Don Harris (Ed.)

LNAI 8020

Engineering Psychology and Cognitive Ergonomics

Applications and Services

10th International Conference, EPCE 2013
Held as Part of HCI International 2013
Las Vegas, NV, USA, July 2013, Proceedings, Part II

2 Part II

HCI2013
INTERNATIONAL

Springer

Lecture Notes in Artificial Intelligence 8020

Subseries of Lecture Notes in Computer Science

LNAI Series Editors

Randy Goebel
University of Alberta, Edmonton, Canada
Yuzuru Tanaka
Hokkaido University, Sapporo, Japan
Wolfgang Wahlster
DFKI and Saarland University, Saarbrücken, Germany

LNAI Founding Series Editor

Joerg Siekmann
DFKI and Saarland University, Saarbrücken, Germany

Don Harris (Ed.)

Engineering Psychology and Cognitive Ergonomics

Applications and Services

10th International Conference, EPCE 2013
Held as Part of HCI International 2013
Las Vegas, NV, USA, July 21-26, 2013
Proceedings, Part II

Springer

Volume Editor

Don Harris
Coventry University
Faculty of Engineering and Computing
Priory Street
Coventry, CV1 5FB, UK
E-mail: don.harris@coventry.ac.uk

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-39353-2 e-ISBN 978-3-642-39354-9
DOI 10.1007/978-3-642-39354-9
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2013941401

CR Subject Classification (1998): H.5, H.1, K.4, J.2, J.3, J.7, H.4, I.6

LNCS Sublibrary: SL 7 – Artificial Intelligence

© Springer-Verlag Berlin Heidelberg 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The 15th International Conference on Human–Computer Interaction, HCI International 2013, was held in Las Vegas, Nevada, USA, 21–26 July 2013, incorporating 12 conferences / thematic areas:

Thematic areas:

- Human–Computer Interaction
- Human Interface and the Management of Information

Affiliated conferences:

- 10th International Conference on Engineering Psychology and Cognitive Ergonomics
- 7th International Conference on Universal Access in Human–Computer Interaction
- 5th International Conference on Virtual, Augmented and Mixed Reality
- 5th International Conference on Cross-Cultural Design
- 5th International Conference on Online Communities and Social Computing
- 7th International Conference on Augmented Cognition
- 4th International Conference on Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management
- 2nd International Conference on Design, User Experience and Usability
- 1st International Conference on Distributed, Ambient and Pervasive Interactions
- 1st International Conference on Human Aspects of Information Security, Privacy and Trust

A total of 5210 individuals from academia, research institutes, industry and governmental agencies from 70 countries submitted contributions, and 1666 papers and 303 posters were included in the program. These papers address the latest research and development efforts and highlight the human aspects of design and use of computing systems. The papers accepted for presentation thoroughly cover the entire field of Human–Computer Interaction, addressing major advances in knowledge and effective use of computers in a variety of application areas.

This volume, edited by Masaaki Kurosu, contains papers focusing on the thematic area of Human–Computer Interaction, and addressing the following major topics:

This volume, edited by Don Harris, contains papers focusing on the thematic area of Engineering Psychology and Cognitive Ergonomics, and addressing the following major topics:

- Driving and Transportation Safety
- Cognitive Issues in Aviation
- Military Applications
- Cognitive Issues in Health and Well-Being

The remaining volumes of the HCI International 2013 proceedings are:

- Volume 2, LNCS 8005, Human–Computer Interaction: Applications and Services (Part II), edited by Masaaki Kurosu
- Volume 3, LNCS 8006, Human–Computer Interaction: Users and Contexts of Use (Part III), edited by Masaaki Kurosu
- Volume 4, LNCS 8007, Human–Computer Interaction: Interaction Modalities and Techniques (Part IV), edited by Masaaki Kurosu
- Volume 5, LNCS 8008, Human–Computer Interaction: Towards Intelligent and Implicit Interaction (Part V), edited by Masaaki Kurosu
- Volume 6, LNCS 8009, Universal Access in Human–Computer Interaction: Design Methods, Tools and Interaction Techniques for eInclusion (Part I), edited by Constantine Stephanidis and Margherita Antona
- Volume 7, LNCS 8010, Universal Access in Human–Computer Interaction: User and Context Diversity (Part II), edited by Constantine Stephanidis and Margherita Antona
- Volume 8, LNCS 8011, Universal Access in Human–Computer Interaction: Applications and Services for Quality of Life (Part III), edited by Constantine Stephanidis and Margherita Antona
- Volume 9, LNCS 8012, Design, User Experience, and Usability: Design Philosophy, Methods and Tools (Part I), edited by Aaron Marcus
- Volume 10, LNCS 8013, Design, User Experience, and Usability: Health, Learning, Playing, Cultural, and Cross-Cultural User Experience (Part II), edited by Aaron Marcus
- Volume 11, LNCS 8014, Design, User Experience, and Usability: User Experience in Novel Technological Environments (Part III), edited by Aaron Marcus
- Volume 12, LNCS 8015, Design, User Experience, and Usability: Web, Mobile and Product Design (Part IV), edited by Aaron Marcus
- Volume 13, LNCS 8016, Human Interface and the Management of Information: Information and Interaction Design (Part I), edited by Sakae Yamamoto
- Volume 14, LNCS 8017, Human Interface and the Management of Information: Information and Interaction for Health, Safety, Mobility and Complex Environments (Part II), edited by Sakae Yamamoto
- Volume 15, LNCS 8018, Human Interface and the Management of Information: Information and Interaction for Learning, Culture, Collaboration and Business (Part III), edited by Sakae Yamamoto
- Volume 16, LNAI 8019, Engineering Psychology and Cognitive Ergonomics: Understanding Human Cognition (Part I), edited by Don Harris
- Volume 18, LNCS 8021, Virtual, Augmented and Mixed Reality: Designing and Developing Augmented and Virtual Environments (Part I), edited by Randall Shumaker
- Volume 19, LNCS 8022, Virtual, Augmented and Mixed Reality: Systems and Applications (Part II), edited by Randall Shumaker
- Volume 20, LNCS 8023, Cross-Cultural Design: Methods, Practice and Case Studies (Part I), edited by P.L. Patrick Rau

- Volume 21, LNCS 8024, Cross-Cultural Design: Cultural Differences in Everyday Life (Part II), edited by P.L. Patrick Rau
- Volume 22, LNCS 8025, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Healthcare and Safety of the Environment and Transport (Part I), edited by Vincent G. Duffy
- Volume 23, LNCS 8026, Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: Human Body Modeling and Ergonomics (Part II), edited by Vincent G. Duffy
- Volume 24, LNAI 8027, Foundations of Augmented Cognition, edited by Dylan D. Schmorrow and Cali M. Fidopiastis
- Volume 25, LNCS 8028, Distributed, Ambient and Pervasive Interactions, edited by Norbert Streitz and Constantine Stephanidis
- Volume 26, LNCS 8029, Online Communities and Social Computing, edited by A. Ant Ozok and Panayiotis Zaphiris
- Volume 27, LNCS 8030, Human Aspects of Information Security, Privacy and Trust, edited by Louis Marinos and Ioannis Askoxylakis
- Volume 28, CCIS 373, HCI International 2013 Posters Proceedings (Part I), edited by Constantine Stephanidis
- Volume 29, CCIS 374, HCI International 2013 Posters Proceedings (Part II), edited by Constantine Stephanidis

I would like to thank the Program Chairs and the members of the Program Boards of all affiliated conferences and thematic areas, listed below, for their contribution to the highest scientific quality and the overall success of the HCI International 2013 conference.

This conference could not have been possible without the continuous support and advice of the Founding Chair and Conference Scientific Advisor, Prof. Gavriel Salvendy, as well as the dedicated work and outstanding efforts of the Communications Chair and Editor of HCI International News, Abbas Moallem.

I would also like to thank for their contribution towards the smooth organization of the HCI International 2013 Conference the members of the Human–Computer Interaction Laboratory of ICS-FORTH, and in particular George Paparoulis, Maria Pitsoulaki, Stavroula Ntoa, Maria Bouhli and George Kapnas.

May 2013

Constantine Stephanidis
General Chair, HCI International 2013

Organization

Human–Computer Interaction

Program Chair: Masaaki Kurosu, Japan

Jose Abdelnour-Nocera, UK
Sebastiano Bagnara, Italy
Simone Barbosa, Brazil
Tomas Berns, Sweden
Nigel Bevan, UK
Simone Borsci, UK
Apala Lahiri Chavan, India
Sherry Chen, Taiwan
Kevin Clark, USA
Torkil Clemmensen, Denmark
Xiaowen Fang, USA
Shin'ichi Fukuzumi, Japan
Vicki Hanson, UK
Ayako Hashizume, Japan
Anzai Hiroyuki, Italy
Sheue-Ling Hwang, Taiwan
Wonil Hwang, South Korea
Minna Isomursu, Finland
Yong Gu Ji, South Korea
Esther Jun, USA
Mitsuhiko Karashima, Japan
Kyungdoh Kim, South Korea
Heidi Krömker, Germany
Chen Ling, USA
Yan Liu, USA
Zhengjie Liu, P.R. China
Loïc Martínez Normand, Spain
Chang S. Nam, USA
Naoko Okuizumi, Japan
Noriko Osaka, Japan
Philippe Palanque, France
Hans Persson, Sweden
Ling Rothrock, USA
Naoki Sakakibara, Japan
Dominique Scapin, France
Guangfeng Song, USA
Sanjay Tripathi, India
Chui Yin Wong, Malaysia
Toshiki Yamaoka, Japan
Kazuhiko Yamazaki, Japan
Ryoji Yoshitake, Japan
Silvia Zimmermann, Switzerland

Human Interface and the Management of Information

Program Chair: Sakae Yamamoto, Japan

Hans-Jorg Bullinger, Germany
Alan Chan, Hong Kong
Gilsoo Cho, South Korea
Jon R. Gunderson, USA
Shin'ichi Fukuzumi, Japan
Michitaka Hirose, Japan
Jhilmil Jain, USA
Yasufumi Kume, Japan
Mark Lehto, USA
Hiroyuki Miki, Japan
Hirohiko Mori, Japan
Fiona Fui-Hoon Nah, USA
Shogo Nishida, Japan
Robert Proctor, USA
Youngho Rhee, South Korea
Katsunori Shimohara, Japan

Michale Smith, USA
Tsutomu Tabe, Japan
Hiroshi Tsuji, Japan
Kim-Phuong Vu, USA
Tomio Watanabe, Japan
Hidekazu Yoshikawa, Japan

Engineering Psychology and Cognitive Ergonomics

Program Chair: Don Harris, UK

Guy Andre Boy, USA
Joakim Dahlman, Sweden
Trevor Dobbins, UK
Mike Feary, USA
Shan Fu, P.R. China
Michaela Heese, Austria
Hung-Sying Jing, Taiwan
Wen-Chin Li, Taiwan
Mark A. Neerincx, The Netherlands
Jan M. Noyes, UK
Taezoon Park, Singapore
Paul Salmon, Australia
Axel Schulte, Germany
Siraj Shaikh, UK
Sarah C. Sharples, UK
Anthony Smoker, UK
Neville A. Stanton, UK
Alex Stedmon, UK
Xianghong Sun, P.R. China
Andrew Thatcher, South Africa
Matthew J.W. Thomas, Australia
Rolf Zon, The Netherlands

Universal Access in Human–Computer Interaction

Program Chairs: Constantine Stephanidis, Greece, and Margherita Antona, Greece

Julio Abascal, Spain
Ray Adams, UK
Gisela Susanne Bahr, USA
Margit Betke, USA
Christian Bühler, Germany
Stefan Carmien, Spain
Jerzy Charytonowicz, Poland
Carlos Duarte, Portugal
Pier Luigi Emiliani, Italy
Qin Gao, P.R. China
Andrina Granić, Croatia
Andreas Holzinger, Austria
Josette Jones, USA
Simeon Keates, UK
Georgios Kouroupetroglou, Greece
Patrick Langdon, UK
Seongil Lee, Korea
Ana Isabel B.B. Paraguay, Brazil
Helen Petrie, UK
Michael Pieper, Germany
Enrico Pontelli, USA
Jaime Sanchez, Chile
Anthony Savidis, Greece
Christian Stary, Austria
Hirotada Ueda, Japan
Gerhard Weber, Germany
Harald Weber, Germany

