sharman R, Raghav Rao H, Upadhyaya S (2007) Self-healing systems – survey and synthesis. Decis Support Syst Emerg Econ 42(4):2164–2185

Goebl M, Althoff M, Buss M, Färber G, Hecker F, Heißing B, Kraus S, Nagel R, León FP, Rattei F, Russ M, Schweitzer M, Thuy M, Wang C, Wünsche HJ (2008) Design and capabilities of the Munich cognitive automobile. In: Intelligent vehicles symposium, IEEE, pp 1101–1107

Goldhammer M, Strigel E, Meissner D, Brunsmann U, Doll K, Dietmayer K (2012) Cooperative multi sensor network for traffic safety applications at intersections. In: Intelligent transportation systems conference, IEEE, pp 1178–1183

Gregor R (2002) Fähigkeiten zur Missionsdurchführung und Landmarkennavigation (Abilities for mission execution and landmark navigation). Dissertation, Universität der Bundeswehr

Himmelsbach M, Wünsche HJ (2012) Tracking and classification of arbitrary objects with bottom-up/top-down detection. In: Intelligent vehicles symposium, IEEE, pp 577–582

Hitchcock A (1995) Intelligent vehicle/highway system safety: multiple collisions in automated highway systems. Forschungsbericht, University of California

Hock CJL (1994) Wissensbasierte Fahrzeugführung mit Landmarken für autonome Roboter (Knowledge-based vehicle guidance with landmarks for autonomous robots). Dissertation, Universität der Bundeswehr

Homm F, Kaempchen N, Burschka D (2011) Fusion of laserscannner and video based lanemarking detection for robust lateral vehicle control and lane change maneuvers. In: Intelligent vehicles symposium, IEEE, pp 969–974

Hörwick M, Siedersberger, KH (2010a) Aktionspläne zur Erlangung eines sicheren Zustandes bei einem autonomen Stauassistenten (Action plans for achieving a safe state for an autonomous traffic jam assistance system) In: 4. Tagung Fahrerassistenz

Hörwick M, Siedersberger KH (2010b) Strategy and architecture of a safety concept for fully automatic and autonomous driving assistance systems. In: Intelligent vehicles symposium, IEEE, pp 955–960

Huang AS, Moore D, Antone M, Olson E, Teller S (2009) Finding multiple lanes in urban road networks with vision and lidar. Auton Robot 26(2–3):103–122

Hundelshausen Fv, Himmelsbach M, Hecker F, Müller A, Wünsche H-J (2009) Driving with tentacles – integral structures for sensing and motion. Springer Tracts Adv Robot 56:393–440

Isermann R (2006) Fault-diagnosis systems: an introduction from fault detection to fault tolerance. Springer, Berlin/Heidelberg

Isermann R, Schwarz R, Stölzl S (2002) Fault-tolerant drive-by-wire systems. IEEE Control Syst 22(5):64–81

Kammel S, Ziegler J, Pitzer B, Werling M, Gindele T, Jagzent D, Schröder J, Thuy M, Goebl M, Hundelshausen Fv, Pink O, Frese C, Stiller C (2008) Team AnnieWAY's autonomous system for the 2007 DARPA Urban Challenge. J Field Robot 25(9):615–639

Kim J, Rajkumar R, Jochim M (2013) Towards dependable autonomous driving vehicles: a system-level approach. SIGBED 10:29–32

Knapp A, Neumann M, Brockmann M, Walz R, Winkle T (2009) Code of practice for the design and evaluation of ADAS. Preventive and Active Safety Applications, eSafety for road and air transport, European Commission Integrated Project, Response 3

Lategahn H, Stiller C (2012) Experimente zur hochpräzisen landmarkenbasierten Eigenlokalisierung in unsicherheitsbehafteten digitalen Karten (Experiments for highly precise landmark-based self-localization in uncertain digital maps). In: Workshop Fahrerassistenzsysteme Walting

Laurgeau C (2012) Intelligent vehicle potential and benefits. In: Eskandarian A (ed) Handbook of intelligent vehicles. Springer, London, pp 1537–1551

Leonard J, How J, Teller S, Berger M, Campbell S, Fiore G, Fletcher L, Frazzoli E, Huang A, Karaman S (2008) A perception-driven autonomous urban vehicle. J Field Robot 25(10):727–774

Levinson JS (2011) Automatic laser calibration, mapping, and localization for autonomous vehicles. Dissertation, Stanford University

Levinson J, Askeland J, Becker J, Dolson J, Held D, Kammel S, Kolter JZ, Langer D, Pink O, Pratt V, Sokolsky M, Stanek G, Stavens D, Teichman A, Werling M, Thrun S (2011) Towards fully autonomous driving: systems and algorithms. In: IEEE Intelligent vehicles symposium, IEEE, pp 163–168

Luettel T, Himmelsbach M, Hundelshausen F, Manz M, Mueller A, Wuensche HJ (2009) Autonomous offroad navigation under poor GPS conditions. In: Planning, perception and navigation for intelligent vehicles, p 56

Luettel T, Himmelsbach M, Manz M, Mueller A, Hundelshausen F, Wuensche HJ (2011) Combining multiple robot behaviors for complex off-road missions. In: Intelligent transportation systems conference, IEEE, pp 674–680

Luettel T, Himmelsbach M, Wuensche HJ (2012) Autonomous ground vehicles' concepts and a path to the future. Proc IEEE 100(13):1831–1839

Manz M (2013) Modellbasierte visuelle Wahrnehmung zur autonomen Fahrzeugführung (Model-based visual perception for autonomous vehicle guidance). Dissertation, Universität der Bundeswehr München

Manz M, Himmelsbach M, Luettel T, Wuensche HJ (2011) Detection and tracking of road networks in rural terrain by fusing vision and LIDAR. In: Intelligent robots and systems, IEEE/RSJ, pp 4562–4568

Matthaei R (2015) Wahrnehmungsgestützte Lokalisierung in fahrstreifengenauen Karten für Assistenzsysteme und autonomes Fahren (Perception-based localization in lane-level maps for ADAS and autonomous driving). Dissertation, Braunschweig, accepted/forthcoming

Matthaei R, Maurer M (2015) Autonomous driving – a top-down approach. Automatisierungstechnik 63(4):155–167 accepted

Matthaei R, Bagschik G, Maurer M (2014a) Map-relative localization in lane-level maps for ADAS and autonomous driving. In: Intelligent vehicles symposium, IEEE, pp 49–55

Matthaei R, Bagschik G, Rieken J, Maurer M (2014b) Stationary urban environment modeling using multi-layer-grids. In: Proceedings of the 17th international conference on information fusion

Matthaei R, Reschka A, Bagschik G, Escher M, Menzel T, Rieken J, Scheide T, Schuldt F, Ulbrich S, Wendler JT, Hecker P, Maurer M (2015) Das Projekt Stadtpilot – Automatisiertes Fahren an der TU Braunschweig (The project Stadtpilot – automated driving at TU Braunschweig). Z Gesamte Wertschöpfungskette Automobilwirtschaft 18(1):12–22

Maurer M (2000) Flexible Automatisierung von Straßenfahrzeugen mit Rechnersehen (Flexible automation of road vehicles with computer vision). VDI-Verlag, Düsseldorf

Maurer M (2013) Autonome Automobile – Wer steuert das Fahrzeug der Zukunft? (Who drives the vehicle of the future?) Daimler und Benz Stiftung, Vortrag Forschungsprojekt „Villa Ladenburg“, Untertürkheim. https://www.daimler-benz-stiftung.de/cms/images/dbs-bilder/foerderprojekte/villa-ladenburg/Dialog_im_Museum_Vortrag_Prof_Maurer.pdf

Miller I, Campbell M, Huttenlocher D, Kline FR, Nathan A, Lupashin S, Catlin J, Schimpf B, Moran P, Zych N, Garcia E, Kurdziel M, Fujishima H (2008) Team Cornell’s Skynet: Robust perception and planning in an urban environment. J Field Robot 25(8):493–527

Montemerlo M, Becker J, Bhat S, Dahlkamp H, Dolgov D, Ettinger S, Haehnel D, Hilden T, Hoffmann G, Huhnke B, Johnston D, Klumpp S, Langer D, Levandowski A, Levinson J, Marcil J, Orenstein D, Paefgen J, Penny I, Petrovskaya A, Pflueger M, Stanek G, Stavens D, Vogt A, Thrun S (2008) Junior: the Stanford entry in the Urban Challenge. J Field Robot 25 (9):569–597

Moore DC, Huang AS, Walter M, Olson E, Fletcher L, Leonard J, Teller S (2009) Simultaneous local and global state estimation for robotic navigation. In: International conference on robotics and automation, IEEE, pp 3794–3799

Müller A, Himmelsbach M, Lüttel T, Hundelshausen Fv, Wünsche HJ (2011) GIS-based topological robot localization through LIDAR crossroad detection. In: Intelligent transportation systems conference, IEEE, pp 2001–2008

NDMV – Nevada Department of Motor Vehicles (2012) Adopted regulation of the Department of Motor Vehicles LCB File No R084-11

Nilsson JO, Zachariah D, Skog I (2012) Global navigation satellite systems: an enabler for in-vehicle navigation. In: Eskandarian A (ed) Handbook of intelligent vehicles. Springer, London, pp 311–342

Nothdurft T, Hecker P, Ohl S, Saust F, Maurer M, Reschka A, Böhmer JR (2011a) Stadtpilot: first fully autonomous test drives in urban traffic. In: Intelligent transportation systems conference, IEEE, pp 919–924

Nothdurft T, Hecker P, Frankiewicz T, Gacnik J, Köster F (2011b) Reliable information aggregation and exchange for autonomous vehicles. In: Vehicular technology conference, IEEE, pp 1–5

Nunen Ev, Kwakkernaat R, Ploeg J, Netten B (2012) Cooperative competition for future mobility. IEEE Trans Intell Transp Syst 13(3):1018–1025

Pellkofer M (2003) Verhaltensentscheidung für autonome Fahrzeuge mit Blickrichtungssteuerung (Behavior decision-making for autonomous vehicles with gaze control). Dissertation, Universität der Bundeswehr München

Pomerleau D, Jochem T (1996) Rapidly adapting machine vision for automated vehicle steering. IEEE Expert 11(2):19–27

Rauch A, Maier S, Klanner F, Dietmayer K (2013) Inter-vehicle object association for cooperative perception systems. In: Intelligent transportation systems conference, IEEE, pp 893–898

Rauskolb FW, Berger K, Lipski C, Magnor M, Cornelsen K, Effertz J, Form T, Graefe F, Ohl S, Schumacher W, Wille JM, Hecker P, Nothdurft T, Doering M, Homeier K, Morgenroth J, Wolf L, Basarke C, Berger C, Gülke T, Klose F, Rumpe B (2008) Caroline: an autonomously driving vehicle for urban environments. J Field Robot 25(9):674–724

Reschka A (2015) Safety concept for autonomous vehicles. In: Maurer M, Gerdes JC, Lenz B, Winner H (eds) Autonomous driving. Springer Vieweg, forthcoming, Springer-Verlag GmbH Berlin Heidelberg, Berlin Heidelberg

Reschka A, Böhmer JR, Nothdurft T, Hecker P, Lichte B, Maurer M (2012a) A surveillance and safety system based on performance criteria and functional degradation for an autonomous vehicle. In: Intelligent transportation systems conference, IEEE, pp 237–242

Reschka A, Böhmer, JR, Saust F, Lichte B, Maurer M (2012b) Safe, dynamic and comfortable longitudinal control for an autonomous vehicle. In: Intelligent vehicles symposium, IEEE, pp 346–351

Rosenblatt J (1997) DAMN: a distributed architecture for mobile navigation. Dissertation, Robotics Institute, Carnegie Mellon University

SAE – Society of Automotive Engineers (2014) Taxonomy and definitions for terms related to on-road motor vehicle automated driving system. SAE International, Warrendale

Saust F, Bley O, Kutzner R, Wille JM, Friedrich B, Maurer M (2010) Exploitability of vehicle related sensor data in cooperative systems. In: Intelligent transportation systems conference, IEEE, pp 1724–1729

Saust F, Wille JM, Maurer M (2012) Energy-optimized driving with an autonomous vehicle in urban environments. In: Vehicular technology conference, IEEE, pp 1–5

Shladover S (2007) PATH at 20 – history and major milestones. IEEE Trans Intell Transp Syst 8(4):584–592

Siedersberger KH (2003) Komponenten zur automatischen Fahrzeugführung in sehenden (semi-) autonomen Fahrzeugen (Components for automated vehicle guidance in perceiving (semi-) autonomous vehicles). Dissertation, Universität der Bundeswehr München

Singh, S (ed) (2006) Special Issue on the DARPA Grand Challenge, Part 2. J Field Robot 23 (9):655–835. Wiley Subscription Services, Inc., 2008

Singh, S (ed) (2008a) Special Issue on the 2007 DARPA Urban Challenge Part I. J Field Robot 25 (8):432–566. Wiley Subscription Services, Inc., 2008

Singh, S (ed) (2008b) Special Issue on the 2007 DARPA Urban Challenge Part II. J Field Robot 25 (9):567–724. Wiley Subscription Services, Inc., 2008

Singh, S (ed) (2008c) Special Issue on the 2007 DARPA Urban Challenge Part III. J Field Robot 25(10):725–860. Wiley Subscription Services, Inc., 2008

Smith BW (2014) Lawyers and engineers can speak the same robot language. In: Robot law. Forthcoming, Edward Elgar Publishing Ltd., Cheltenham, UK

Spieß E (2014) Kooperation. In: Wirtz MA (ed) Dorsch – Lexikon der Psychologie (Dorsch – Lexicon of psychology). https://portal.hogrefe.com/dorsch/kooperation/. Accessed 05 Dec 2014

Stanek G, Langer D, Müller-Bessler B, Huhnke B (2010) Junior 3: a test platform for advanced driver assistance systems. In: Intelligent vehicles symposium, IEEE, pp 143–149

Stiller C, Burgard W, Deml B, Eckstein L, Flemisch F, Köster F, Maurer M, Wanielik G (2013) Kooperativ interagierende Automobile (Cooperative interacting automobiles). Schwerpunktprogramm der Deutschen Forschungsgemeinschaft (DFG priority program)

Strang T, Linnhoff-Popien C (2004) A context modeling survey. In: Workshop on advanced context modeling, reasoning and management, 6th international conference on ubiquitous computing, pp 1–8

Thorpe C, Jochem T, Pomerleau D (1997) The 1997 automated highway free agent demonstration. In: Intelligent transportation system conference, IEEE, pp 496–501

Thrun S, Montemerlo M, Dahlkamp H, Stavens D, Aron A, Diebel J, Fong P, Gale J, Halpenny M, Hoffmann G, Lau K, Oakley C, Palatucci M, Pratt V, Stang P, Strohband S, Dupont C, Jendrossek L-E, Koelen C, Markey C, Rummel C, Niekerk Jv, Jensen E, Alessandrini P, Bradski G, Davies B, Ettinger S, Kaehler A, Nefian A, Mahoney P (2006) Stanley: the robot that won the DARPA Grand Challenge. J Field Robot 23(9):661–692

Tsugawa S (1993) Vision-based vehicles in Japan: the machine vision systems and driving control systems. In: IEEE international symposium on industrial electronics. Conference proceedings, ISIE'93 – Budapest, pp 278–285

Tsugawa S (1994) Vision-based vehicles in Japan: machine vision systems and driving control systems. IEEE Trans Ind Electron 41(32):398–405

Ulbrich S, Maurer M (2013) Probabilistic online POMDP decision making for lane changes in fully automated driving. In: Intelligent transportation systems, IEEE, pp 2063–2067

Ulmer B (1992) VITA-an autonomous road vehicle (ARV) for collision avoidance in traffic. In: Intelligent vehicles symposium, IEEE, pp 36–41

Ulmer B (1994) VITA II-active collision avoidance in real traffic. In: Intelligent vehicles symposium, IEEE, pp 1–6

Urmson C, Anhalt J, Bagnell D, Baker C, Bittner R, Clark MN, Dolan J, Duggins D, Galatali T, Geyer C, Gittleman M, Harbaugh S, Hebert M, Howard TM, Kolski S, Kelly A, Likhachev M, McNaughton M, Miller N, Peterson K, Pilnick B, Rajkumar R, Rybski P, Salesky B, Seo Y-W, Singh S, Snider J, Stentz A, Whittaker WR, Wolkowicki Z, Ziglar J, Bae H, Brown T, Demitrish D, Litkouhi B, Nickolaou J, Sadekar V, Zhang W, Struble J, Taylor M, Darms M, Ferguson D (2008) Autonomous driving in urban environments: Boss and the Urban Challenge. J Field Robot 25(8):425–466

Visintainer F, Darin M (2008) Final requirements and strategies for map feedback/Ertico. Forschungsbericht, D2.2

Wachenfeld W, Winner H, Gerdes C, Lenz B, Maurer M, Beiker S, Fraedrich E, Winkle T (2015) Use cases for autonomous driving. In: Autonomous driving – technical, legal and social aspects. Springer, Berlin/Heidelberg

Wei J, Dolan JM, Litkouhi B (2010) A prediction- and cost function-based algorithm for robust autonomous freeway driving. In: 2010 I.E. intelligent vehicles symposium, IEEE, pp 512–517

Wei J, Dolan JM, Snider JM, Litkouhi B (2011) A point-based MDP for robust single-lane autonomous driving behavior under uncertainties. In: International conference on robotics and automation, IEEE, pp 2586–2592

Werling M, Ziegler J, Kammel S, Thrun S (2010) Optimal trajectory generation for dynamic street scenarios in a frenèt frame. In: International conference on robotics and automation, IEEE, pp 987–993

Wille JM (2012) Manöverübergreifende autonome Fahrzeugführung in innerstädtischen Szenarien am Beispiel des Stadtpilotprojekts (Maneuver-comprehensive autonomous vehicle guidance in urban scenarios exemplified by the Stadtpilot-project). Dissertation, Technische Universität Braunschweig

Winkle T (2015a) Safety benefits of automated vehicles: extended findings from accident research for development, validation and testing. In: Maurer M, Gerdes C, Lenz B, Winner H (eds) Autonomous driving – technical, legal and social aspects. Springer, Berlin/Heidelberg

Winkle T (2015b) Development and approval of automated vehicles: considerations of technical, legal and economic risks. In: Maurer M, Gerdes C, Lenz B, Winner H (eds) Autonomous driving – technical legal and social aspects. Springer, Berlin/Heidelberg

Zapp A (1988) Automatische Straßenfahrzeugführung durch Rechnersehen (Automated road vehicle guidance with computer vision). Dissertation, Universität der Bundeswehr München

Zhang W-B, Parsons RE, West T (1990) An intelligent roadway reference system for vehicle lateral guidance/control. In: American control conference, pp 281–286

Ziegler J, Bender P, Lategahn H, Schreiber M, Strauß T, Stiller C (2014) Kartengestütztes automatisiertes Fahren auf der Bertha-Benz-Route von Mannheim nach Pforzheim (Map-supported automated driving on the Bertha-Benz-Route from Mannheim to Pforzheim). In: Workshop Fahrerassistenzsysteme Walting

ADAS, Quo Vadis?

63

Hermann Winner

Contents

Abstract

Gazing into the future of driver assistance systems means looking for stimuli for future developments and identifying specific challenges and effects DAS may be exposed to. One specific “problem area” is the validation of autonomous driving. Statistical considerations show how very difficult it will be to obtain proof of

---

H. Winner (✉)
Institute of Automotive Engineering, Technische Universität Darmstadt, Darmstadt, Germany
e-mail: winner@fzd.tu-darmstadt.de

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3_62

safety on a level comparable to human drivers and that a metric is needed to be able to certify autonomous driving in future. The evolution to autonomous driving is shown as a triangle with three different starting bases. The priorities for future research are given in the last section.

1 Introduction

In the first edition of the German handbook for driver assistance systems in 2009, most driver assistance systems described were already in series production. However, apart from a few exceptions, such as brake assist, parking assist, and navigation, their actual penetration in the market was still very low. In recent years, advances in technology and manufacturing have led to significant reductions in production costs with the result that currently available assistance packages with four or five major functions often cost the car buyer no more than an individual function would have cost previously. In addition, as described in ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems" due to consumer tests, such as the NCAP rating and regulatory requirements for commercial heavy goods vehicles, such vehicles are already fitted with assistance functions as standard equipment. Knowing these trends, it is not difficult to predict that driver assistance systems will become a matter of course in new vehicles and, along with measures for driving efficiency, they will be responsible for the highest increase in value in road vehicles. Relating to the ambitions of pioneering developers, we could say mission accomplished. But of course, development has not come to an end and, as will be shown here, it is not merely the final step to autonomous driving that is missing. On the one hand, when considering present-day designs, it is still possible to identify many incremental improvements which it is not our intention to discuss in detail here. On the other hand, the development of driver assistance systems is influenced by other technological developments and, at least as importantly, by interaction with the developments in society. These stimuli to development were analyzed by members of the Uni-DAS e. V. association in 2012 and described in a position paper (Bengler et al. 2012) (see also a review article resulting from this (Blumenthal 2013)). The contents of two following sections are fully based on this original work and also cite large parts of it. This edition of the handbook already shows that test methods are playing a greater role. They are the focus for the newly added chapters. However, since still considerably more has to be done to achieve autonomous driving, as pointed out in previous issues at this point, we have devoted a detailed section to this aspect and now also explain the basic statistical concepts used for dimensioning validation routes. Once again a view of the evolution of driver assistance systems is presented, albeit in a new illustration as a triangle of autonomous driving. Finally, again taken from the position paper, very specific recommendations are given for future research, and it becomes more than clear that this topic still has much potential for the future, although a great deal of research work will be involved.