Virtual, Augmented and Mixed Reality

Program Chair: Randall Shumaker, USA

Waymon Armstrong, USA
Juan Cendan, USA
Rudy Darken, USA
Cali M. Fidopiastis, USA
Charles Hughes, USA
David Kaber, USA
Hirokazu Kato, Japan
Denis Laurendeau, Canada
Fotis Liarokapis, UK
Mark Livingston, USA
Michael Macedonia, USA
Gordon Mair, UK
Jose San Martin, Spain
Jacquelyn Morie, USA
Albert "Skip" Rizzo, USA
Kay Stanney, USA
Christopher Stapleton, USA
Gregory Welch, USA

Cross-Cultural Design

Program Chair: P.L. Patrick Rau, P.R. China

Pilsung Choe, P.R. China
Henry Been-Lirn Duh, Singapore
Vanessa Evers, The Netherlands
Paul Fu, USA
Zhiyong Fu, P.R. China
Fu Guo, P.R. China
Sung H. Han, Korea
Toshikazu Kato, Japan
Dyi-Yih Michael Lin, Taiwan
Rungtai Lin, Taiwan
Sheau-Farn Max Liang, Taiwan
Liang Ma, P.R. China
Alexander Mädche, Germany
Katsuhiko Ogawa, Japan
Tom Plocher, USA
Kerstin Röse, Germany
Supriya Singh, Australia
Hsiu-Ping Yueh, Taiwan
Liang (Leon) Zeng, USA
Chen Zhao, USA

Online Communities and Social Computing

Program Chairs: A. Ant Ozok, USA, and Panayiotis Zaphiris, Cyprus

Areej Al-Wabil, Saudi Arabia
Leonelo Almeida, Brazil
Bjørn Andersen, Norway
Chee Siang Ang, UK
Aneesha Bakharia, Australia
Ania Bobrowicz, UK
Paul Cairns, UK
Farzin Deravi, UK
Andri Ioannou, Cyprus
Slava Kisilevich, Germany
Niki Lambropoulos, Greece
Effie Law, Switzerland
Soo Ling Lim, UK
Fernando Loizides, Cyprus
Gabriele Meiselwitz, USA
Anthony Norcio, USA
Elaine Raybourn, USA
Panote Siriaraya, UK
David Stuart, UK
June Wei, USA

Augmented Cognition

Program Chairs: Dylan D. Schmorrow, USA, and Cali M. Fidopiastis, USA

Robert Arrabito, Canada
Richard Backs, USA
Chris Berka, USA
Joseph Cohn, USA
Martha E. Crosby, USA
Julie Drexler, USA
Ivy Estabrooke, USA
Chris Forsythe, USA
Wai Tat Fu, USA
Rodolphe Gentili, USA
Marc Grootjen, The Netherlands
Jefferson Grubb, USA
Ming Hou, Canada
Santosh Mathan, USA
Rob Matthews, Australia
Dennis McBride, USA
Jeff Morrison, USA
Mark A. Neerincx, The Netherlands
Denise Nicholson, USA
Banu Onaral, USA
Lee Sciarini, USA
Kay Stanney, USA
Roy Stripling, USA
Rob Taylor, UK
Karl van Orden, USA

Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management

Program Chair: Vincent G. Duffy, USA and Russia

Karim Abdel-Malek, USA
Giuseppe Andreoni, Italy
Daniel Carruth, USA
Eliza Yingzi Du, USA
Enda Fallon, Ireland
Afzal Godil, USA
Ravindra Goonetilleke, Hong Kong
Bo Hoege, Germany
Waldemar Karwowski, USA
Zhizhong Li, P.R. China
Kang Li, USA
Tim Marler, USA
Michelle Robertson, USA
Matthias Rötting, Germany
Peter Vink, The Netherlands
Mao-Jiun Wang, Taiwan
Xuguang Wang, France
Jingzhou (James) Yang, USA
Xiugan Yuan, P.R. China
Gülcin Yücel Hoge, Germany

Design, User Experience, and Usability

Program Chair: Aaron Marcus, USA

Sisira Adikari, Australia
Ronald Baecker, Canada
Arne Berger, Germany
Jamie Blustein, Canada
Ana Boa-Ventura, USA
Jan Brejcha, Czech Republic
Lorenzo Cantoni, Switzerland
Maximilian Eibl, Germany

Anthony Faiola, USA
Emilie Gould, USA
Zelda Harrison, USA
Rüdiger Heimgärtner, Germany
Brigitte Herrmann, Germany
Steffen Hess, Germany
Kaleem Khan, Canada
Jennifer McGinn, USA
Francisco Rebelo, Portugal
Michael Renner, Switzerland
Kerem Rızvanoğlu, Turkey
Marcelo Soares, Brazil
Christian Sturm, Germany
Michele Visciola, Italy

Distributed, Ambient and Pervasive Interactions

Program Chairs: Norbert Streitz, Germany, and Constantine Stephanidis, Greece

Emile Aarts, The Netherlands
Adnan Abu-Dayya, Qatar
Juan Carlos Augusto, UK
Boris de Ruyter, The Netherlands
Anind Dey, USA
Dimitris Grammenos, Greece
Nuno M. Guimaraes, Portugal
Shin'ichi Konomi, Japan
Carsten Magerkurth, Switzerland
Christian Müller-Tomfelde, Australia
Fabio Paternó, Italy
Gilles Privat, France
Harald Reiterer, Germany
Carsten Röcker, Germany
Reiner Wichert, Germany
Woontack Woo, South Korea
Xenophon Zabulis, Greece

Human Aspects of Information Security, Privacy and Trust

Program Chairs: Louis Marinos, ENISA EU, and Ioannis Askoxylakis, Greece

Claudio Agostino Ardagna, Italy
Zinaida Benenson, Germany
Daniele Catteddu, Italy
Raoul Chiesa, Italy
Bryan Cline, USA
Sadie Creese, UK
Jorge Cuellar, Germany
Marc Dacier, USA
Dieter Gollmann, Germany
Kirstie Hawkey, Canada
Jaap-Henk Hoepman, The Netherlands
Cagatay Karabat, Turkey
Angelos Keromytis, USA
Ayako Komatsu, Japan
Ronald Leenes, The Netherlands
Javier Lopez, Spain
Steve Marsh, Canada
Gregorio Martinez, Spain
Emilio Mordini, Italy
Yuko Murayama, Japan
Masakatsu Nishigaki, Japan
Aljosa Pasic, Spain
Milan Petković, The Netherlands
Joachim Posegga, Germany
Jean-Jacques Quisquater, Belgium
Damien Sauveron, France
George Spanoudakis, UK
Kerry-Lynn Thomson, South Africa

Julien Touzeau, France
Theo Tryfonas, UK
João Vilela, Portugal
Claire Vishik, UK
Melanie Volkamer, Germany

External Reviewers

Maysoon Abulkhair, Saudi Arabia
Ilia Adami, Greece
Vishal Barot, UK
Stephan Böhm, Germany
Vassilis Charissis, UK
Francisco Cipolla-Ficarra, Spain
Maria De Marsico, Italy
Marc Fabri, UK
David Fonseca, Spain
Linda Harley, USA
Yasushi Ikei, Japan
Wei Ji, USA
Nouf Khashman, Canada
John Killilea, USA
Iosif Klironomos, Greece
Ute Klotz, Switzerland
Maria Korozi, Greece
Kentaro Kotani, Japan
Vassilis Kouroumalis, Greece
Stephanie Lackey, USA
Janelle LaMarche, USA
Asterios Leonidis, Greece
Nickolas Macchiarella, USA
George Margetis, Greece
Matthew Marraffino, USA
Joseph Mercado, USA
Claudia Mont'Alvão, Brazil
Yoichi Motomura, Japan
Karsten Nebe, Germany
Stavroula Ntoa, Greece
Martin Osen, Austria
Stephen Prior, UK
Farid Shirazi, Canada
Jan Stelovsky, USA
Sarah Swierenga, USA

HCI International 2014

The 16th International Conference on Human–Computer Interaction, HCI International 2014, will be held jointly with the affiliated conferences in the summer of 2014. It will cover a broad spectrum of themes related to Human–Computer Interaction, including theoretical issues, methods, tools, processes and case studies in HCI design, as well as novel interaction techniques, interfaces and applications. The proceedings will be published by Springer. More information about the topics, as well as the venue and dates of the conference, will be announced through the HCI International Conference series website: http://www.hci-international.org/

General Chair
Professor Constantine Stephanidis
University of Crete and ICS-FORTH
Heraklion, Crete, Greece
Email: cs@ics.forth.gr

Table of Contents – Part II

Driving and Transportation Safety

Cognitive Issues in Aviation

Military Applications

Cognitive Issues in Health and Well-Being

Table of Contents – Part I

Measuring and Monitoring Cognition

Cognitive Issues in Complex Environments

Productivity, Creativity, Learning and Collaboration

Part I
Driving and Transportation Safety

An Evaluation of the Interior Design of the Stockholm Bypass Tunnel – A Driving Simulator Study

Ruggero Ceci, Christopher Patten, and Selina Mårdh

VTI – Swedish National Road and Transport Research Institute, Sweden
{ruggero.ceci,christopher.patten,selina.mardh}@vti.se

Abstract. Maintaining high levels of road traffic safety is always important and when the road is in a tunnel, and especially in a long tunnel, maintaining the highest possible level of safety is paramount. In Sweden, the Stockholm bypass tunnel has been scheduled to commence construction in 2013. The tunnel will be approximately 18 km in length. The length of the tunnel is expected to affect the drivers' experiences pertaining drowsiness, arousal, distraction and feelings of safety and security. The study included 24 participants, 12 men and 12 women, aged 30-45. All of the participants drove two versions of the tunnel, one version with a decoration design in terms of string lighting in the ceiling of the tunnel and one version of the tunnel without any decoration design. Eye tracking behaviour was measured during the study. CR10 ratings of four subjective dimensions (distraction, visually cluttered, visually stimulating/arousal and safety and well-being) during the drive were also measured. The results revealed that 58 per cent of the participants preferred the tunnel with the strings of light in the ceiling and 29 per cent preferred the tunnel without the ceiling lighting. 13 per cent prefer neither one design more than the other. The participants perceived feelings of their driving through the tunnel suggested that the tunnel with the ceiling light design was experienced as being more ”visually cluttered” than the tunnel without the light strings but at the same time it was also experienced as more ”arousing/stimulating”. Mean glance duration times suggested that although there was a significant main effect of the tunnel with the string lighting and in specific areas of the tunnel, the drivers were looking at the ceiling lighting but using short glances (445.3 ms with lighting and 234.3 ms without lighting). The negative safety implications of the elaborate interior lighting features would appear to be minimal in terms of distraction and irritation whereas the safety benefits in this particularly long road tunnel, in terms of subjective feelings of visual stimulation is encouraging. Based on the participants' experiences of the interior design concept of the 18 km long tunnel, having stimulating lighting features in different locations alone the length of the tunnel is recommended.

Keywords: tunnel safety, tunnel interior design, driving simulator, category ratio scale (CR-10), eye tracking, string lighting, Stockholm bypass tunnel.

1 Introduction

Maintaining high levels of road traffic safety is always important and when the road is in a tunnel, and especially in a long tunnel, maintaining the highest possible level of

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 3–11, 2013.
© Springer-Verlag Berlin Heidelberg 2013

safety is paramount (Patten & Mårdh, 2012). The Stockholm By-pass (FSS) project is a new road project that will create a new bypass of central Stockholm. The entire project includes motorways, bridges and two tunnels; one of which will be 18 km. The FFS is the largest infrastructure project in Sweden to date. The planning of the project includes the choice of the exact route, the road geometry and also the interior design of the 18 km tunnel, including the aesthetics of all aspects of the tunnel. Earlier studies have suggested that the drivers' ability to gauge speed can be affected by visual design concepts (Manser and Hancock, 2007). Other forms of driver behaviour, such as eye-glance behaviour and mental workload has also been suggested as being affected by the lighting colours and patterns of the tunnel walls as well as the strength of the lighting (Kircher and Ahlström, 2012; Kircher and Lundkvist, 2011; Patten, Ceci, Engström and Anund, in press). Based on these studies, the interior design and decoration of the tunnel is expected to affect drivers. The drivers' subjectively perceived experiences pertaining drowsiness, arousal, distraction and feelings of safety and security were studied in VTI's advanced driving simulator in Linköping, Sweden.