2 Stimuli for Future Developments

2.1 Data Communication

So far, the Internet has played an only marginal role in vehicles. Up until now, the use of data links has been restricted mainly to infotainment and navigation support. In the future, along with developments in the infotainment area, new driving assistance capabilities can be expected. The “driving office” is certainly an interesting concept for managers and business people. It may be assumed that the mobile office and autonomous driving package will be in high demand for company cars. Other potential uses for data communication include the allocation of parking spaces before the actual arrival of the vehicle at the parking lot or communication in intermodal traffic.

Communication-based DAS for guidance or stabilization support on the other hand require an independent network concept. In field operational tests, like SIM-TD (www.simtd.org), and research projects, like Ko-FAS (www.kofas.de) and Koline (Saust et al. 2012), the foundations for a comprehensive implementation of such technology are being laid down. Integrating all traffic participants of a certain area into a common network, a new stage of driving assistance can be realized, based on the vastly improved quality and quantity of information about the local traffic situation. This would, for example, impact traffic infrastructure considerably (Tank and Linnartz 1997; Tischler and Hummel 2005; Nagel et al. 2007; Dietl et al. 2001; Eichler 2007; Kosch 2004).

A traffic light, for example, could be replaced by a wireless access point that directs the vehicles through the junction. While this approach would be more efficient and effective with automated vehicles than with human operators, it should yield positive results irrespective of who is driving the vehicle.

Vehicles equipped with sensors and v2v communication devices could expand their horizon via cooperative sensing. Given sufficient bandwidth and integrity of the data sources and the communication network, information from all the vehicles and the infrastructure, if available, could then be fused into a detailed dynamic map, as demonstrated in the EU project DRIVE C2X (www.drive-c2x.eu). In analogy to the IT-cloud concept, shared sensor data could be described as cloud sensors.

Of course, it is also possible to integrate information sources from “the big cloud,” as can already be seen from the diverse activities of Google & Co. In a cloud-sensor technology concept, the demands on the local environment-sensor technology might be lowered, e.g., in terms of reach, so that presumably, the overall economic costs are more likely to fall, despite the investment in cloud technology.

A necessary precondition, however, is a reliable network with a high quality of service and integrity, which is yet to be developed.

Seeing the enormous potential to increase driving safety and efficiency with respect to energy, time, and traffic infrastructure, such networks will hopefully soon become a reality for the benefit of all traffic participants (European Union 2010).

2.2 Electromobility

Electromobility, too, presents new challenges for driver assistance. It is very important, therefore, to secure a combined parking and loading space in good time. Guaranteeing the energy required for the intended journey will probably remain a challenge to electromobility for decades with the result that the type and manner of use will be different from today's.

This relates to the most efficient use possible of the "gas" pedal both from the microscopic angle and from the macroscopic point of view in terms of which means of transport or which business model will be used for which transport task. Accordingly, assistance functions will have to adapt to altered basic conditions of transportation usage and provide additional functions for intermodal mobile assistants and range extension.

One approach toward improved energy efficiency, discussed particularly often in the e-mobility sector, is the reduction of vehicle weight. Besides the weight of the engine and the transmission, a high proportion of a vehicle's mass is attributed to passive passenger safety. Increasing the emphasis on and improving the performance of active and integrated safety systems can thus pave the way to a significant reduction of vehicle weight.

2.3 Societal Changes and Market Trends

Technology changes society, as is abundantly clear from the famous example of the industrial revolution in the nineteenth century. But technology can also be seen as reflecting society: It reflects demand which, when technology can reflect it at acceptable costs, is also satisfied. In affluent societies, this demand is by no means geared to elementary needs. The technology used more or less consciously reflects the individual "way of life." If this changes over the generations, the products also change with them, often in parallel with technological progress: Two obvious trends are the change in demographics and the redefinition of status symbols.

The world of older people is currently changing drastically. Often weakened in the past by austerity and difficult working conditions, they mostly remained in a family environment with fewer outwardly visible activities. Today, the family environment is increasingly diminishing, either through greater residential mobility, increasing alienation, or childlessness. Instead, the proportion of "mobile older people" and "active older people" is continually increasing. This generation that has used individual mobility almost their whole lives will not only *want* to maintain it as long as possible, but they will have to maintain it as working lives become longer. Moreover, the trend reminds us that, while hard to imagine today, future generations accustomed to technology like "the cloud" will be demanding intelligent vehicles with high support for driving tasks.

Another major societal change regarding mobility occurs within the young generation. Having grown up with a high degree of individual mobility as a common standard, they tend to take mobility for granted. In conjunction with increasing urbanization, the car as an important symbol of societal status is being replaced by other values, such as real estate, group affiliation, or design icons.

The consequences of this trend are not yet clear. It could lead to possessing a vehicle becoming increasingly less important in the course of increasingly rational choice of transport. However, changes could also come about to the effect that cars with special design features will become attractive and by analogy they might then be marketed as “iCars.” The key factors for success in all such waves of products were radically modified operating concepts. These new operating paradigms then dominated competitor products in rapid succession, with the result that a product group changed so vastly, often within fewer than three product generations, that everything previous was no longer marketable.

In vehicle design, this could cause the steering wheel and pedal interaction concepts, developed more than 110 years ago, to be abandoned.

In the course of such a reinvention of driver-vehicle interaction, new elements like assistance functions and partial automation could find their way into the proverbial “iCar” not as an optional feature, but an integral and defining part. Provided such vehicles share the success of today’s models, conventional vehicles could quickly become “old” rather than “classic.”

Another market trend results from changes in the value-creation chain. Companies generate revenues by brokering product deliveries. The Apple App Store, for example, provides a distribution platform as its own investment, but does not take over the risk of product development and warranty obligation. Similarly, billions of euros can be earned through the routing of advertising. These platform business models lead to large monopolistic companies with a market power that forces the other branches of the trade chain into “eat-or-be-eaten behavior.” These business models have not yet been applied to the mobility sector in large scale. Currently, there are only a small number of successful mobility platforms like organized ridesharing or used-car portals. Smartphone-based approaches, such as the app “taxi.eu” (www.taxi.eu) or Uber (Uber Technologies Inc. 2015), demonstrate how the product mobility can turn into a brokered good.

As a result, platforms can also determine the equipment, which will significantly restrict individual choice. Such services can prove to be an obstacle or a catalyst for the further development of DAS. In vehicles optimized for cost reduction, requirements are likely to be fulfilled with minimal effort in the least expensive way possible, which could prove fatal for the budget for innovations. On the flip side, certain automation technologies, for example, driving to a parking lot, the next customer, or even a mobile ordering office for an online retailing corporation, can serve as the technical basis for future business models (Bläser et al. 2012; Terporten et al. 2012).

2.4 The Role of Culture and Media

Other stimuli for future developments can come from the adaptation of traditions or new developments from other cultures. Increased globalization accelerates the rate of such transfer. Regarding market and technology, German and Japanese companies currently dominate the driving assistance scene; this is primarily due to the customers and automobile companies being sufficiently willing and financially strong to invest in vehicle technology. As time passes, older, more saturated markets may be overtaken by newer, emerging ones. This results in a change of numbers, customer needs, and usage conditions, as well as the initial difficulty to appraise willingness to pay and foresee possible regulative interventions. For megacities in emerging markets, traffic jam assistance and mobile offices are certainly of high market interest to the affluent sections of the population.

Finally, the role of media should not be underestimated. It is obvious that the presentation of Google's self-driving car in the media has both changed the attitudes of users and the effort of established car manufacturers with respect to this technology.

3 Challenges and Effects

For future transportation technology, simple roadmaps showing different developmental steps can be derived. Usually, these plans culminate in an interconnected autonomous vehicle, able to drive unsupervised in any possible environment. Along the paths toward such a vehicle, many arduous issues of homologation and liability have to be addressed. Today's testing and approval methods are unsuitable for the evaluation of intelligent machines, and new metrics assessing the performance of driving robots are required. Some experts consider this an even greater challenge, than the development of artificial intelligence for autonomous driving itself.

Another aspect impeding advancements are their costs. Technological development requires large investments that can only be redeemed by an appropriate market demand. Even if the car as a product is not the stimulus for innovations, as shown by computer or communications technology, nevertheless, automotive-specific technologies need to be developed so that technology finds its way out of the laboratory or special applications into the car as a volume product, as the recent examples of ESC and ACC RADAR have shown. If the market does not accept the developed product, financing further developmental steps may prove difficult.

This risk is increased by several mediating influences on a product's way from its development to its end user. So-called specialist magazines often prefer to rave over the sound and power of a combustion engine instead of covering meaningful technological innovations in an appropriate manner. Even trade chains and car salespeople often fail to promote the technology properly (e.g., "It doesn't need ESC, it's already safe even without it"). With future products, therefore, this backlash must always be expected, causing the development risk to increase yet again. It is worth pointing out, however, that development of DAS has not always been user-oriented and is therefore, in part, co-responsible for low user acceptance (cf. ► Chap. 47, "Development Process of Forward Collision Prevention Systems"). In order to maximize the success of human-machine interaction, the increasing number of assistance functions requires the development of integrated display and control concepts, providing a consistent user interface.

As discussed in Sect. 3, development is not driven forward in a static society, but in one that is changing and which, in turn, expects other mobility products. Market response to changed conditions can result in a product line shift or entirely new business models, creating new markets and suppressing old ones. DAS appearing at the right time and along with a fitting business model can be a key element to revolutionize individual mobility. Such revolutionary market changes may prove especially challenging for the well-established German automotive industry. The example of the IT sector (dominated by companies like IBM/DEC/Nixdorf, then Microsoft/Intel/Nokia, then Google/Apple/Facebook) demonstrates that decades of success are transient if circumstances and business models change.

As long as cars are used the way they are today, the market is not expected to change remarkably. But the development of DAS, especially toward autonomous vehicles, opens up different usage options. However, because the development of driver assistance, specifically that relating to autonomous vehicles, offers different opportunities for use, it may trigger a change that threatens the car world as it dominates today. As, of course, we can’t stand in the way of progress, driving forces from other areas will ensure upheavals, as the activities of Google alone have shown. The automatic conclusion is that only proactive development which shapes the future will provide protection against this effect. It must not be too widely diversified; it needs a good scientific base to prevent undesirable developments, good boundary conditions for implementation, a positive technology climate, and open-minded observation of the market.