In this study, two versions of the tunnel's interior design have been meticulously created from the blue prints of the real tunnel (this hasn't been built yet); one version with ceiling lighting and the other version without. The real tunnel's construction starts in 2013 and is expected to take about ten years to complete. The research questions ask if the ceiling string lighting and other visual design features have a negative effect on vehicle based performance (not reported in this paper), eye tracking data and subjectively rated experiences of tunnel design according to the following:

- Subjective rating of distraction.
- Subjective rating of visual clutter or messiness.
- Subjective rating of arousal/vigilance.
- Subjective rating of safety and well-being.

Moreover, the opinions of specific design features were also rated in a post experimental survey (not reported here).

2 Method

2.1 Participants

The study included 24 participants, 12 men and 12 women, aged between 30-45 years and were recruited to the study from VTI's database of voluntary participants. Their mean age was 38 years and they had held their licences for 18.5 years. The selection criteria for the participants were 1) an annual mileage > 5000 km, 2) a full category B (car) driving licence held for at least 5 years, 3) no spectacles (contact lenses were fine), and 4) no predispositions to motion sickness.

Four of the participants indicated in a pre- experimental survey that they feel unease when driving in tunnels. One of these participants specified the reason as being that the GPS stopped working in a tunnel; another submitted reasons for this unease as being due to darkness and the possibility slipperiness in winter. Three other participants thought that driving in tunnels was interesting and fun whilst the remaining 18 participants held no specific opinions of driving in real-life tunnels. All of the participants had previous experience of driving in real-life tunnels.

2.2 Equipment and Materials

*Simulator***.** The study was performed in VTI's driving simulator III in Linköping, using the car set-up pictured in Figure 1. The simulator utilises all of the controls of a real car and is mounted on a full motion based platform. The visual experience is created using six projectors with a forward field of view of 120 degrees. There are also three rearward facing LCD screens instead of mirrors that provide images of the road scene behind the simulator-vehicle.

Eye Tracking. The SMI IviewX is a helmet mounted eye tracking device. The device records the position of the pupil and predicts the eye's direction with a cross-hair. The device also films the direction of sight. All of the participants were donned with the eye tracker before entering the simulator and individual calibration is required for all participants.

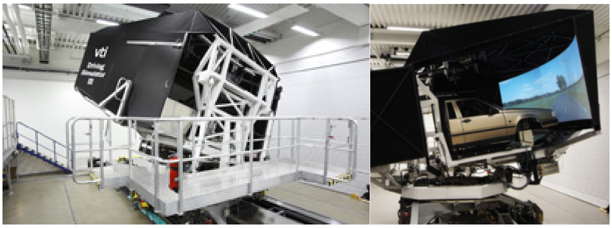


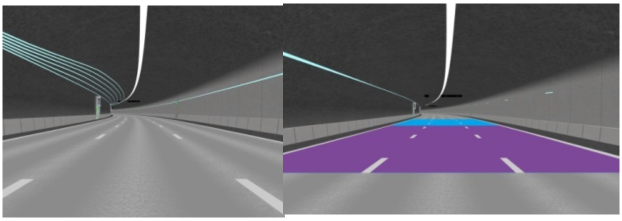
Fig. 1. Simulator III, VTI in Linköping, Sweden

2.3 Procedure

A motorway tunnel replica based on the blue-prints of the Stockholm Bypass tunnel was recreated in the VTI advanced driving simulator no. III. The simulated tunnel included all of the original road topography, including curvature, gradient, length and breadth. It also included the planned surface texture of the walls, road signage, emergency exits and other road furniture such as extraction fans and standard lighting fixtures. The simulated tunnel comprised a three lane motorway. The visual design feature comprised a series of ceiling lighting strings that travelled the length of the tunnel and varied from one line to five or six. The position of the string of ceiling lights also varied and could travel from left to right and vice versa, as well as travelling down the upper section of the tunnel wall (approx. ≥ 2.5 m). The visual design concept also includes special artistic lighting features at certain points in the tunnel such as the subterraneous junctions at Ekerö, the Mid-way point, the Lung and Vällingby.

For the purpose of this study, two versions of the tunnel were used; one with the string lighting feature and one without. The other design features were identical thus isolating the pros and cons of the string lighting feature in the tunnel. The study design was a within-subject design where all of the participants drove all of the different experimental conditions. The order was balanced for lighting design and gender. All participants drove the route in the same direction (from south to north) through the tunnel.The study procedure started when the participants arrived at VTI in Linköping, Sweden whereupon written instructions and informed consent forms were signed. The participants were also drilled with the rating scales as well as completing background questionnaires before being allowed to enter the driving simulator.

Once inside the simulator the participants would familiarise themselves with the basic controls. The eye tracking equipment was also calibrated. The participants then started their familiarisation or practice drive. The speed limit in the tunnel was sign-posted at 100 km/h. The participants were instructed to drive 'normally' and to drive in the centre lane (there were three lanes). There was also simulated light traffic in the tunnel. There were no overtaking situations where the vehicle to the left of the participant maintained slightly slower and the vehicles to the right of the participant drove slightly faster (driving in right-hand traffic). After each tunnel condition, the participants stopped to rate four CR10 questions. After the final tunnel condition, a post-questionnaire was answered (not reported in this paper). The whole process took approximately 1.5 hours. All participants received 300 SEK in compensation.



(a) (b)

Fig. 2. A screen capture of a) the string lighting visual design feature and b) the Middle section or mid-way point in the tunnel

2.4 Rating Scale CR10

The rating scale used in this study was the Category Ratio Scale 10 (CR10). The scale ranges from 0 to >11 although in practice 11 is the highest rating (see Appendix 1). More importantly with the CR10 scale that was developed by Borg (2008), is the verbal anchors associated with each ratings. Moreover, all of the CR10 ratings were measured directly after exiting the different experimental conditions to reduce the likelihood of confusion and memory loss regarding the ratings of the conditions. The four dimensions were rated whilst the participants remained seated in the stationary vehicle. The questions were read aloud on the loudspeaker from the simulator control room and recorded by the operator. The following four ratings were rated by the participants:

- How distracting did you experience the tunnel's interior design features?
- How visually cluttered did you experience the tunnel's interior design features?
- How stimulating or arousing did you experience the tunnel's interior design features?
- How safe did the tunnel's interior design features make you feel?

2.5 Data Analysis

There were 24 participants that completed the study. The order of presentation was balanced and interaction effects from order were also investigated between the conditions.

The tunnel design, including the specifics details of subterraneous junctions was not altered as their location, design and geographic position are all fixed. All of the separate road sections with visual design features (seven with string lighting and five junctions or other special features) were analysed as separately due to their uniqueness's. The statistical analyses used were with ANOVA repeated measures and t-tests using SPSS (version 17.0). Due to technical problems with the eye tracking equipment and its calibration, only data from 12 of the participants was valid (7 women and 5 men).

The string lighting design features in the tunnel was divided into seven episodes or segments as well as the additional five junctions or other special features. The seven separate segments of string lighting was present either on the left (L) or the right (R). Alternatively, the string lighting was present on first the left and then the right (LR) of the ceiling or vice versa (RL). The seven string lighting segments were as follows:

Segment 1 (string lighting LR) Segment 2 (string lighting LR)
Segment 3 (string lighting L) Segment 4 (string lighting RL)
Segment 5 (string lighting RL) Segment 6 (string lighting L)
Segment 7 (string lighting R)

3 Results

3.1 CR10 Borg Scale

The participants were asked to rate four dimensions on the CR10 scale (0 - >11) in regard to the tunnel they had just driven through (one with string lighting design features and one without the design features). The results are shown in figure 3.

The difference in the mean CR10 rating scores for the "distraction" dimension between the tunnel with the visual string light design feature was 2.4 and 1.97 for the tunnel without the design features. This was tested with a t-test and was not

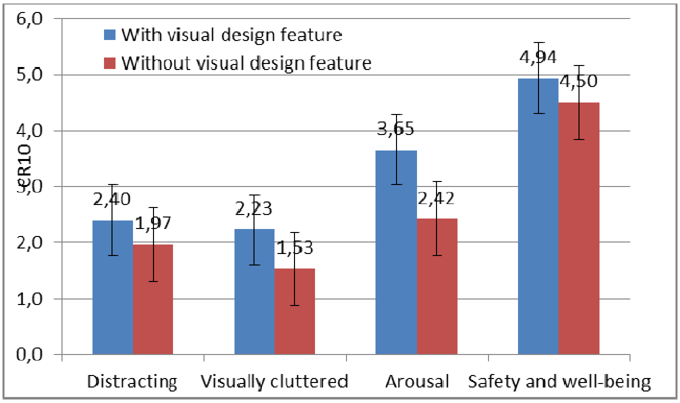


Fig. 3. CR10 rating scale. There were four dimensions that were rated by the participants. Standard error bars and the mean CR10 rating are indicated (n = 24).

statistically significant (t(23) = 1.54, p = .137). The verbal anchors for the different scores can be found in figure 2.

The difference in the mean CR10 rating scores for the "visually cluttered" dimension between the tunnel with the visual string light design feature was 2.23 and 1.53 for the tunnel without the design features. This was tested with a t-test and was statistically significant (t(23) = 2.232, p = .036).

The difference in the mean CR10 rating scores for the "visually stimulating/arousal" dimension between the tunnel with the visual string light design feature was 3.65 and 2.42 for the tunnel without the design features. This was tested with a t-test and was statistically significant (t(23) = 2.612, p = .016).

The difference in the mean CR10 rating scores for the "safety and well-being" dimension between the tunnel with the visual string light design feature was 4.94 and 4.5 for the tunnel without the design features. This was tested with a t-test and was not statistically significant (t(23) = .968, p = .343).

3.2 Mean Glance Duration

In table 1 below, the mean glance durations (total glance time/total glance frequency). String number 4 was statistically significant, F(1, 10) 7.782 p = .019 with a large effect size (Eta2 = .438) according to Cohen's guidelines (1988). String 7 was also significant, F(1, 10) 8.92 p = .014 with a large effect size (Eta2 = .471) according to Cohen's guidelines (1988), whilst strings 1, 2, 3, 5 and 6 were not significant. There was a significant main effect between the tunnel with the design concept (445.34 ms) and the tunnel without (234.33 ms), (F(1,11) = 9.81, p = .01) and the effect size was large (Eta2 = .471) according to Cohen's guidelines (1988).

Table 1. Mean glance duration for the different segments of tunnel with the string light design concept and an identical tunnel segment without the string lighting feature. (n = 12).

Lighting string	With design concept (ms)	Without design concept (ms)	p value	Effect size (Eta^2)
String 1	420.6	206.1	n.s.	-
String 2	404.4	266.3	n.s.	-
String 3	533.8	263.7	n.s.	-
String 4	403.9	186.0	*	.438
String 5	375.0	211.1	n.s.	-
String 6	360.2	168.4	n.s.	-
String 7	619.4	338.7	*	.471

* significance at .05 level; n.s. = not significant.

3.3 Correlations between Mean Glance Durations and CR-10 Ratings

In order to further analyse the correlations between eye-tracking data (mean glance durations) and the four different perceived dimensions according to the CR-10 scale a Pearson Correlation analysis was performed over the range of lighting string (1-7). A significant correlation only between mean glance duration (1-7) with string lighting and CR-10 ratings of *distraction* was found (p <0.05, 1-tailed). This could be interpreted as a support for the assumption that the string lighting features in this condition affect the visual (glance) behaviour while driving through the simulated tunnel in a somewhat distracting manner. However, the magnitude of this perceived distraction seems fairly low according to the CR-10 scale (about 2 with the verbal anchor *weak*).