While future DAS are expected to contribute greatly to traffic safety and efficiency, they are also likely to entail side effects. Depending on the pace at which new assistance systems are implemented, a segregation of high-tech-assisted automobiles and still operable older vehicle models may occur to an extent surpassing the already given situation. While such conditions could act as an incentive to buy a new model, they could as well exacerbate discrimination and envy between vehicle holders. Since any market change will produce winners and losers, an impact assessment of technological innovations should be conducted beforehand, to ensure that technological progress does not get stifled and that its advantages are made visible.

Since they are a vision of the distant future, the effects of interconnected autonomous vehicles cannot yet be fully estimated. Traffic flow and traffic safety will increase, while “old” vehicles could be considered a traffic obstacle or safety risk. Here, a legal obligation to new technology may be worth discussing. On the plus side, the need for parking space close to a destination would become less relevant, as vehicles could drive to and from any external parking space by themselves; this would additionally benefit the environment, reducing the amount of land required to build parking lots. Transportation centrals could work as bookable resources managed within a network, which would create new opportunities for the industry as well as public authorities.

The new quality in freedom of movement gained, quite literally, by autonomous cars can easily be compared to the introduction of mobile telephony in the 1990s – a technology initially only available for a few users that has now become ubiquitous and in the process has completely changed communication and with it social life.

4 Problem Area: Validation of Autonomous Driving

Autonomous driving is understood to be handing over the vehicle guidance function and authority to a machine, also referred to in the following as a driving robot. The transfer of the guidance function may be for a limited location or a limited time and can potentially be interrupted by the driver. Basically, the autonomous vehicle is capable, without human assistance, of deciding the route, lane, and interventions in the driving dynamics (in the definition presented in ▶ Chap. 3, "Framework Conditions for the Development of Driver Assistance Systems" this corresponds to fully automated driving). Such a function is subject to both technical and social requirements. According to the legal principles in force now and in all likelihood in the future, an autonomous vehicle must pose no greater danger than a vehicle controlled by a person. This applies to all groups of road users and to all areas of use in which vehicles are driven by humans today.

The well-known concepts relating to autonomous driving are described in ▶ Chap. 62, "Autonomous Driving". So far, no concept has provided a strategy for validation; instead, it is repeatedly pointed out which routes or what length of route has been driven autonomously. Although initially this may be impressive, these routes are, nevertheless, virtually meaningless in terms of general use in traffic as the considerations outlined below will demonstrate.

4.1 Requirements for the Validation of Autonomous Driving in Widespread Use

If autonomous driving is actually intended to improve road safety by the widespread use of vehicles with such capabilities, the safety objective must be measured against the status quo of road safety. Since damage to property is offset against any benefits (e.g., working time, relaxation, entertainment), from the author's perspective, it is almost irrelevant for the safety assessment. Therefore, accidents involving personal injury remain the benchmark, with the result that it is necessary to validate the hypothesis:

> Due to the widespread use of autonomous driving, the damage in terms of the number of people injured and killed will not be higher than without this capability.

"Widespread" means that the use of autonomous driving is in the same order of magnitude as human-guided driving, to distinguish it from the trial phase (3–4 orders of magnitude lower) or the introductory phase (1–2 orders of magnitude lower). For the other two categories, it may be necessary to apply other standards if, for example, users were willing to accept a higher intrinsic risk due to the greater benefit of autonomous driving. An example of this are motorized two-wheelers, which are exposed to at least one order of magnitude higher risk of injury and death per distance traveled; users accept this risk, either from lack of alternatives or due to the pleasure of riding motorbikes. Similarly, it might be justifiable for people previously excluded from driving to be mobile in a vehicle that is unsafe compared to the unrestricted control group. But as long as the group with unsafe vehicles accounts for an insignificant proportion of the exposure for other road users, the additional hazard for other road users can be seen as irrelevant. There are two aspects, however, that make it difficult to set a benchmark: In today's private transport, drivers can influence the risk. With autonomous driving, they and the other passengers are passively exposed to the "external risk" of the driving robot, such as when traveling on public transport. In this area, however, the personal injury risk per distance is another order of magnitude lower, which may still lead, of course, to new discussions about the acceptance level of the risk.

The following section considers only the approach valid for widespread use and for which general current road safety, as yet without widespread use of autonomous driving, is taken as a reference.

4.1.1 Statistical Consideration

As input values of this calculation, the average number of accidents in each category i (e.g., with personal injury or, in greater detail, with slightly or seriously injured casualties or fatalities) is used per distance s_i of the reference system $a_{i,\mathrm{aut}} = k_{i,\mathrm{aut}}/s_{i,\mathrm{aut}}$ and the autonomous vehicle $a_{i,\mathrm{ref}} = k_{i,\mathrm{ref}}/s_{i,\mathrm{ref}}$. Thus, for a distance s, the result is an expected value of $k_{i,j}(s) = s \cdot a_{i,j}$. As a first result of the consideration, the aim is to discover which distance would be adequate in proportion to the reference level with a given number of registered accidents in order to demonstrate, for an acceptable significance level, that the accident risk of autonomous driving is no greater than it is in the reference case.

For the statistical calculations, the Poisson distribution $P_\lambda(k) = \lambda^k e^{-\lambda}/k!$ is used, which emerges for a non-exhaustive entity, that is to say as a limit case of the binomial distribution for infinite elements. Thus, it is possible to calculate the probability with which k events will occur for an expected value $\lambda = k$.

As Fig. 1 shows, occurrence figures which are significantly different from the expected value are not so uncommon: Thus, with the expected value of three events (accidents), there is approximately 5 % probability that no event (accident) occurred at all. Or in the case of an expected value of 6.3, 0–2 events (accidents) taken together are represented with a probability of 5 %.

If 5 % is assumed as an accepted significance level ε_{acc}, as is often the case in empirical science, it can be concluded that if no event occurs over three times the reference distance or no more than two occur over 6.3 times the distance, there is a 95 % probability that the expected value of the system would be ≤ 1 per reference distance. This can also be calculated for other numbers of events by solving

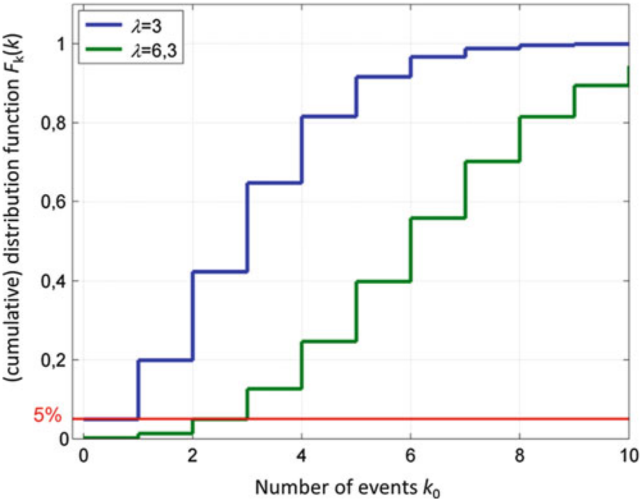


Fig. 1 Frequency of occurrence of a certain number of events k_0 for two different expected value λ

equation $\sum_{k=0}^{k_0} P_\lambda(k) = \varepsilon_{acc}$ numerically according to λ. Figure 2 shows the result for numbers of events k_0 from 0 to 5.

However, if autonomous driving should have the same risk as the reference, i.e., the expected values $\lambda_{i,ref} = \lambda_{i,aut}$ should match, then the case $k_0 = 0$ would be just as unlikely as the assumed significance level, that is to say only 5 % here, i.e., with a system that is equally good, one would need to have at least as much “luck” as the significance level. To have a realistic chance (e.g., 50 %) for the proof by demonstration, autonomous driving must have a much smaller expected value for an accident, but by how much? For this reason, one must look for the expected value factor $\alpha_{50\ \%}$, at which no more than k_0 events occur with a probability of 50 %. Accordingly, $\sum_{k=0}^{k_0} P_\lambda(k) = 50\ \%$ must then be solved according to λ. Figure 3 shows the result.

If the expected value factor is about a quarter, then over three times the reference distance the case with 0 events may actually occur with a probability of 50 %. With a system “only” twice as good as the reference system (= expected value factor of 0.5), however, four events occur which, according to Fig. 2, involve a reference distance more than nine times as long. Other combinations can be formed in exactly the same way, as shown in Fig. 4 for numbers of events up to five. From the semilogarithmic representation, it becomes clear that the distance factor grows exponentially in this range such that, from this alone, the deducible target is that a system which is to be qualified according to these standards should have an expected value factor ≤ 0.5, that is to say it should be at least twice as good as the reference system.

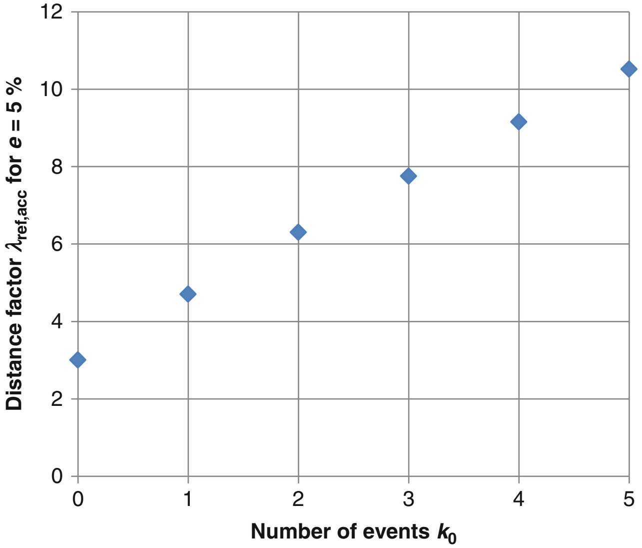


Fig. 2 Distance factor $\lambda_{ref,\ acc}$ by which the reference distance must be multiplied in order to demonstrate, with a significance level of 5 % and numbers of events k_0 occurring, that the expected value for accidents is less than 1/reference distance