4 Discussion

The main research question addressed in this study was whether or not the string lighting design feature that ran along the upper walls and ceiling of the conceptualised Stockholm Bypass tunnel had an adverse effect on driver performance and road safety. Moreover, the participants' opinions and subjective ratings of the tunnel's visual design features were explored in detail. Based on the overall results of the present study, the effects of the design with light strings in the ceiling of the tunnel was considered to be more beneficial from a traffic safety point of view as compared to the tunnel design without light strings. This conclusion is based on the fact that the light string design did not significantly change the driver behaviour but was perceived as breaking the monotony of the tunnel. The results of the study can be used as a basis for further design aspects of long tunnels.

The participants' experiences, opinions and comments regarding the visual design features in the tunnel were measured with different rating scales. The first rating scale was the CR10 scale that was used to rate four dimensions of the tunnel, directly after experiencing the design features. The CR10 was easy and quick to administer, as well as being tangible because of the verbal anchor words, for the participants' when putting a number to a feeling or experience. It is also generally advantageous to record the rating as soon as possible after the event, to avoid memory/recollection issues or confusion regarding the object being rated when there are potentially several conditions to choose from. The CR10 dimensions for "visually cluttered" and "visually stimulating/arousing" were statistically significant between the two conditions (with the string lighting and without). The tunnel with the string lighting design feature was rated as being more visually cluttered than the tunnel without. It was however, also significantly rated as being more visually stimulating than the tunnel without the string lighting which was deemed as a positive result because the aim of the string lighting was, in part, to break the visual monotony of an 18 km long tunnel wall/ceiling.

The post study questionnaire revealed that 58 per cent of the participants preferred the tunnel with the strings of light in the ceiling and 29 per cent preferred the tunnel without the ceiling lighting. 13 per cent preferred neither one design more than the other. The eye tracking data was extracted and analysis for the sections of tunnel where the string lighting was present as well as an equivalent section of the tunnel without string lighting. The participants' visual behaviour was studied to see whether or not the drivers looked at the string lighting, and if they did, how often and how long? The safety threshold for non-road related vehicle design glance times and visual distraction varies. In the USA, a maximum of 2.0 second (or 2000 ms) glance times are stipulated (AAM, 2003). The mean glance time ranged between 360 ms and 619 ms in the tunnel with the string lighting (mean 445.3 ms) and was significantly longer than the tunnel without the string lighting (234.3 ms). It should be noted, however, that the mean glance durations are unlikely to pose any serious safety concerns.

5 Conclusions

The construction of long tunnels brings issues regarding driver behaviour and traffic safety into focus. In Sweden, the Stockholm Bypass tunnel (FFS), has been scheduled to commence construction in 2013. The tunnel will be approximately 18 km in length and the length and stretch of the tunnel is expected to affect the drivers' experiences pertaining drowsiness, arousal, distraction and feelings of safety and security. An attempt to meet these issues was made by focusing on the interior design of the tunnel. The present study was performed at VTI's driving simulator III in Linköping. A tunnel based on the blue-prints of the FFS tunnel (version FST Plan 2011-11-08) was designed in the simulator. The study included 24 participants, 12 men and 12 women, aged 30-45. All of the participants drove two versions of the tunnel, one version with a decoration design in terms of strings of lights in the ceiling of the tunnel and one version of the tunnel without any decoration design. Measures of lateral position, speed, eye movement data (mean glance duration), perceived feelings of distraction, arousal/stimulation, drowsiness, safety and security as well as perceived design preferences were collected.

The participants perceived feelings of their driving through the tunnel suggested that the tunnel with the ceiling light design was experienced as being more ”visually cluttered” than the tunnel without the light strings but at the same time it was also experienced as more ”visually stimulating/arousing”.

Appendix 1

0	Nothing at all	
0, 3		
0.5	Extremely weak	Just noticeable
0.7		
1	Very weak	
1.5		
2	Weak	Light
2.5		
3	Moderate	
4		
5	Strong	Heavy
6		
7	Very strong	
8		
9		
10	**Extremely strong**	**"Maximal"**
11		
+		
●	Absolute maximum	Highest possible

The Borg CR10 scale®
© Gunnar Borg, 1982, 1998, 2004

References

1. Alliance of Automobile Manufacturers. Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communications Systems. Alliance of Automobile Manufacturers, Washington, DC (2003)
2. Borg, G., Borg, E.: Borg CR Scales Folder. Stockholm, Borg Perception (2008)
3. Cohen, J.W.: Statistical power analysis for the behavioural sciences, 2nd edn. Lawrence Erlbaum Associates, Hillsdale (1988)
4. Dingus, T.A., Antin, J.F., Hulse, M.C., Wierwille, W.W.: Attentional demand requirements of an automobile moving map navigation system. Transportation Research A23(4), 301–315 (1989)
5. Green, P.: Visual and task demands of driver information systems (Technical Report No. UMTRI-98-16). UMTRI (The University of Michigan, Transportation Research Institute), Ann Arbor, MI (1999)
6. Kircher, K., Ahlstrom, C.: The impact of tunnel design and lighting on the performance of attentive and visually distracted drivers. Accident Analysis and Prevention 47, 153–161 (2012)
7. Kircher, K., Lundkvist, S.-O.: The influence of lighting, wall colour and inattention on traffic safety in tunnels – A simulator study. VTI rapport 724A, Linköping, Sweden (2011)
8. Maner, M.P., Hancock, P.A.: The influence of perceptual speed regulation on speed perception, choice, and control: Tunnel wall characteristics and influences. Accident Analysis 39, 69–78 (2007)
9. Patten, C., Ceci, R., Engström, J., Anund, A.: Tunnel Driving and the Effects of Visual Design (2012) (in press)
10. Patten, C., Mårdh, S.: Förbifart Stockholm Tunnel Utvärdering 2012. VTI Rapport R759. VTI, Linköping (2012)
11. Wierwille, W.: Visual and manual demands of in-car controls and displays. In: Peacock, B., Karkowski, W. (eds.) Automotive Ergonomics, pp. 299–320. Taylor & Francis, Washington, DC (1993)

Comprehension of Vibrotactile Route Guidance Cues

Andre Garcia[1], Jesse Eisert[1], Carryl L. Baldwin[1], and Victor Finomore[2]

[1] Department of Psychology, George Mason University, United States
[2] Human Effectivness / Battlespace Acoustics Branch, Air Force Research Lab, United States
agarciagmu@gmail.com, {jeisert,cbaldwi4}@gmu.edu,
victor.finomore@wpafb.af.mil

Abstract. Two experiments with 24 participants each evaluated comprehension of vibrotactile route guidance instructions via a tactile seat in a driving simulator. Vibrotactile patterns were presented from an array of 8 tactors arranged in two rows of 4 tactors located in the seat pan. A faster pulse rate and a slower pulse rate as well as four distinct locations on the tactile seat (Front-Left, Front-Right, Back-Left, Back-Right) created 8 different combinations of stimuli. Across all participants, the most consistent interpretation was that the faster pulse rate played from the back two tactors was perceived as an instruction to make the next most immediate turn while a slow pulse rate from the front two tactors was interpreted as a cue directing the user to the direction of the next eventual turn. Results have direct implications for design of effective vibrotactile and multimodal route guidance systems.

1 Introduction

Vibrotactile technology for in-vehicle use has shown increasing promise and popularity of late (Scott & Gray, 2008; Mohebbi, Gray, Tan, 2009). General Motors currently offers a feature on their Cadillac XTS sedan where the seat pan vibrates if there is a potential rear-end collision while you are reversing. This is just one example among several other current production vehicles that come equipped with vibrotactile technology. The tactile modality offers a way to relay information that is privileged to only the user. Tactile collision warning systems have been shown to effectively reduce reaction time (Scott and Gray, 2008), and may be particularly effective in multimodal systems (Mohebbi, Gray, & Tan, 2009).

The tactile modality is a way to provide the user information without relying on visual or auditory attentional resources that are often in high demand in many operational settings. Recent studies investigating vibrotactile route guidance systems have shown great potential. Van Erp and Van Veen (2004) demonstrated how a tactile navigation system display can reduce a driver's perceived workload compared to a visual display, particularly in high workload settings. Van Erp, Van Veen, Jansen, and Dobbins (2006) investigated the efficacy and feasibility of a tactile navigation waist belt and found that directional information is easy, intuitive, and requires almost no training, although their results on how to map distance were inconclusive. Vibrotactile systems for in-vehicle technology have generally been limited to collision warning

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 12–18, 2013.
© Springer-Verlag Berlin Heidelberg 2013

systems or lane departure warning systems and relatively few studies have investigated the use of vibrotactile systems for in-vehicle route guidance.

Garcia, Finomore, Burnett, Baldwin & Brill (2012) conducted a study to investigate waypoint navigation via a visual, auditory, tactile, or multimodal route guidance system in dismounted soldiers. Participants were lead via the various uni- or multimodal route guidance system from waypoint to waypoint and were instructed to look for certain landmarks throughout the environment. For the tactile modality, a vibrotactile belt was used, which consisted of 8 tactors equally spaced around the waist (For more information on this belt see: Merlo, Duley, & Hancock, 2010; Cholewiak, Brill, & Schwab, 2004). Overall, the unimodal geocentric visual condition was the slowest and least accurate at guiding the user from waypoint to waypoint to complete a course through a virtual environment. Additionally, every multimodal condition was as fast as its fastest unimodal condition, i.e. there was no additive effect. This experiment provides evidence for tactile navigation and its effectiveness compared to other modalities to guide dismounted soldiers. The current experiment is intended to build on this knowledge and investigate how to best design a tactile navigation for in-vehicle use.

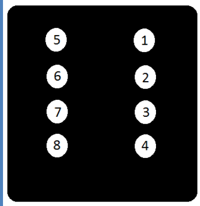
The goal of this investigation was to determine the most intuitive mapping of different vibrotactile patterns for use as route guidance instructions. It was predicted that a redundant mapping consisting of presenting a slower pulse rate from the front two tactors to represent a preliminary cue and a faster pulse rate played from the back two tactors to represent an immediate cue, indicating to turn at the next available location would lead to the most consistent interpretation relative to formats providing information using only tactor location or pulse rate.

2 Experiment 1

Procedure. After providing written informed consent, participants sat in a high fidelity driving simulator equipped with a tactile seat pan. A schematic of how the tactors are arranged on the seat pan is available on the right side of Table 1. The driving simulator was created by RealTime Technologies, Inc. The vibrotactile seat was custom designed and constructed by Engineering Acoustics, Inc and contained an array of 8 C2 tactors. Although no motion was used for this study, the simulator is capable of yaw and pitch motion. The yaw motion allows for 180 degrees of motion, 90 left and 90 right and the pitch motion allows for 1.5 degrees of pitch motion to simulate abrupt acceleration and braking. The simulator features 3 screens that allows for 180 degree forward field of view. The cab was built from a 2002 Ford Taurus and is operated similar to a real car with an automatic transmission.

Before the experiment began, a variety of vibrotactile patterns were presented to the participant to familiarize them with the seat. Participants were then shown an image of an overhead view of a street with a stationary car and six possible turn options. Each turn was labeled with a corresponding response choice (letters A-F). This can be seen in Figure 1. Eight combinations of stimuli (front or back, left or right, slow or fast pulse rate) were presented twice each in randomized order. For the two pulse rates the “slow” stimuli had an interpulse-interval (IPI) of 475 ms and the “fast” stimuli had an IPI of 118 ms. The properties of each of the stimuli are described in table 1. After receiving each stimulus participants were asked to identify which direction they would turn, for a total of 16 questions.