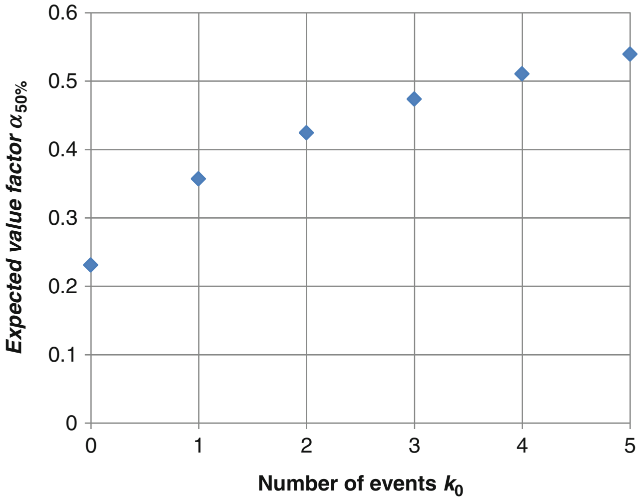


Fig. 3 Expected value factor $\alpha_{50\ \%}$ as a function of the specified maximum number of events k_0

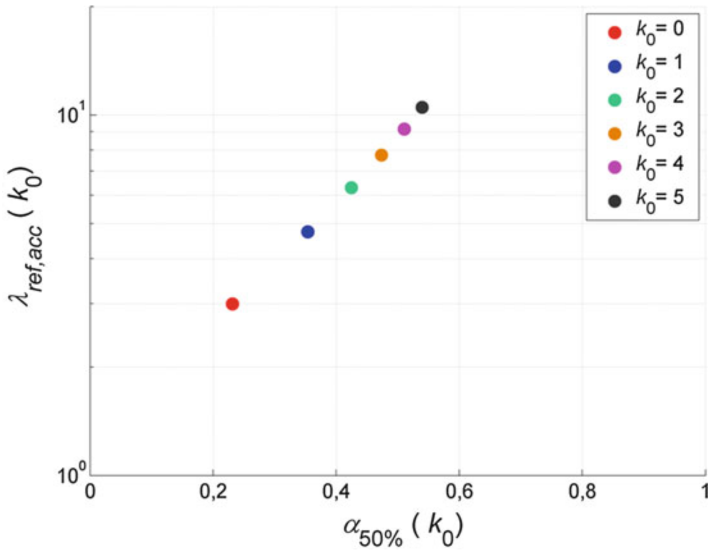


Fig. 4 Distance factor $\lambda_{ref,\ acc}$ for demonstration using 5 % significance level as a function of the expected value factor $\alpha_{50\ \%}$

For this condition (expected value factor of ≤ 0.5), a distance factor of roughly 10 can be assumed for the first test. Although with an expected value factor $\alpha_{50\,\%} \leq 0.23$, three times the distance would be sufficient for a 50 % probability; nevertheless, it is not possible to assume from a first instance using three times the distance that the same distance would be sufficient in case of recurrence, because even with $\alpha_{50\,\%} = 0.5$, it is possible to achieve the result of 0 events for three times the reference distance with 22 % probability. Only with about 10–20 times the reference distance is it possible to safely assume a basic assumption for the expected value factor that is so good that a lower distance factor would actually be sufficient in follow-up tests.

4.1.2 Reference Distance

Once the distance factor has generally been derived, it is then possible to define the reference distance for the approval of autonomous driving.

This theoretically necessary number of kilometers for the approval varies, as has been shown, with the control group and other factors. These factors, broken down into the categories below, will now be described:

Type of accident consequences (only property damage, personal injury: all types of injuries, only seriously injured casualties, only fatalities)

The accidents involving damage to property without personal injury, which occur far more frequently, cannot be used – at least not for the introductory period. This is because it is not known whether the ratio of accidents involving

Table 1 Reference distances as a function of the area of use and the consequences of accidents (Source: (Statistisches Bundesamt 2013; Lerner et al. 2013) (rounded)). Regarding information on seriously injured casualties and fatalities, the figures refer to distance per person. However, as more than one person per accident is affected in the category, the given value is a lower estimate

	Total	Extra-urban (w/o motorway)	Urban	Motorways
Distance traveled/billion km	724	110[a]		225
Total accidents/million	2.42	0.49	1.77	0.15
Personal injury/1000	291	73/23.5[a]	200	18.4
Distance between two accidents/million km				
Total	0.34			1.67
Involving personal injury	2.5	4.6[a]		12
Involving serious casualties	>11			>40
Involving fatalities	>200	>140[a]		>500

[a] Only nonurban highways

personal injury to accidents only involving property damage will remain unchanged even with the introduction of autonomous driving, with the result that there can be no extrapolation to accidents involving personal injury. Otherwise, damage to property during autonomous driving is offset, as previously mentioned, against the benefits, so that this represents a discussion more about economics than safety.

Accidents involving personal injury are predominantly accidents with light casualties. Thus, it is not worth making a distinction between the total number of accidents involving personal injury and that involving light casualties. Of course, accidents involving seriously injured casualties or fatalities have greater relevance. In number, they are less frequent by almost one (seriously injured casualties) or two (fatalities) orders of magnitude, as a result of which reference distances relating to this class rise correspondingly steeply.

Area of use (total, only extra-urban (w/o motorways), only urban, only motorways)

A limited area of use is to be expected, particularly when automated driving starts, e.g., only automation on motorways. Therefore, the reference values of these areas of use should be used. Table 1 shows the reference values for passenger cars in the reference area of Germany for 2013: They emerge from the distance traveled in the area of use divided by the number of accidents for each accident class. If the data were not appropriately available for all fields, they were estimated in accordance with the manner specified in the footnote.

Based on the figures shown in Table 1, two statistical dilemmas are evident:

1. The more severe, and therefore more relevant, the type of accident is, the larger the reference distance.
2. The simpler the function appears, e.g., for motorways, the further has to be driven to demonstrate safety.

Cause of accident (no distinction, only the main contributor)

On average, vehicle drivers are the main contributor in about 60 % of cases if they are involved in an accident, with the middle-age group as the main contributor appearing to be responsible for a lower proportion (approx. 50 %) than the groups of young and old drivers (cf. Statistisches Bundesamt 2013). If only the accidents in which drivers appear as the main contributors were used for a validation, the testing effort would initially increase by 5/3 in line with the extended distance between two accidents "caused," while it must be noted here that this distance is usually much further than people can drive in their lifetime.

But what if we still want to stick with the total number of accidents because the contributor split is irrelevant to safety. First, it is assumed that the number of accidents not caused will change only a little, for in most cases, they are still caused by human-driven vehicles. Thus, even with an autonomous vehicle that has caused no accidents, an expected value factor of 0.4 can be achieved at best. For the previously required expected value factor of 0.5, the autonomous vehicle may only cause accidents one sixth as frequently ($\alpha_{50\,\%} = 0.5 = 40\,\% + 60\,\%/6$). Alternatively, only an expected value factor ($\alpha_{50\,\%} = 0.7 = 40\,\% + 60\,\%/2$) remains for a vehicle which is twice as good in relation to the cause of accidents, which would result in an approximately three times higher distance factor $\lambda_{ref,acc} \approx 27$.

An argument against the sole choice of accidents caused is the uncertainty of the aforementioned assumption that the number of accidents not caused is not affected. It is still unclear whether the behavior of autonomous vehicles compensates for the mistakes of other road users more or less than human drivers did previously. It is conceivable that, in spite of significantly fewer accidents caused, the total number of accidents in which autonomous vehicles are involved will increase, which is the reason why, in turn, the overall number would have to be considered.

Reference vehicle group (current vehicle stock in each case, only advanced vehicles)

A distinction according to reference vehicle is also no easy task. The reference event is related to the accident frequency of all vehicles of a class (e.g., all cars). As past experience has shown, automotive progress has resulted in a higher safety level and therefore to a reduction in the number of accidents involving personal injury. The only recently started rollout to the broader market of safety systems that prevent and alleviate accidents means that we can expect a significant reduction in accidents involving personal injury. For this reason, it would be necessary to use a much higher reference distance (1.2–2 times higher) than that generated by the current incidence of accidents.

4.1.3 Example Calculation

In spite of the many options obtained when considering statistics and reference values, the size of the proof distance will be calculated on the basis of an example, although this is by no means a worst-case scenario:

For this, an automated motorway is selected. Thus, the current reference distance for accidents involving personal injury is 12 million km. If we then assume the number of main contributor accidents, a value which is more favorable for the statistics, and a very conservative improvement value for modern control vehicles of 1.2, the result is a reference distance of 24 million km (= 12 × 1.2/0.6). In addition, for the object under test (OUT), we assume only half the number of accidents caused involving personal injury (i.e., expected value factor $\alpha_{50\,\%} = 0.5$ and consequently distance factor $\lambda_{ref,acc} \approx 10$). The proof range thus multiplies to 240 million km. It has to be stressed in this context that this is only one of many other values; nevertheless, it is representative of the order of magnitude required in order to have a proof of safety "over distance" and if the aim is to reduce the risk compared to the control group.

4.1.4 Conclusions

Proof distances of the length outlined above exceed the technical, personal, and economic possibilities of today's companies. Even if venturing into such large inputs and expenditures was dared for a first system, it must nevertheless be borne in mind that this test would have to be carried out again with at least one third of the initial expenditure after each system modification, which is obviously not economically feasible. Even with a much better system with expected value factor $\alpha_{50\,\%} \leq 0.23$, at least one third of the expenditure still remains (see Sect. 4.1.1).

These arguments lead to the conclusion that no economically feasible development or approval of autonomous vehicles is possible using the known testing procedures for measuring risk over a continuous distance. This aspect certainly has the potential for being a "showstopper" and is also described as the "approval trap of autonomous driving."

One special irony of this dilemma is that the proof distance becomes especially long when automating situations are apparently simple. This is because only a few accidents happen per distance in such situations. Conversely, the recommendation derived for testability reasons is to start with areas of use that are particularly susceptible to accidents. However, this susceptibility to accidents will become clear even before the validation phase, namely, in the development phase. With each improvement stage, an ever greater distance is needed to raise safety. As a result, statements referring to 100,000 km or 1 million km without accidents involving personal injury are very impressive from a technical standpoint but of little relevance compared to the safety goals which are two to three orders of magnitude higher.

4.2 Way Out of the Test Dilemma

This test dilemma can only be overcome by achieving a drastic reduction in the distance required. In component durability tests, it is common on the one hand to select those parts of the operating load spectrum which place stress on the component in a relevant manner and to considerably shorten the tests by omitting the irrelevant portions. On the other hand, acceleration methods are resorted to, i.e., higher loads or environmental conditions exerting greater stresses are used to load the component. However, it appears difficult to adapt these strategies to the proof of safety for autonomous driving, as the failure mechanisms are not based on failure of the function, but rather on wrong decisions that lead to accidents. Of course, a system simulation, whether as software-in-the-loop (SIL) or as hardware-in-the-loop (HIL), is conceivable to validate the function and essential for development. It will be hardly possible to incorporate a diversity of possible variations in road traffic that correspond to a driving distance of several million kilometers and that are representative for all user groups. Nevertheless, this latter consideration regarding the test dilemma is the one that points to a way out: Even if all relevant states can be included in a test program, in certain situations, it would not be possible to decide which system response is right or wrong because this question cannot be answered by the ego system alone. In particular, wherever the actions and reactions of other road users are to be anticipated, it is impossible to arrive at an assumption that is 100 % correct because the reaction models of individual road users do not provide us with any individual and instantaneous correctness. The system responses thus assume a probabilistic character, and assessment of the extent to which the current decision may be correct is time-dependent and will probably only be determinable in simple situations. All other actions and reactions possible in this particular situation will not be repeated in this way anywhere or ever again, and even the conclusion as to whether the reaction was correct cannot be inferred from the situation's result. Even if an accident occurs following a reaction, the reaction may still have been correct in terms of minimizing damage. It is equally possible that a wrong decision does not as such have a negative outcome and does not lead to an accident because the environmental constellations are favorable. However, this raises the question of what can be used to counter the previous thinking of "right or wrong." The answer is as simple in principle as it is difficult to implement: The driving robot has the task of executing the driving task more safely than the human control group, who, for example, may be experienced, high mileage drivers in the peak of health. For this task, the driving robot's overall performance, consisting of perception, cognition, and action performance, must be at least as high as that of the control group. If it is possible to measure this performance, then the driving robot can be approved; this statement can also be transferred to other fields of robotics, such as humanoid household robots.

So far, there is no known general metric for expressing the perception, cognition, and action performance of robots and humans: An example can be found, however, in the games of chess and the game Go, an area in which the computer has achieved the performance of humans and in some cases even exceeded it. The game of chess is not fundamentally probabilistic, but it is not predictable within a finite time due to the sheer number of possible move combinations. As a result, the chess computer has to make decisions according to heuristic algorithms, it being impossible to evaluate these decisions as right or wrong at the time of the decision. However, if the computer decides something "more correctly," it can be expected that it will win more games than a human player. This expected playing strength of a Go or a chess player is expressed using his/her Elo rating (officially FIDE rating): It describes the expected scores of a match and is part of an objective evaluation system developed by Elo (1978). It should be mentioned as a limitation of this example that, although both computer and human players receive an Elo rating, in each case, these ratings are determined only from games between identical categories (human vs. human or computer vs. computer). Nevertheless, it can be said that for a small area, two of the requirements which are placed on a metric for the approval of robot functions are met: On the one hand, it is possible with the Elo scale to compare (theoretically at least) human performance with that of the robot, and on the other hand, an absolute classification can be made using this metric since it is possible with the Elo rating, for example, to assign whether someone is an amateur or a grand master. If there was also such a metric for driving robots, it would be possible to specify a defined skill class in accordance with ISO 26262 for certain levels of automation.

However, this approach from the field of chess cannot be transferred directly to the driving robot because the Elo value is determined via direct comparison, that is to say via the win/loss record of opponents of a given strength. Furthermore, only the cognitive performance is measured; the performance of perception is carried out in an idealized manner because the position of the chess pieces is transmitted correctly and completely to the chess computer, while in road traffic, all this information will not be available in this manner to either the driver or the driving robot. Moreover, it would not be possible in practice to filter and process such a wealth of information, if it were available. For a technical system such as a driving robot, therefore, the overall task would have to be broken down into three domains with a separate metric assigned to each.

As the chapters on actuator systems illustrate, the execution capabilities possible with machines are already very close to human capabilities: In some areas, they already go beyond them, such as individual wheel brake control or rear axle steering. Machine perception has already achieved a remarkable performance, although the perception of very complex situations, e.g., traffic around the Arc de Triomphe in Paris, has not yet succeeded. Drivers who are not native to Paris, however, may likewise feel overwhelmed in this situation and at the limit of their performance. Nevertheless, the relatively small number of accidents that happen – apart from harmless car body damage – shows that humans are also able to handle such situations. Machine-based cognitive performance, particularly with regard to decision flexibility, is still low at present. Above all, it still seems to be very difficult to simulate the learning process of humans. All motorists experience this learning process after their driving instruction, and without this broadening of driving skills, we would certainly be exposed to a higher traffic risk. A breakdown into the three domains could be used advantageously to isolate the assessment: A change in the sensor range can be certified on the perception metric alone, without compulsorily having to certify the others at the same time. For the same reason, there is corresponding modularization in the development of autonomous vehicles (see ▶ Chap. 62, "Autonomous Driving" or Langer et al. (2008) and Darms et al. (2008)).

Returning to the previous considerations that in order to have a chance of approval, only those driving robots that replace humans in vehicle guidance need to be superior to them in terms of safety, we can draw two conclusions: The driving robots still have a lot of development ahead of them but, assuming a recognized metric for driving performance, they can become superior to humans. This metric, which can definitely be very specific for certain areas of use, is an indispensable prerequisite for targeted development of autonomic functions and, in the author's view, its development represents the critical path of the autonomous vehicle's development:

> As long as no metric exists in a generally accepted form, wider use of autonomous vehicles will not be achieved on public roads.

4.3 Possible Route to a Metric

The requirements for such a metric are:

The metric is valid for the relevant application area.

Basically, this requirement cannot be achieved because the abilities required only become completely clear when the metric is used. However, this also applies equally to today's developments. Here, transfers from similar areas are used as a help, but at the same time, this approach makes it necessary to introduce many intermediate stages en route to autonomous driving. Only when sufficient experience with similar systems is available can the metric be calibrated and transferred to the next expansion with justifiable residual risk: It is the validation strategy, therefore, that determines the migration and introduction strategy and not the development of technically possible functions.

The metric allows a comparison between the driving abilities of humans and robots.

This is perhaps the most difficult requirement to implement because it assumes that human abilities are measured and weighted in a manner appropriate to the driving task. Although a breakdown into the three domains is indeed carried out in human engineering models, perception performance cannot be separated from cognitive performance. This is possible, however, for the execution performance even if cross-coupling may occur due to retroactive effects. For these reasons, there is no choice but to compare the combined performance of perception and cognition for human and machine, at least until the metrics have been established. Once the relevant levels for a classification have been established, the breakdown of perception and cognition performance in machines can be considered separately.

The metric allows for clear class levels.

This requirement is needed for certification so that a classification can be made analogously to the Automotive Safety Integrity Levels of ISO 26262. To this end, we must work out appropriate limit values and weightings of individual features.

The metric uses economically feasible test methods for classification.

Prohibitive costs in particular, as already explained above, were the reason for the departure from the established approval methodology. The new procedure must therefore be significantly cheaper. Real and virtual test tracks with a high level of difficulty may offer a way out here, although the difficulties must be representative for the area of use.

The metric itself must not favor any pattern of action but must determine precisely the ability to act appropriately in unfamiliar conditions.

This means that no training may be done on the test pattern because it would lead to a reduction in the flexibility of action. Such training must definitely be prevented as it is this flexibility that actually allows the extrapolation of a test track to the whole area of use.

All these requirements are very demanding. However, since in the author's view introducing autonomous vehicles into use on public roads can only be possible with such a metric, its development will determine the timing and strategy of introduction. The preparatory work still be carried out certainly has the order of magnitude of the genome project and will claim hundreds of person-years in research. It appears that a realignment of computer intelligence research will be necessary because current research activities still give too little priority to this issue of validation.

5 Evolution to Autonomous Driving

Deviating from the presentations in the first two editions of the German handbook, in which an evolutionary roadmap was presented with temporal and functional dependencies, an approach is introduced here which emanates from three evolutionary starting points. Autonomous driving can be depicted as a color triangle, as a composition of three basic forms:

Simple scenarios:

The starting point for this direction is formed by the two systems adaptive cruise control (ACC; see ▶ Chap. 45, "Adaptive Cruise Control") and lane-keeping assist system (LKAS; see ▶ Chap. 48, "Lateral Guidance Assistance"), whose function concentrates on free travel and traveling in a moving line within a lane at higher speeds. Limiting the level of intervention with regard to acceleration (ACC) and steering wheel torque (LKAS) enables automation within the comfort zone but not for situations with higher intervention dynamics; this also applies to the combined longitudinal and lateral guidance systems. The basic design of the systems is based on the driver's ability to take over.

Low speed:

Based on parking assist by means of active lateral guidance which has already been available for some time, this strand will develop via full automation of the entire parking process through to automatic valet parking. The great advantage of automating driving at low speeds lies in the simple fail-safe strategy. It is possible to decelerate to a stop within a short distance. As soon as there is no guarantee that the driving space is free over this distance, not only is it possible to create a safe situation by way of such a stop but it is also possible to hand over to a driver or, remotely, to an operator who can then take over responsibility for continuing to drive.

High-risk situations:

As set out in ▶ Chaps. 46, "Fundamentals of Collision Protection Systems," and ▶ 47, "Development Process of Forward Collision Prevention Systems," automatic interventions can significantly reduce the risk of accidents if the driving situation is classified as so critical that the risk without intervention appears to be higher than that with intervention. The first systems of this evolutionary strand were the emergency braking systems which mitigated the consequences of collision. In subsequent systems, the functionality was also extended to collision avoidance braking and approaches for emergency evasion are also foreseeable in defined situations. An Emergency Stop Assistant, upon detecting the driver's incapacity to drive due to health problems, continues this orientation but requires a much larger repertoire of responses.

Combinations:

Driver assistance (packages) already combine functionalities from the starting directions mentioned. Examples include full-speed range ACC (see ▶ Chap. 45, "Adaptive Cruise Control") and the traffic jam assistant (see ▶ Chap. 51, "Traffic Jam Assistance and Automation"). The lane-correcting braking intervention, which only prevents the vehicle from leaving the lane when it detects oncoming traffic (see ▶ Chap. 48, "Lateral Guidance Assistance"), may also be regarded as a combination. A functionality called Safety Corridor in the project PRORETA 3 (Cieler et al. 2014) continues this by permanently monitoring the safety margin and by recovering the safety margin in an informative and, if necessary, intervening manner when approaching the margin's limits. Concepts of maneuver-based cooperative automation are considered to be cooperative guidance (see ▶ Chap. 59, "Cooperative Guidance, Control, and Automation"), specifically conduct-by-wire (see ▶ Chap. 60, "Conduct-by-Wire"). Due to the differently designed arbitration of the H-mode concept (see ▶ Chap. 61, "H-Mode"), the boundaries between a safety corridor function and an automated longitudinal and lateral guidance are fluid, but without getting closer to full automation, only the driver's involvement is designed to be more extensive.