Table 1. Details of each condition and type of cue

Location	Slow	Fast	
Front	Pulse rate 3.69 Tactor 5+6 For Left Turn Tactor 1+2 for Right Turn	Pulse rate 7.87 Tactor 7+8 for Left Turn Tactor 3+4 for Right Turn	
Back	Pulse rate 3.69 Tactors 6+7 for Left Turn Tactors 2+3 for Right Turn	Pulse rate 7.87 Tactors 6+7 for Left Turn Tactors 2+3 for Right Turn	

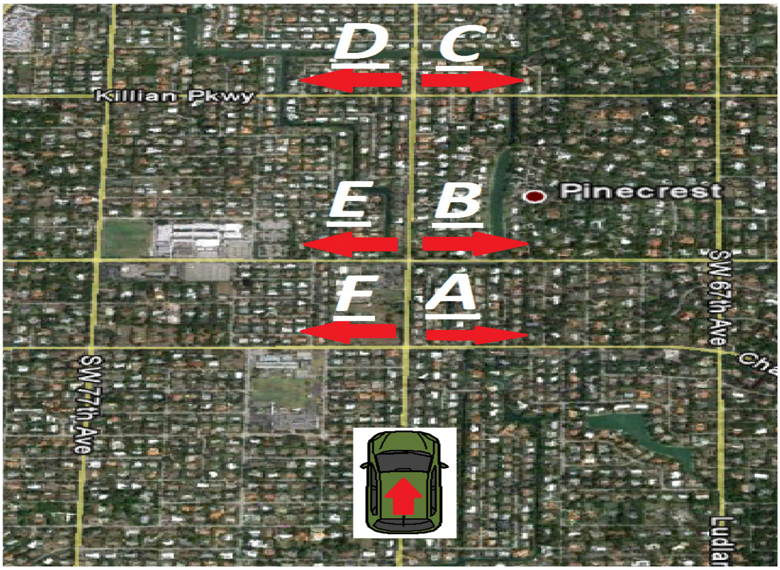


Fig. 1. Overhead view of Response Options

2.1 Results and Discussion

The results from experiment one suggest that the most agreed upon responses were that a slow pulse rate played in the front two tactors best represent a preliminary cue (45% Front-Right-“Slow”, 37.5% Front-Left-“Slow”) whereas a fast pulse rate played in the back two tactors best represent an immediate navigational cue (83% Back-Right-"Fast", 83% Back-Left-"Fast").

When the front right tactors were activated at the slow pulse-rate (tactors 1 and 2), 8 participants indicated they would turn at option A, 5 at option B, and 11 at option C. When the front left tactors were activated at the slow pulse rate, 8 participants indicated they would turn at option F, 9 indicated they would turn at option E, and 7 responded with option D. This suggests that there is no clear consensus on what participants perceived as a slow pulse rate vibrating on either side of the front half of the seat meant. When the back two tactors on the right side (tactors 3 and 4) were activated at a slow pulse rate, 13 participants indicated they would take turn A, 6 participants indicated turn B, and 5 participants indicated turn C. When the back two tactors on the left side (tactors 5 and 6) were activated at a slow pulse rate, 4 participants indicated they would turn at option D, 8 at option E, and 12 indicated they would turn at option F.

When the front right tactors (tactors 1 and 2) were activated at a fast pulse rate, 11 participants indicated they would make turn A, 4 indicated they would turn at option B, and 9 indicated they would turn at option C. For the back-right-fast pattern, 20 participants indicated they would make turn A, 3 indicated they would turn at option B, and 1 indicated they would turn at option C. For the front-left-fast pattern, 8 participants indicated they would turn at option D, 7 indicated they would turn at option E, and 9 indicated they would turn at option F. For the back-left-fast combination, 2 participants indicated they would turn at option D, 2 participants indicated they would turn at option E, and 20 indicated they would turn at option F. These results can be seen in Table 2.

Table 2. Participant responses indicating which turn location they thought each stimulus represented

Turn	Front Left		Front Right		Back Left		Back Right	
	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate
A			11 (46%)	8 (33%)			20 (84%)	13 (54%)
B			4 (17%)	5 (21%)			3 (12%)	6 (25%)
C			9 (37%)	11 (46%)			1 (4%)	5 (21%)
D	8 (33%)	4 (17%)			2 (8%)	4 (17%)		
E	7 (30%)	8 (33%)			2 (8%)	8 (33%)		
F	9 (37%)	12 (50%)			20 (84%)	12 (50%)		

The results of experiment one suggest that a slow pulse rate on the front half of the seat indicates a preliminary cue giving the participant a "heads-up" as to which direction the next eventual turn will be, but not necessarily when that turn will be. Conversely, a fast pulse rate to either side on the back half of the seat most clearly indicated an immediate instruction to make the next possible turn.

3 Experiment 2

Experiment two followed the same procedure as experiment one except the perspective of the image with the response options was changed to a third-person view instead of a birds-eye view. The viewing angle was manipulated to give the participants a more realistic point of view compared to the overhead view of experiment 1, as shown in figure 2 below.

3.1 Results and Discussion

The results from experiment two are consistent with experiment one in that, the most agreed upon responses were that a slow pulse rate played in the front two tactors best represent a preliminary cue whereas a fast pulse rate played in the back two tactors best represent an immediate turn instruction. The results from experiment two are summarized in table 3 below. In sum, there was no difference in how participants responded between a birds-eye view and a third person view. Table 3 shows results on which turn a participant indicated they would make, regardless of pulse-rate.

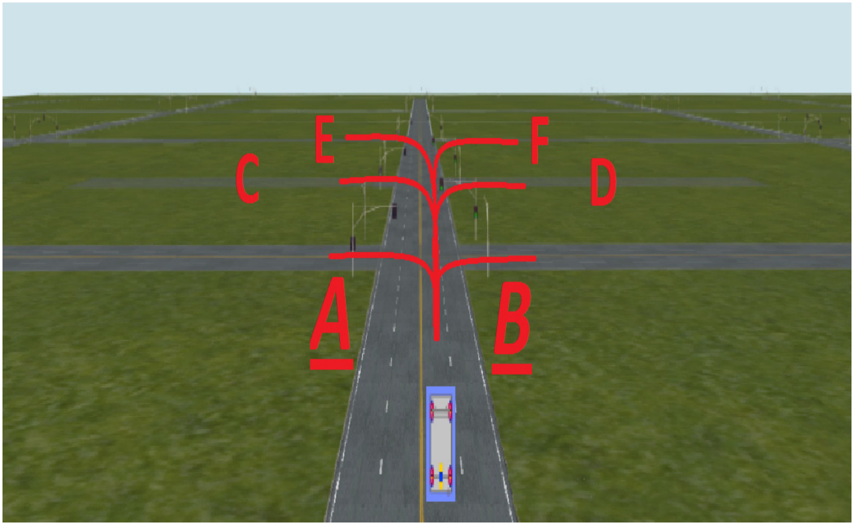


Fig. 2. Birds Eye View of Response Option

4 General Discussion

As new technology based systems begin to trickle into modern vehicles, attention should be paid towards seamlessly integrating these systems. The goal of this study was to better understand how to design a route guidance system using a tactile seat. The results suggest that both pulse rate and location and potentially the interaction between the two can affect perception of route guidance instructions. Results indicated that without training mappings between tactors located in the seat and navigational instructions are not consistent. Results of the current investigation demonstrate that depending on the location and pulse rate of the cue, the perceived meaning will differ. Some people intuitively perceive a vibration towards the front of the leg to mean make an immediate turn while others perceive the same location to map to a more distal turn. Further research is currently being conducted to examine how quickly users may learn to comprehend a designed mapping and if they can use this mapping during simulated driving as a a navigation system. Future research should also strive to increase fidelity and assess the additive effect of multimodal route guidance systems. It is also suggested that these systems be assessed concurrently with the simultaneous use of other types of in-vehicle technology such as infotainment systems and collision warning systems.

Table 3. - Results by Location from Experiment 2

Turn	Front Left		Front Right		Back Left		Back Right	
	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate	Fast Pulse Rate	Slow Pulse Rate
A	6 (30%)	5 (25%)			16 (80%)	17 (85%)		
B			9 (45%)	4 (20%)			16 (80%)	11 (55%)
C	8 (40%)	9 (45%)			3 (15%)	3 (15%)		
D			6 (30%)	8 (40%)			3 (15%)	4 (20%)
E	6 (30%)	6 (30%)	1 (5%)		1 (5%)			
F			4 (20%)	8 (40%)			1 (5%)	5 (25%)

Currently, our lab is conducting additional research extending these results to an examination of wayfinding performance and spatial memory. Another future direction of this line of research would be to investigate individual differences in wayfinding display preferences based on spatial abilities and sense of direction. Garcia et al. (2012) suggest that individuals differ in their ability to understand and use certain navigational display formats depending on their sense of direction. Additional individual differences research conducted by Baldwin and Reagan (2009) suggests that individuals with low spatial abilities may rely on verbal working memory when learning a route while navigating where as those with a good sense of direction rely more on visuospatial working memory. They assessed this by having participants learn a route while performing either a concurrent verbal task (articulator suppression) or a visuospatial tapping task. They found that those with a poor sense of direction had more difficulty while having to perform a concurrent verbal task, suggesting an interference with their verbal working memory. Conversely, those with a good sense of direction experience more interference while attempting to learn a route while performing a concurrent visuospatial tapping task, suggesting that the two tasks - navigation and the tapping task - were both fighting for visuospatial working memory resources at the same time. Vibrotactile stimuli may induce an egocentric mapping since they require direct contact with the touch receptors. However, coding via location may induce a visuospatial code. Individual differences in navigation strategy may be able to predict the preferred or most effective way to display navigational information to the user based on individual spatial abilities. However, it will also be important to examine the effectiveness and potential impact of vibrotactile navigational systems if vibrotactile stimuli are being used to present other forms of time critical information like collision warnings. Future systems must ensure that tactile overload does not supplant visual or auditory overload.

References

1. Baldwin, C.L., Reagan, I.: Individual Differences in Route-Learning Strategy and Associated Working Memory Resources. Human Factors 51(3), 368–377 (2009); Garcia, A., Finomore, V., Burnett, G., Baldwin, C.L., Brill, C.: Individual Differences in Multimodal Waypoint Navigation. In: Proceedings of the 56th Annual Meeting of the Human Factors & Ergonomics Society, Boston, MA (2012)
2. Cholewiak, R.W., Brill, J.C., Schwab, A.: Vibro-tactile localization on the abdomen: Effects of place and space. Perception and Psychophysics 66, 970–987 (2004)
3. Garcia, A., Finomore, V., Burnett, G., Baldwin, C.L., Brill, C.: Individual Differences in Multimodal Waypoint Navigation. In: Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting (2012)
4. Scott, J.J., Gray, R.: A comparison of tactile, visual, and auditory warnings for rear-end collision prevention in simulated driving. Human Factors 50(2), 264–275 (2008)
5. Merlo, J.L., Duley, A.R., Hancock, P.A.: Cross-modal congruency benefits for combined tactile and visual signaling. American Journal of Psychology 123(4), 413–424 (2010)
6. Mohebbi, R., Gray, R., Tan, H.Z.: Driver reaction time to tactile and auditory rear-end collision warnings while talking on a cell phone. Human Factors 51(1), 102–110 (2009)
7. Van Erp, J.B.F., Van Veen, H.A.H.C., Jansen, C., Dobbins, T.: Waypoint Navigation with a Vibrotactile Waist Belt. ACM Transactions on Applied Perception 2(2), 106–117 (2006)
8. Van Erp, J.B.F., Van Veen, H.A.H.C.: Vibrotactile in-vehicle navigation system. Transportation Research Part F 7, 247–256 (2004)

The Safe System Approach – A Road Safety Strategy Based on Human Factors Principles

Peter Larsson[1,*] and Claes Tingvall[2]

[1] Swedish Transport Agency, Borlänge, Sweden
[2] Swedish Transport Administration, Borlänge, Sweden
peter.larsson@transportstyrelsen.se,
claes.tingvall@trafikverket.se

Abstract. In most safety critical domains, safety has been improved through the application of contemporary human error models and management methods. But the common strategic approach to improve road safety has so far mainly been built on the view that individual road-users utterly are responsible when crashes occur and countermeasures have consequently been aimed at changing the behaviour of the road-user. This approach is however slowly shifting and there is a growing understanding that the strategies must be based on human factors principles. In this paper the human factors principles of the Safe System approach are outlined and important implications for the design and regulation of the road transport system will be presented. It is concluded that the Safe System approach share vital foundations with the human factors concept. But it is argued that the Safe System approach takes the human factors approach further by regarding the capability of the human body to withstand external influences with a potential to induce bodily harm.

Keywords: Human Factors, road safety, Vision Zero, system safety, road users, Safe System approach, safety.

1 Introduction

The development of road safety is undergoing major changes now and most certainly in the next ten years. The global community has reacted strongly on the predictions of the impact of poor safety and the growth of road traffic, on the society and the health of the population. It has been estimated that death through a traffic accident will become the third or fourth most common source of death within 10-20 years, unless major and effective actions are taken. The UN has declared 2011-2020 as "the Decade of Action" asking for contributions from all countries and stakeholders to diminish a world epidemic of road casualties that has not only an impact on health but also on economy and economic growth in particular in low and middle income countries. The concern is related to safety, but the overall aim of the future is to develop a sustainable transport system where safety, environment, energy and accessibility are integrated. Such integration is complex and system design necessary as a tool to find synergies and limitations.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 19–28, 2013.
© Springer-Verlag Berlin Heidelberg 2013

To meet this epidemic there is a growing understanding that the strategic approach to road safety must evolve. In most safety critical domains, safety has been improved through the application of contemporary human error models and management methods (Sheridan, 2008). However it seems that this approach to a rather high degree has been neglected by the road safety community (Stanton and Salmon, 2009). The common approach to improve road safety has so far mainly been built on the view that individual road-users utterly are responsible when crashes occur. As a consequence countermeasures have been aimed at changing the behaviour of the road-user in order to adapt him/her to the road transport system and making their behaviour free from human errors (e.g. Salmon et al., 2010; Larsson et al., 2010). Since the road transport system can be seen as a complex socio-technical system (e.g. Salmon et al., 2012; Larsson et al., 2010), such an approach has its limitations in extensively reducing the number of fatalities and severe injuries.

There has been a shift in the strategic approach to road safety for the last ten years. The pace of this process has been slow but there is now a growing understanding in the road safety community that the strategies must be based on human factors principles and system theories.

Current approach to road safety in large parts of the world is “Vision Zero” or “Safe System”, two expressions of an identical policy. Recently, in the white paper on transport “Roadmap to a single European transport area —Towards a competitive and resource efficient transport system “ the European Commission has adopted Vision Zero, with the target that by 2050, the number of fatalities due to road traffic crashes should be close to zero. Also the guiding principles underlying the global Plan for the Decade of Action are those included in the Safe System approach. The forthcoming ISO 39001 management standard for traffic safety specifies that the standard is only relevant for organizations that wish to eliminate death or serious injury in road traffic crashes. OECD/ITF (2008) has recommended that the Safe System approach should be used to manage road safety. In the private sector, Volvo Cars has set a target of zero deaths and serious injury in or by a Volvo car 2020. Other car manufacturers have expressed zero as their vision, but not specified when this is supposed to be fulfilled. All these examples have one thing in common, except from explicitly aiming for elimination of death as a result of road traffic crashes, and that is the human factors perspective.

2 Human Factors

Going through the literature it is hard to find a clear, simple and commonly agreed definition of the human factors concept.

According to the Swedish Human Factors Network (HFN) human factors is:

“the scientific discipline that investigates and produces knowledge concerning human physical, cognitive and psychological prerequisites in relation to the character of the task, technical requirements with respect to design, complexity, and organizational premises regarding resource allocation, organizational culture, methods, competence and leadership as well as following up and evaluating systems.

The utilization of the knowledge about human capabilities and needs in design, implementation, deployment, operation and maintenance of products, systems (of humans, machines and organizations) in order to optimize system functionality as well as human wellbeing, health and safety."

From this definition and other similar definitions it can easily be understood that human factors is a wide and rather complex scientific discipline embracing different subjects from ergonomic issues, e.g. Human Machine Interface (HMI), to safety culture and resilience in large organizations. It is however clear that the discipline has a common aim of managing human error by setting out from the physical, cognitive and psychological prerequisites and limitations of the human being in her interaction with other components of a complex socio-technical system in a social, organizational and often rule based context. The aim is hence to create optimal system conditions for the human being to be able to act as safely as possible in such a system by eradicating or reducing human error.

From the different definitions it is hard to derive some distinguishing and operationalized features to which the Safe System approach easily can be compared. But the following two central "axiom" or principles, can be derived from the definition.

2.1 Human Capability

The individual or person approach to safety management is still common in complex socio-technical systems e.g. occupational safety (Leveson, 2011). Leveson (2011) states that this approach is built on the assumption that most accidents are caused by operator error and rewarding ''correct" behavior and punishing ''incorrect" behavior will eliminate or reduce accidents significantly. According to Read et al. (in press) the individual approach views the person as another component of the system and recommendations are made for increasing the reliability of this component. Read et al. (in press) further states that little consideration is given to the context of behavior and its influence. Consequently the individual approach leads to proposals for countermeasures aimed at behavior change through education and enforcement that increase compliance with laws, internal regulations, routines etc.

The human factors discipline instead acknowledge the human frailty and clearly shows that human beings cannot physically, cognitively or psychologically always cope with the complex demands of socio-technical systems (e.g. Dekker, 2002). For that reason there is a need to understand the human capabilities in relation to the system and how to adapt the properties of the system to these capabilities.

2.2 Systems Approach

Acknowledging human frailty the human factors discipline treats human error as a systems failure, rather than solely an individual operator´s failure (Salmon et al., 2010). It considers the interactions between humans and between humans and technology within a system and the presence of system wide latent conditions and their role in shaping the context in which operators make errors. Human error is no longer seen as the primary cause of accidents. Instead it is considered as a consequence of latent failures created by decisions and actions within the broader organizational, social or political system in which processes or operations take place (e.g. government, local authorities, organizations/companies and their different management levels) (Salmon et al., 2010). At least in principle, the systems approach is now the dominant approach in most safety critical domains where it is often denoted Human Factors or MTO (Man, Technology and Organization).

Elaborating the concept of systems approach further, Leveson (2002) and Hollnagel (2004) mean that accidents can be seen as emergent phenomena. Accidents occur when components of a system interact with each other and these interactions are not possible to foreseen because of their complexity. According to Leveson (2002) systems theory provides the theoretical foundation for systems engineering, which views each system as an integrated whole even if it is composed of diverse individual and specialized components. A basic and important assumption of systems engineering is according to Leveson (2002) ‘‘that optimization of individual components or subsystems will not in general lead to a system optimum; in fact improvement of a particular sub-system may actually worsen the overall system performance because of complex, non-linear actions among the components”. This means e.g. that safety cannot be optimized through the optimization of the safety performance of the individual components and according to Leveson (2002) ‘‘attempts to improve long-term safety in complex systems by analyzing and changing individual components have often proven to be unsuccessful over the long-term”.

3 Basic Design Principles of the Safe System Approach

According to Langford (2009) Sweden’s *Vision Zero* and the Netherlands’ *Sustainable Safety* represent the longest established Safe System approaches. Langford (2009) describes that such approaches around the world have the following common key features:

- “they recognize that the human body has a limited tolerance of violent forces and when crash energies exceed this tolerance, death or serious injury will be a probable outcome;
- they accept that crashes will continue to occur, accident prevention efforts notwithstanding, given that humans make mistakes when using the road system;
- the challenge for any Safe System in the event of a crash is to ensure that no fatalities will occur (and that serious injuries will be reduced) for road users behaving appropriately; and
- this challenge can be best met by managing the road infrastructure, vehicles and speeds to reduce crash energies to levels that can be tolerated by the human body.”

Langford (2009) means that Safe System approaches clearly differ from traditional views of road users. One example is the Netherlands’ *Sustainable Safety* which considers the road user the weakest link in the transport chain. According to Langford (2009) the individual road user “is largely unpredictable and cannot be relied upon to behave safely over the long term, all of his or her best intentions notwithstanding. People make mistakes. Training, education and even enforcement measures which rely upon correcting road user behavior will not succeed in achieving Safe Systems' ambitious goals."

The basic principles of the Safe System approach is further elaborated and summarized by the following design principles.

- *The design of the road transport system should guide the road user to a safe behaviour and mitigate the consequences of common human errors.*

The overarching concept of a safe transport system contains two imperatives, known for thousands of years. The first is that "it is human to err" (errare humanum est) meaning in this particular case that the human can never be trusted to repeatedly perform correct in all traffic situations, even if the intention is to maneuver in a safe manner. Hence the capabilities and limitations of the human being must to a great extent be taken into consideration when designing the road transport system. Road users will always make errors and mistakes for various reasons. These errors and mistakes in many cases originate from the interaction between the road user and the complex social, organisational and technical context in which the behaviour of the road users take place. They hence may be reduced by understanding these interactions and designing the road transport system from these conditions in order to guide the road user to an as safe as possible behaviour.

The other imperative, today well and thoroughly verified, is the role of kinetic energy in case of a human error. Hippocrates wrote around 400 B.C (Adams (ed.), 1886):

"Of those who are wounded in the parts about the bone, or in the bone itself, by a fall, he who falls from a very high place upon a very hard and blunt object is in most danger of sustaining a fracture and contusion of the bone, and of having it depressed from its natural position; whereas he that falls upon more level ground, and upon a softer object, is likely to suffer less injury in the bone, or it may not be injured at all."

What is behind Hippocrates sentences is simply the strong relationship between the speed/energy and the object that finally stops us in case of a human error leading to a crash, and how such errors can be counteracted with lower speed and/or substituting or modifying surfaces that we hit. A high risk of human error can therefore be matched by reduced kinetic energy or less harmful contact surfaces.

Since human errors and mistakes cannot fully be eradicated, the infrastructure components and vehicles of the road transport system must be designed to mitigate the consequences of common human errors and mistakes. While this may be clear and logic, the road transport system has not been designed from ground with the aim to absorb or mitigate common human error or to absorb the consequences of it.

The Safe System approach combines the two imperatives with back casting. If zero deaths and serious injuries are to be achieved, how do we combine human error with human biomechanical tolerances by minimizing human error, but when it occur to make sure that the human biomechanical tolerance is not exceeded? In doing so, it is simply necessary to develop design principles for system safety, and not simply treat each component individually. The characteristics of the road user, the vehicles, the road design and the speeds on a road all have to work together to achieve safety.

- *The setting of speed limits must be in accordance to the safety standard of the infrastructure and the type of vehicle in such way that normal and common human errors and mistakes can be managed as to eliminate the risk of serious injuries.*

The preconditions for designing a safe road transport system are twofold; the biomechanical tolerance to mechanical force and the possible crash scenarios that can be foreseen. In working out possible scenarios, the human behavior is the key for understanding what might lead to a crash with energy enough to harm the human. A high risk of human error can be matched by reduced kinetic energy or less harmful contact surfaces. The balancing act is to maintain accessibility and mobility of the road transport system, but limitations in safety should be counteracted by reduced kinetic energy, which in most cases mean reduce speed. The alternative to reduce speed is an investment into the system that leads to maintained or even increased speed. This is why progressing in safety in the end is an investment in mobility.

- *New rules and regulations with the purpose to change human behavior must be developed from a human factors perspective taking into account the limitations and capabilities of the human being.*

There is still a fairly widespread belief that accidents are caused by human errors and that these could be significantly reduced by introducing additional regulations and procedures to ensure a "correct" behaviour and punish an "improper" behaviour of those who "violate" the rules. This approach presupposes that human errors are more or less intentional violations, i.e. that the road user in all situations can make a deliberate or conscious decision to act right or wrong. Running a red light or trying to cross an intersection despite there is conflicting traffic are typical examples of serious traffic offences that might have no intention behind. Forgetting to put on the seat belt, not turning on headlamps, losing control on a road with invisible ice are other such examples of violating the traffic rules with no real intention behind, but possibly leading to lethal consequences.

On a general level human error in road traffic hence can be divided into unintentional errors (mistakes, slips, lapses etc.) and intentional violations. Contemporary Human Factors research clearly shows that regulating human behaviour and making the individual accountable for accidents will only have marginal effect on unintentional errors (e.g. Dekker, 2002). In depth analyses of road traffic accidents show that such errors are common contributing factors.

When it comes to intentional violations Svensson (2008) shows that such regulating activities will have an effect but it varies considerably with the risk of being caught and the level of sanctions.