The combination of risk assessment and low-speed automation could lead to an extended application of a city shuttle (e.g., INDUCT 2014), as already being tested in pilot tests in Singapore and Stanford (Beiker 2015). The advantage here is that such vehicles do not automate existing driver-vehicle units, but rather that new mobility services are created which are designed to be driverless from the outset. They do not have to compete with conventionally driven cars, either in terms of driving performance or even in terms of driving safety, as they serve a different driving profile for which no reference values exist and, thanks to the low speed, it is generally possible to assume a lower latent baseline risk. Technical development can then gradually extend the traveling speed upward and also enlarge the areas of use.

Synthesis:

Each of the directions previously mentioned provides marketable basic applications and thus the basis for technological advancement and increasing functional maturity. However, with regard to autonomous driving (irrespective of whether or not a driver would be available), all applications come up against conceptual boundaries that cannot be crossed without “importing” from the other corners. Nevertheless, many issues of technology, legal aspects, and user acceptance still remain open, which also relegate such an evolutionary consideration to the level of gazing into a crystal ball, although with a trained eye (Fig. 5).

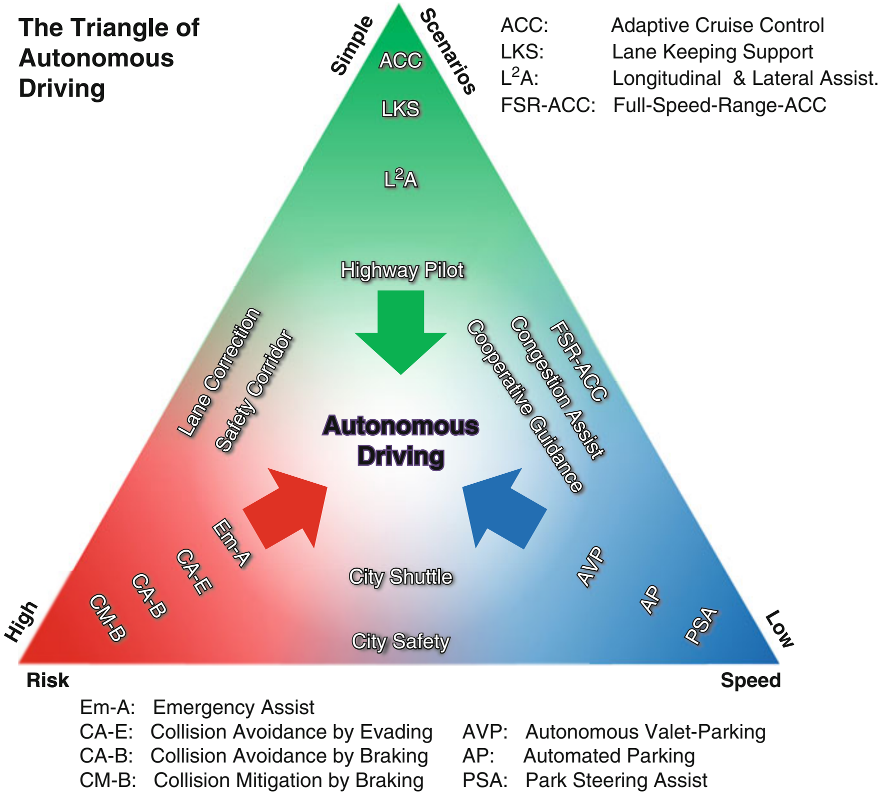


Fig. 5 Evolution to autonomous driving, beginning from the three-corner starting points to the center of autonomous driving

6 Future Research Priorities

Following our overview of future impacts, detailed illumination of validation issues, and presentation of the evolutionary triangle for autonomous driving, we conclude by formulating future research priorities. These focal areas are intended as a recommendation for action and have also been taken from the Uni-DAS position paper (Bengler et al. 2012).

6.1 Individualization

At present, driver assistance is not directed at all road user groups, nor is it geared to individual needs or preferences. For existing functionalities, this deficiency can probably be resolved in most cases by way of suitable human-machine interaction concepts and can be made available to a larger user group. However, appropriate functional designs and also functions are still lacking that specifically address the needs of individual user groups. This need becomes apparent in regard to elderly drivers who need to preserve their individual mobility for as long as possible, but is equally applicable to young drivers who are disproportionally frequently involved in accidents, motorbike drivers as users of a vehicle class inherently different from four-wheelers, and commercial bus and truck conductors with their unusually high driving frequencies and heavy vehicles. Regarding the latter, traffic safety should be reevaluated after the introduction of the new emergency braking and lane-keeping systems, in order to determine what additional support is required.

Overall, functional development should be directed more toward this aspect of use geared to the individual. With future assistance functions, particularly with a higher level of automation, attention must be paid to optimal human-machine interaction. New interface technologies for display and operation should allow the driver's "immersion" in an integrated driver-vehicle system which – in the figurative sense – is controlled by the driver's intention. However, this concerns not only the interface elements but also the overall functionality for intuitive task distribution without the danger of confusion, even when a higher level of automation can be used. All intermediate stages for the fully automated, then driverless vehicle require interaction concepts, both for assigning and taking back the driving function and also for explicit or implicit control. Since the advances can only be achieved via commercial success, these systems require a high hedonic quality, a challenge not only for the "look and feel" of the system design. At the latest when reconsidering this point, the location-specific character of the assistance functions will emerge as a new challenge. Locally adapted or even locally different concepts will be necessary both for market success and also for further increased traffic safety. This work must be tackled internationally, to satisfy the different societies, economic areas, and legal systems.

Future priorities:

- Assistance functions addressing specific requirements of specific user groups, especially young or elderly drivers as well as motorcyclists
- Analysis of traffic accidents after introduction of emergency braking and lane-keeping support systems
- New human-machine interfaces to support immersion into an integrated driver-vehicle system
- Driver intention recognition
- Concepts for cooperation between driver and vehicle in automated mode
- Raise hedonic quality in regard to acceptance and market success
- International approach to DAS to achieve acceptance in different countries and cultures

6.2 Machine Perception and Cognition

Today's sensors are capable of collecting detailed data of a car's surrounding environment, but machine cognition and situational awareness are still in their infancy. To improve them, significant progress is required in symbolic scene classification, e.g., object recognition under dynamic conditions, as well as in contextual scene understanding, e.g., inference of the relationship between different dynamic objects and with traffic infrastructure elements. Last but not least, the uncertainty and vagueness of the information from and interpretation of the traffic scene needs to be made explicit. Managing the above is crucial for the realization of appropriate driving functions and corrective actions in complex traffic situations. The acquisition of information should be based on more sources than are available in today's cars. High-precision ego-localization in rich 3D digital maps will play a special role (Nothdurft et al. 2011). New hardware concepts and algorithms for sensor data acquisition and interpretation could pave the way for performance improvements at reduced costs. Machine vision techniques for image sequence analysis, as well as microwave and active optical sensor technologies, still exhibit large potentials to enhance spatiotemporal resolution and situational awareness of the perceived traffic scene. Methods for scene representation, including measures for quality, need to be further elaborated as a basis for situational awareness.

Another future focus will be on a probabilistic prediction of likely future behavior of the ego-vehicle and other traffic participants, based on comprehensive intention inference and behavior modeling (Liebner et al. 2013).

Future priorities:

- Improved algorithms for vehicle situational awareness in complex traffic scenarios, especially in urban environment
- Improvement of sensor hardware and software to yield richer high-quality information

- Development of methods and algorithms to acquire situational awareness at a safety-relevant integrity level
- Automated generation, updating, and distribution of local dynamic maps
- Intention and behavior models to predict the behavior of the driver and other traffic participants

6.3 Methods of Assessment

In the past, driving assistance research focused on technological breakthroughs. The emphasis is now shifting, as methods of assessment (e.g., Fecher et al. 2008; Schöner et al. 2011; Aparicio et al. 2012; Brahmi et al. 2013) become increasingly important. Without suitable and generally accepted methods of assessment, potentially distracting or unsafe functions cannot be introduced to the market.

Conventional testing procedures are insufficient to ensure the safety of increasingly complex future assistance functions involving machine perception and cognition. For this reason, only apparently "harmless" assistance functions, like ACC or systems with short intervention periods like emergency braking assistance, are currently available. However, the number of DAS and their functional range are expected to grow considerably in the near term. If testing and assessment methods cannot keep pace with this functional growth, they will become the bottleneck of the introduction of advanced DAS to the market (Maurer and Winner 2013). There is a lack of concepts that allow economically viable implementation, starting with the test for machine perception and the test for desired as well as faulty functional behavior through to acceptance assessment. A particular challenge is posed by functions that delegate decisions from the human operator to the machine in unexpected scenarios. In these cases, market introduction requires prior proof that the risk taken by handing vehicle control over to the machine is at most equal to the risk taken when the human driver is in control (Färber and Maurer 2005; Bock et al. 2007). Two yet unresolved issues arise: how to measure performance of the machine and that of the human operator (Damböck et al. 2012). Valid assessment methodologies exist for neither, not even to mention the case of shared or cooperative control by the human operator and the machine (Bengler et al. 2012).

Since a solution to these challenges is not to be expected in the near term, research on suitable assessment methods constitutes a part of the critical path to be taken in order to avoid DAS development being held up for decades.

Future priorities:

- Testing and evaluation methods for machine cognition and (semi-)automated assistance functions
- Concepts for the assessment of human and machine driving performance

6.4 Interconnection

Unlike the first three priorities, the fourth area does not focus on assistance in the vehicle but rather on integration in the overall transport network. Existing communication networks and in particular near-future vehicle2x networks open up a wide spectrum of improvements to the holistic performance of transportation systems. Hence, existing approaches should be further developed toward a level of maturity that allows market introduction and increases the safety of all traffic participants through the benefits of shared information.

New concepts in the future should soon follow this which, assuming a high penetration rate of driver assistance systems, will enable an optimal transport system with minimal use of resources and the highest possible level of safety. These concepts should furthermore not focus merely on individual rides, but also provide interfaces to currently inactive or intermodal traffic in order to interlink alternative transportation systems. When DAS are modified to promote cooperative traffic, appealing visions such as “deterministic traffic” can come true. This would mean that a trip is carried out according to an interactive schedule with traffic participants moving in imaginary spatiotemporal slots.

Future priorities:

- Integration of vehicle2X networks for the sake of traffic safety and efficiency
- Collective provision of accurate local traffic information
- Collective traffic control based on individually operated cooperative systems
- Continual joint mission planning with reliable prediction of the individual vehicle trajectories
- Usage optimization of deterministic traffic system concepts

6.5 Social Priorities of Research

The foci addressed so far concentrated on research areas from a technological point of view. They are based on the expertise of the authors of this technical report. However, not all relevant topics have been exhaustively addressed. DAS not only reflect the progress of technology, but are developed for humans who purchase and use them; they make a difference for individual and collective safety as well as for the mobility of groups and individuals.