Speeding, driving under influence of alcohol or other drugs, not using restraint systems or not using protective equipment are in many cases serious intentional violations but in some cases unintentional errors (especially when it comes to speeding, not using restraint systems and protective equipment). These violations and errors may lower the effects of the system design and must be met with special attention.

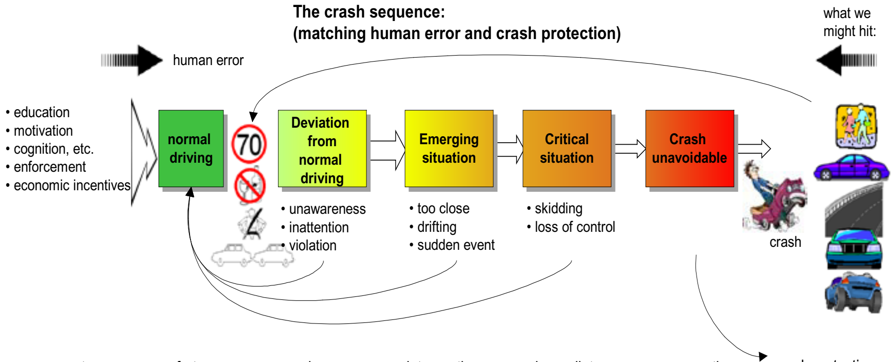
- *Design solutions of the road transport system and rules and regulations with the purpose to change human behavior must be evidence based and based on an integrated systems approach.*

The development of road safety should and must be based on scientific evidence and best evidence from experience. This should apply to all stages of the development, from target setting and management to detailed design solutions and regulating efforts to diminish or eradicate trauma. In doing so, it is necessary to integrate safety solutions to all factors of an accident and injury prevention process. This is a general trend in the automotive sector since a few years back, but needs to be broadened to the entire road transport system. To seriously reduce e.g. pedestrian casualties, road user rules and behaviour, road and traffic environment, speed management, systems to brake a car automatically and "pedestrian friendly" front ends cars must be combined in an optimised way. In isolation, the effect of each component may have some effect, but as they give each other preconditions to maximise benefit, the whole combination might give more effect than the sum of each component.

The possibility to build a safe road traffic system on an overall level is not complicated. The aim of such a system is to make sure that the biomechanical tolerance for a fatal or serious injury is not exceeded. Setting out from the so called integrated safety chain in fig. 1 the system should be designed backwards from a possible event where injury might occur. The challenge is to prevent the hazardous event at as early as possible stages of the integrated safety chain but simultaneously acknowledge that this is not always possible and hence take measures to limit the amount of energy that might be exchanged in case of a crash. At a later stage in the chain, safety systems that can limit the risk of an injury given a crash can be activated. The key is to link the possible outcome with the speed that can be tolerated under normal driving conditions so that the amount of kinetic energy is not larger than what can be managed through the chain. The real challenge is both to combine the links of the chain to what is most effective, and to evaluate the combined effects.

For each step in the chain, a systems approach must be taken in order to understand and integrate road user rules and behaviour, road and traffic environment, speed management, vehicle systems with the aim to combine them in an optimised way.

Even if harmful events cannot be prevented and returned to an earlier stage of the chain they can be transformed in order to make it easier to mitigate their consequences in following stages by e.g. a more favourable crash configuration. One example is Electronic Stability Systems (ESC). Even if a vehicle comes over into the oncoming traffic lane, the risk of doing it with the side hitting oncoming traffic resulting in an unfavourable crash configuration is minimized. It is important to understand that efforts should be put in all parts of the chain but it is not possible to always rely on solving the problems in earlier stages of the chain. There will always be events that "falls through" the chain and the only way is then to mitigate the consequences.



• access to road transport system	• comfort • economy • social conformity	• warning system • supporting system	• intervention in driving	• immediate correction	• preparation for crash	
Vehicle	promote normal driving	(ISA, SBR, alcohol interlock)	(AICC, LDW)	(ESC, LDA, AICC2)	(pre-safe, emergency braking)	(seat belts, airbag, whiplash protection, pedestrian protection)
Infrastructure	promote normal driving	(speed warning, tactile warning, humps)	tactile edge lines	high friction surface		barrier design, roundabouts
Others	promote normal driving	• enforcement • insurance • contracts				• emergency service

Fig. 1. The integrated safety chain (Tingvall et al., 2009)

4 Discussion

From the design principles it seems clear that the Safe System approach share many features of the view of human capability and the systems approach with the human factors concept.

The maybe most obvious common feature of the two concepts is that they both acknowledge that human beings cannot physically, cognitively or psychologically always cope with the complex demands of socio-technical systems, like the road transport system. Instead the human factors discipline and the Safe System approach urge us to understand that when the human is a part of a complex system he/she cannot always act safely or perfectly according to the rules even if his/her intentions are correct. Such systems must consequently be designed and operated with the capabilities and limitations of human beings as a starting point.

Another important common feature, which is closely related to the previous feature, is that accidents can be seen as emergent phenomena (Leveson, 2002 and Hollnagel, (2004) and that optimization of individual components for that reason will not in general lead to a system optimum. This is an important conclusion since it states that safety cannot be optimized by e.g. trying to optimize the safety performance and behavior of the human being or any other component of a complex system. The Safe System approach consequently emphasizes the importance of understanding and integrating road user rules and behaviour, road and traffic environment, speed management, vehicle systems with the aim to combine them in an optimised way.

However it can be argued that the Safe System approach takes the concept further creating a more holistic systems approach to road traffic accidents.

According to Read et al. (in press) the field of human factors has traditionally focused on the physical and cognitive capabilities and limitations of humans aiming at the understanding of human error and how it can be managed or controlled. Read et al. (in press) however claims that modern human factors approaches are moving away from the psychological approach that considers humans as limited information processors. Still human capabilities and limitations are important to understand but a greater focus must be put on the context of behavior and the constraints imposed by the environment.

From these statements and the previous mentioned definition by HFN it can be argued that the human factors discipline seems to be mainly focusing on managing and controlling human error in order to avoid accidents. It seems that the main focus of human factors is the understanding of behavioral aspects in relation to the system context.

The Safe System approach also embraces the management and control of human error. However it also focuses on mitigating the consequences of such errors. This dual view on human error is illustrated by the integrated safety chain (fig. 1).

The integrated safety chain share many common features with nuclear safety and the principle of "defense in depth". According to IAEA (1996):

> "Defense in depth is generally structured in five levels. Should one level fail, the subsequent level comes into play. The objective of the first level of protection is the prevention of abnormal operation and system failures. If the first level fails, abnormal operation is controlled or failures are detected by the second level of protection. Should the second level fail, the third level ensures that safety functions are further performed by activating specific safety systems and other safety features. Should the third level fail, the fourth level limits accident progression through accident management, so as to prevent or mitigate severe accident conditions with external releases of radioactive materials. The last objective (fifth level of protection) is the mitigation of the radiological consequences of significant external releases through the off-site emergency response."

But there seems also to be an important difference between the integrated safety chain and "defense in depth". The outcome of each step of the integrated safety chain, once it did not stop the event, is an input to the next step. Therefore, each step must have to design properties, either to stop the chain or prepare for the next phase. It should though be noted that the principle should be not to let many cases run to far in the chain of events and run the risk to be stopped by the last barrier.

From the literature it is quite unclear if the Human Factors discipline embraces these aspects of system perspective regarding the physical frailty of the human being and the way it can be mitigated even if e.g. Hollnagel (2004) is considering material or physical barrier systems important for the blocking or mitigation of an unexpected event. This may be due to the fact that in systems where the concept of human factors and system safety have evolved in safety critical systems where the levels of harmful energy have not been possible to mitigate in an effective and economic way.

However it seems that the last step in the integrated safety chain, preventing or mitigating the consequences of human error, is omitted or at least not clearly elucidated in the human factors approach. Thus the Safe System approach takes the view of human capability and systems approach a step further compared to the human factors approach by adding the biomechanical tolerance and its implications on systems design.

References

1. Adams, F. (ed.): The Genuine Works of Hippocrates. William Wood and Company, New York (1886)
2. Dekker, S.W.A.: The Field Guide to Human Error Investigations. Ashgate Publishing Limited, Hampshire (2002)
3. Hollnagel, E.: Barriers and Accident Prevention. Ashgate Publishing Limited, Hampshire (2004)
4. IAEA.: Defence in Depth in Nuclear Safety, INSAG-10. International Atomic Energy Agency, Austria (1996)
5. Langford, J.: Towards Zero: Understanding a Safe System Approach to Road Safety. Fact Sheet No. 1. Curtin-Monash Accident Research Centre, School of Public Health, Faculty of Health Sciences, Curtin University of Technology, Perth, Western Australia (2009)
6. Larsson, P., Dekker, S.W.A., Tingvall, C.: The need for a systems approach to road safety. Safety Science 48, 1167–1174 (2010)
7. Leveson, N.G.: System Safety Engineering: Back To The Future (2002), http://sunnyday.mit.edu/book2.pdf (retrieved October 18, 2006)
8. OECD/ITF.: Towards Zero: Ambitious Road Safety Targets and the Safe System approach. Summary document. OECD Publishing (2008)
9. Read, G.J.M., Salmon, P.M., Lenné, M.G.: Sounding the warning bells: The need for a systems approach to understanding behaviour at rail level crossings. Manuscript submitted for publication. Applied Ergonomics (in press)
10. Salmon, P.M., Lenné, M.G., Stanton, N.A., Jenkins, D.P., Walker, G.H.: Managing error on the open road: The contribution of human error models and methods. Safety Science 48, 1225–1235 (2010)
11. Svensson, M.: Sociala normer och regelefterlevnad – Trafiksäkerhetsfrågor ur ett rättssociologiskt perspektiv. Doctoral thesis, Lunds University, Lund, Sweden (2008)
12. Tingvall, C., Eckstein, L., Hammer, M.: Government and Industry Perspectives on Driver Distraction. In: Christie, K., Lee, D., Regan, M. (eds.) Driver Distraction. Theory, Effects and Mitigation. CRC Press (2009)

Actualising a Safe Transport System through a Human Factors Systems Approach

Michael G. Lenné[1], Paul M. Salmon[2], Neville A. Stanton[3], and Elizabeth Grey[4]

[1] Human Factors Group, Monash University Accident Research Centre, Monash University, Victoria, Australia
[2] University of the Sunshine Coast Accident Research, University of the Sunshine Coast, Maroochydore, Queensland, Australia
[3] Transportation Research Group, School of Civil Engineering and the Environment, University of Southampton, Highfield, Southampton, UK
[4] Transport Safety Victoria, Melbourne, Australia
michael.lenne@monash.edu, psalmon@usc.edu.au, N.Stanton@soton.ac.uk, Elizabeth.Grey@transportsafety.vic.gov.au

Abstract. Safe system strategies govern the approaches to road safety in many countries. This is the case for both road and rail safety in Australia. In this paper we take a complex segment of the road and rail system, rail level crossings, to demonstrate why the current approaches to safety in this area need to change. We argue that approaches that are more consistent with real systems thinking are required to generate the new interventions needed to reduce road trauma in this setting. In recognizing the need for new approaches the Victorian road and rail sponsors have partnered with Australian and UK Universities in an exciting four year initiative designed to change the paradigm in RLX safety. In this paper we outline the rationale for this change and describe the four phase analytical approach being used. It is hoped that this approach will help to actualise safe system strategies in ways that are more consistent with systems thinking and that significantly improve safety.

Keywords: Transport safety, systems approach, human factors, rail level crossings.

1 Introduction

In efforts to increase road safety many jurisdictions have followed the early leads by the Swedish and Dutch governments in introducing safe system approaches to road safety. This is certainly the case in Australia where both state and national strategies advocate for a safe system. Yet the approaches within these strategies typically take a reductionist approach where safety issues are considered in isolation. The key to successful implementation of such an approach lies in the consideration of human performance in the context of the wider system in which it takes place, and the acquisition of appropriate evidence to support systems-based strategy development. While these approaches have led to safety improvements there is now a growing body of

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 29–35, 2013.
© Springer-Verlag Berlin Heidelberg 2013

literature that illustrates that more systematic approaches to safety management can produce more significant and long-term benefits. Systems theory has underpinned advances in complex sociotechnical systems [1] and led to the latest drift into failure model of system safety [2]. While road safety is such a system, the important features of systems theory are not found in the current road safety approach. There have been calls for these human factors-driven approaches to be applied to road safety [3].