Development of advanced DAS may be stimulated by market acceptance or stunted by societal reservations (Krüger 2008; Karmasin 2008). On the other hand, ADAS – particularly those with a high degree of vehicle automation – may induce changes in traffic behavior (Freyer et al. 2007; 2008). Furthermore, they may stimulate new business models and have a drastic impact on the nature of future mobility. An early proactive assessment of the consequences of technology may reduce potential conflicts. Here, political and social discussions can begin prior to market introduction, thus reducing the risk of investments loss (Homann 2005). Also in the future, we will see a need for interdisciplinary research with regard to social aspects, such as the social impact and paving the way for highly automated or autonomous driving.

Acknowledgments We gratefully acknowledge colleagues from Uni-DAS e. V. for the opportunity to take over Sects. 2, 3, and 6 from the Uni-DAS position paper.

We also gratefully acknowledge the coauthors of the first edition, Dr. Gabriele Wolf, and the second edition, Dr. Alexander Weitzel, for their previous contributions which have strongly influenced this chapter so far.

References

Aparicio A, Bargalló J, Baurès S, Rodarius C, Vissers J, Bartels O, Seiniger P, Lemmen P, Unselt T, Ranovona M, Okawa T, Schaub S (2012) Pre-crash performance of collision mitigation and avoidance systems – results from the ASSESS project, Paper Nr. AVEC-043. In: Proceedings of the 11th AVEC Conference, Seoul, Sept 2012

Beiker S (2015) Implementation of a self-driving transportation system. In: Maurer M, Gerdes C, Lenz B, Winner H (eds) Autonomes Fahren. Springer, Berlin

Bengler K, Dietmayer K, Färber B, Maurer M, Stiller C, Winner H (2012) Die Zukunft der Fahrerassistenz – Ein Strategiepapier der Uni-DAS (The future of driver assistance – A strategy paper of Uni-DAS), Uni-DAS e. V. (Publ.), ISBN: 978-3-941543-09-6, Internet: http://www.uni-das.de/documents/Strategiepapier_Uni-DAS.pdf

Bläser D, Arch M, Schmidt A (2012) Mobilität findet Stadt. Zukunft der Mobilität für urbane Metropolräume. (Mobility meets City). In: Proff H, Schönharting J, Schramm D, Ziegler J (eds) Zukünftige Entwicklungen in der Mobilität. Betriebswirtschaftliche und technische Aspekte. Springer Gabler, Wiesbaden, pp 501–515

Blumenthal F (2013) Vernetzt und autonom (Connected and autonomous). ATZ Agenda Vol. 2, 20–23

Bock T, Maurer M, Färber B (2007) Vehicle in the Loop (VIL) – a new simulator set-up for testing advanced driving assistance systems. In: Driving simulation conference North America 2007, University of Iowa

Brahmi M, Siedersberger KH, Siegel A, Maurer M (2013) Referenzsysteme für die Bewertung der Umfelderfassung: Anforderungen, Validierung und Einsatz (Reference systems for the assessment of surroundings perception: Requirements, validation and usage). In: 6. Tagung Fahrerassistenz, Lehrstuhl für Fahrzeugtechnik (FTM), Technische Universität München, München

Cieler S, Konigorski U, Lüke S, Winner H (2014) Umfassende Fahrerassistenz durch Sicherheitskorridor und kooperative Automation (Comprehensive driver assistance by safety corridor and cooperative automation), ATZ Ausgabe 10-2014

Damböck D, Farid M, Tönert L, Bengler K (2012) Übernahmezeiten beim hochautomatisierten Fahren (Takeover times at highly automated driving) 5. Tagung Fahrerassistenz, Lehrstuhl für Fahrzeugtechnik (FTM), Technische Universität München, München

Darms M, Baker C, Rybksi P, Urmson C (2008) Vehicle Detection and Tracking for the Urban Challenge. In: 5. Workshop Fahrerassistenzsysteme Walting, Freundeskreis Mess- und Regelungstechnik, Karlsruhe

Dietl M, Gutmann JS, Nebel B (2001) Cooperative sensing in dynamic environments. In: Proceedings IEEE/RSJ international conference on Intelligent Robots and Systems (IROS), Maui, pp 1706–1713, Nov 2001

Eichler S (2007) Extending onboard sensor information by wireless communication. In: Proceedings IEEE intelligent vehicles symposium, Istanbul, pp 541–546, June 2007

Elo A (1978) The rating of chess players, past and present. Arco Pub, New York

European Union (2010) E-Safety, Nov 2010. [Online] Available: http://www.ec.europa.eu/informationsociety/activities/esafety. Accessed June 2013

Färber B, Maurer M (2005) Nutzer- und Nutzen-Parameter von Collision Warning und Collision Mitigation Systemen. (Benefit and usage parameters of collision warning and collision mitigation systems) In: M. Maurer, C. Stiller (eds) 5. Workshop Fahrerassistenzsysteme Walting, Freundeskreis Mess- und Regelungstechnik, Karlsruhe

Fecher N, Fuchs K, Hoffmann J, Abendroth B, Bruder R, Winner H (2008) Analysis of the driver behavior in autonomous emergency hazard braking situations. In: FISITA World Automotive Congress, Munich, 14–19 Sept 2008

Freyer J (2008) Vernetzung von Fahrerassistenzsystemen zur Verbesserung des Spurwechselverhaltens von ACC (Connecting of driver assistance systems for improving the lane change behavior of ACC), Cuvillier

Freyer J, Deml B, Maurer M, Färber B (2007) ACC with enhanced situation awareness to reduce behavior adaptations in lane change situations. In: Intelligent vehicles symposium – IV2007, Istanbul

Homann K (2005) Wirtschaft und gesellschaftliche Akzeptanz: Fahrerassistenzsysteme auf dem Prüfstand (Economics and societal acceptance: driver assistance to reflect). In: Maurer M, Stiller C (eds) Fahrerassistenzsysteme mit maschineller Wahrnehmung. Springer, Berlin, pp 239–244

INDUCT (2014) Navia – A self-driving shuttle at the service of urban mobility. http://induct-technology.com/en/products/navia. Accessed Juli 2014

Karmasin H (2008) Motivation zum Kauf von Fahrerassistenzsystemen (Motivation for buying driver assistance systems). In: 24. VDI/VW-Gemeinschaftstagung Integrierte Sicherheit und Fahrerassistenzsysteme. VDI-Berichte 2048, Wolfsburg

Ko-FAS – Kooperative Sensorik und kooperative Perzeption für die präventive Sicherheit im Straßenverkehr (Ko-FAS – Cooperative sensors and cooperative perception for the preventive safety in road traffic). http://www.ko-fas.de. Accessed on July 2015

Kosch T (2004) Efficient information dissemination in vehicle ad-hoc networks. In: Proceedings of the 11th World Congress on Intelligent Transportation Systems (ITS), Oct 2004

Krüger HP (2008) Hedonomie – die emotionale Dimension der Fahrerassistenz (Hedonomics – the emotional dimension of driver assistance). In 3. Tagung Aktive Sicherheit durch Fahrerassistenz, Lehrstuhl für Fahrzeugtechnik (FTM), Technische Universität München, München

Langer D, Switkes J, Stoschek A, Hunhnke B (2008) Environment Perception in the 2007 Urban Challenge: Utility for Future Driver Assistance Systems. In: 5. Workshop Fahrerassistenzsysteme Walting, Freundeskreis Mess- und Regelungstechnik, Karlsruhe

Lerner M, Schepers A, Pöppel-Decker M, Leipnitz C, Fitschen A (2013) Voraussichtliche Entwicklung von Unfallanzahlen und Jahresfahrleistungen in Deutschland – Ergebnisse 2013 (Prospective trend of accident figures and annually driving mileage in Germany). Bundesanstalt für Straßenwesen. http://www.bast.de/DE/Presse/Downloads/2013-32-langfassung-pressemitteilung.pdf?__blob=publicationFile&v=2. Dec 2013

Liebner M, Klanner F, Baumann M, Ruhhammer C, Stiller C (2013) Velocity-based driver intent inference at urban intersections in the presence of preceding vehicles. IEEE Intell Transp Syst Mag 5(2):10–21

Maurer M, Winner H (2013) Automotive systems engineering. Springer, Berlin

Nagel R, Eichler S, Eberspächer J (2007) Intelligent wireless communication for future autonomous and cognitive automobiles. In: Proceedings IEEE intelligent vehicles symposium, Istanbul, pp 716–721, June 2007

Nothdurft T, Hecker P, Ohl S, Saust F, Maurer M, Reschka A, Böhmer JR (2011) Stadtpilot: first fully autonomous test drives in Urban Traffic. In: 14th international IEEE annual conference on intelligent transportation systems, Washington DC

Saust F, Wille J, Maurer M (2012) Energy-optimized driving with an autonomous vehicle in urban environments. In: IEEE Vehicular Technology Conference (VTC Spring), Yokohama

Schöner HP, Hurich W, Haaf D (2011) Selbstfahrendes Soft Crash Target zur Erprobung von Fahrerassistenzsystemen (Self-driving soft crash target for testing of driver assistance systems). In: Proceedings Konferenz AAET 2011, Braunschweig

simTD-Konsortium: Sichere Intelligente Mobilität – Testfeld Deutschland (Safe Intelligent Mobility – Test Field Germany). Projekt-Homepage http://www.simtd.de

Statistisches Bundesamt (2013): Verkehrsunfälle Zeitreihen 2013 (Traffic accidents time series 2013). Access https://www.destatis.de/DE/Publikationen/Thematisch/TransportVerkehr/Verkehrsunfaelle/VerkehrsunfaelleZeitreihen.html

Tank T, Linnartz M (1997) Vehicle-to-vehicle communications for avcs platooning. IEEE Trans Veh Technol 46:528–536

Terporten M, Bialdyga D, Planing P (2012) Veränderte Kundenwünsche als Chance zur Differenzierung. Herausforderungen für das Marketing am Beispiel neuer Mobilitätskonzepte (Changing requests as chance for differentiation. Requirements for the marketing in case of new mobility concepts). In: Proff H, Schönharting J, Schramm D, Ziegler J (eds) Zukünftige Entwicklungen in der Mobilität. Betriebswirtschaftliche und technische Aspekte. Springer Gabler, Wiesbaden, pp 367–382

Tischler K, Hummel B (2005) Enhanced environmental perception by inter-vehicle data exchange. In: IEEE intelligent vehicles symposium, Las Vegas, June 2005

Uber Technologies Inc (2015) https://www.uber.com

Index

H. Winner et al. (eds.), *Handbook of Driver Assistance Systems*,
DOI 10.1007/978-3-319-12352-3