Given this, what does it mean to implement road safety interventions that are truly safe system compliant? Does it mean introducing measures that are consistent with systems theory? Or does it mean implementing measures that achieve a prescribed safe system standard but yet continue to address road safety issues in isolation? Current road safety practice would suggest the latter. Road safety needs to go through a paradigm shift, inspired by systems approaches, to truly achieve a safe system. How can this paradigm shift be realised, and what is the conceptual framework that can take us there?

This paper considers these questions and the safe system concept in a complex part of the transport system, rail level crossings (RLX; highway-rail grade crossings in US terminology). This is a particularly interesting area of application as it features in both rail and road strategies in Australia. In recognizing the need for new approaches to safety in the RLX context, the Victorian road and rail sponsors have partnered with Australian and UK Universities in an exciting four year initiative designed to change the paradigm in RLX safety. Specifically, the program aims to:

1. Develop a systems-based model of RLX operation, informed by data collection and analysis of established theories, methods and models;
2. Use the model outputs to specify the optimum functionality and characteristics of interventions to prevent RLX crashes, and to prioritise existing interventions for testing and refinement; and
3. Evaluate driver responses to novel interventions, using advanced driving simulation, and propose effective and cost-efficient solutions to shape policy and standards to reduce RLX crashes.

This paper discusses the overarching framework we have adopted to achieve these aims, while specific examples within each analytical phase will be provided in the conference presentation. In the paper we firstly illustrate how existing approaches in this area are having modest success. Secondly, we introduce the paradigm shift needed, namely the systems approach. Thirdly, we present an approach to actualising the systems approach.

2 The Current Approach to RLX Safety

In 2008 there were 58 collisions between trains and vehicles at level crossings in Australia, which led to 33 fatalities and serious injuries [4]. Such incidents typically involve road user errors and violations, traumatic injury, and have a significant economic impact on both networks. This is particularly so for heavy vehicle collisions as they have a much greater potential to derail the train.

Our recent review of RLX research highlighted that achieving acceptable levels of performance and safety at RLXs has proven difficult [5], partly because RLXs are not homogeneous. RLXs are typically classified as one of two types: RLXs with active warnings (e.g., flashing red lights), or passive crossings (protected by stop or give way signs). Further, there are differences in the volume of rail and road traffic, the type and speed of traffic, overall RLX geometry, and so on. All of these factors influence fundamental aspects of human performance (including perceptual processes and expectations) that shape road user behaviour and thus the appropriate solution. Across Australia there are approximately 9,400 rail level crossings, with 6,060 passive (60%), 2,650 (30%) active, and 690 (10%) having other forms of control. Current solutions to the problem, such as grade separation and installation of boom gates, provide significant safety improvements but are cost-prohibitive (Wigglesworth & Uber, 1991). The effectiveness of lower cost interventions, such as education campaigns, speed limit reductions, rumble strips, train strobe lightings and in-vehicle warnings remains largely unknown, with the evaluations conducted to date being poorly designed and lacking a sound theoretical underpinning [5].

Key to developing effective RLX crash interventions is an in-depth understanding of the RLX system, including the performance of, and interactions between, its component parts (road users, vehicles, trains, train drivers, infrastructure etc). Although a limited number of models have been developed, currently we do not possess this systemic understanding. The research to date has been driven by an individual road user viewpoint, and therefore does not fully consider the wider RLX system factors that shape road user performance.

Our understanding of RLX system operation and road user behaviour at RLXs is therefore currently limited [5]. The focus of existing research on individual factors is a critical shortfall given the recent theoretical advances within the discipline of Human Factors that emphasise the need to take a systems perspective when evaluating, modelling, and supporting the performance of complex sociotechnical systems. The need to take the entire system, comprising human operators, tools, artefacts, and technologies and the interactions between them as the unit of analysis rather than the individuals working within it has been advocated [7, 8]. While existing theories of human performance (e.g., information processing) and models of driver behaviour provide a solid foundation, a new approach is needed to drive intervention development that recognises broader systemic influences on RLX crashes. Our recent review of level crossing intervention research identified the lack of a systemic model that describes how road users interact with level crossing infrastructure as a major gap in the area [5]. The traditional approach to this issue, while important, has not taken us forward in terms of improving safety for some time as these interventions for road user behaviour and level crossing safety have not been assessed from a true systems perspective.

3 The Human Factors Approach

The approach we are adopting to reducing RLX trauma involves collection of data to better understand the nature and performance of different RLX systems, and then development of models of RLX system performance using contemporary methods. This involves the use of theoretically underpinned, systems-based methods, including Hierarchical Task Analysis (HTA) [9] and Cognitive Work Analysis (CWA) [10, 11]. Following this, the analyses are used to evaluate existing interventions, and to inform the design and specification of novel interventions designed to treat the problem of RLX crashes. The final phase will use advanced driving simulation and on-road methods to test and refine the interventions proposed. An overview of the process is provided in Figure 1.

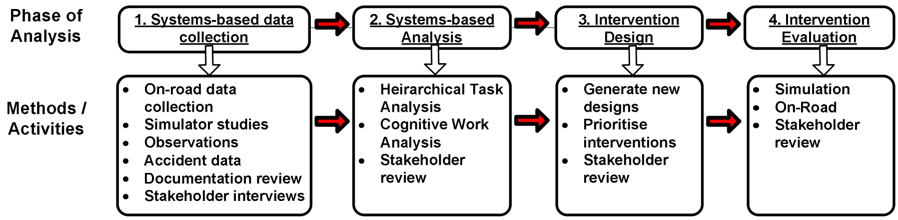


Fig. 1. An overview of the human factors-based systems approach to safety research and intervention design

3.1 Data Collection and Model Development

The application of HTA and CWA in a complementary manner has previously been used for system design and evaluation in other areas such as process control and the military. The outputs from each approach describe the system in a different but complementary manner, which is particularly powerful for system design and evaluation. For example, HTA describes the system normatively, in terms of what currently happens, whereas CWA describes the system formatively, in terms of what could potentially happen. The use of both methods together in an integrated manner, although used in the past in other complex safety critical domains, is novel with regard to road and rail safety.

The development of both models (HTA and CWA) is supported initially by data collected from a range of activities, including observational studies, documentation review, subject-matter interviews, and walkthrough analyses, all of which have previously been used to support previous model development using HTA and CWA [10, 12]. The use of driving simulation and on-road methods to study road user behavior also provides novel insights into the role of the RLX system in shaping behavior.

In this regard we have already conducted two studies using instrumented vehicles to study driver behavior at RLX in both regional and metropolitan settings. In addition to the standard suite of vehicle-based and eye-movement measures [13], measures of driver cognitive process and strategies were also derived via the use of verbal protocol analysis (i.e. think aloud) during the drive [14], and then post drive 'critical decision method' interviews, which are designed to explore the cognitive processes underpinning task performance and decision making. Data of this type have not previously been collected at RLXs and are required to underpin both the development of the RLX models and future research efforts.

3.2 The Systems-Based Methods of Analysis

The data collected are being used to construct HTA descriptions of different RLXs systems (e.g. passive versus active, metro versus rural). The outputs provide goal-based models of RLX system operation and will be used to inform our understanding of RLX system operation. They are also used for evaluating existing interventions and for generating and refining novel intervention designs. Additional human factors analyses modules applied to the HTA descriptions, including interface design and evaluation and error identification will be particularly useful for the intervention development and evaluation phases of the research.

The second analytical component is provided by the Cognitive Work Analysis framework which is currently receiving great attention as a comprehensive complex system evaluation and design approach. CWA is an ecological interface design-based framework that is concerned with constraints (rather than goals), which is based on the notion that making system constraints explicit in interfaces and displays potentially enhances human performance. The underlying premise of CWA is that one cannot understand cognition without first understanding the nature of the work domain. Further, CWA is formative and so, rather than normatively prescribe how work should be done or describe how it is currently being done, it seeks to identify how work could be done if the appropriate tools were made available, which is particularly important for system design efforts.

We have completed the CWA for active and passive RLX. Examples of how we worked through these analyses, and their outputs, will be provided in the conference presentation.

3.3 Design of Interventions

The CWA approach is used to generate new system designs, in terms of tools, interfaces, task allocation, and social organisation. This step will involve the use of the CWA outputs in conjunction with Human Factors, road and rail SMEs, to generate new RLX crash intervention design specifications. Specifically, a workshop will be held with researchers, SMEs and project partners, whereby the CWA outputs will be used to drive novel intervention design specification. Whereas CWA has in the past been used as a revolutionary design approach, the utility of HTA lies in its use as an evolutionary design approach. This step will involve the use of HTA and its range of extended analysis modules (e.g., error identification, interface design) to refine the interventions identified.

The output of these steps will be a series of candidate intervention designs for further testing. Importantly, at this stage the interventions chosen for scientific evaluation during phase 3 of the project will have been selected and refined based on exhaustive, systems-based models of RLX performance and expert opinion.

3.4 Evaluation of Interventions

Simulation is an excellent and established method to test any number of system-based interventions that might emerge during earlier phases of the research.

In preliminary research we have established that controlled and safe exposure to rare events (encounters with trains) can be achieved using simulation [15]. The class of performance measures that can be used include the following:

1. Vehicle-based measures including: mean and standard deviation of lateral position, mean speed and speed profiles on approach to RLXs and reaction time.
2. Compliance with RLXs as measured by proportion of drivers who come to a stop when a train approaches.
3. Driver eye movements, including object detection times and fixation durations to be measured. This provides for more theoretical hypothesis testing related to attentional strategies adopted by road users.
4. Subjective data collected via post-drive interviews regarding driver perceptions of the interventions trialled and understanding of required behaviours at RLXs.

Following simulator evaluation, selected interventions can be implemented in the field, after which point the on-road methods of assessment are appropriate.

4 Conclusion

Approaches to safety research and management that take a systems approach are more likely to generate effective interventions. Several authors have called for such approaches to be applied to transport settings. In recognizing the need for new approaches to safety in the RLX context, the Victorian road and rail sponsors have partnered with Australian and UK Universities in an exciting four year initiative designed to change the paradigm in RLX safety. The approach we are using initially involves the collection of new data on the ways in which the RLX system shapes road user behviour. The second component involves a conceptualisation and analysis of the RLX system using methods that are congruent with systems thinking. The use of these outputs to support new intervention design and evaluation are the final components.

To this point we have completed the first phase identified in Figure 1, the data collection. In doing so we have provided new data on the factors that shape road user behavior at RLX. The phase 2 analyses using CWA are almost complete and provide a very different and novel means for conceptualizing and representing the RLX system. It is our hope and intention that the analytical process we are adopting will serve a model that can guide others in actualizing a safe road and rail transport systems.

Acknowledgements. This project is funded by an Australian Research Council Linkage Grant (LP100200387) to Monash University in partnership with the Victorian Rail Track Corporation, Transport Safety Victoria, Public Transport Victoria, Transport Accident Commission, Roads Corporation (VicRoads) and V/Line Passenger Pty Ltd. We gratefully acknowledge the support of the project partners and community participants. Paul Salmon's contribution to this research and article is funded through the Australian National Health and Medical Research Council post-doctoral training fellowship scheme.

References

1. Larsson, P., Dekker, S.W.A., Tingvall, C.: The need for a systems theory approach to road safety. Safety Science 48, 1167–1174 (2010)
2. Dekker, S.W.A.: Drift into failure: From hunting broken components to understanding-complex systems. Ashgate Publishing Co., Farnham (2011)
3. Salmon, P.M., McClure, R., Stanton, N.A.: Road transport in drift: Applying contemporary systems thinking to road safety. Safety Science 50, 1829–1838 (2012)
4. Australian Transport Safety Bureau.: Australian rail safety occurrence data, January 2001 to December 2008. Canberra, Australia: Australian Transport Safety Bureau (2009)
5. Edquist, J., Stephan, K.L., Wigglesworth, E., Lenné, M.G.: A literature review of human factors safety issues at Australian level crossings: Monash University Accident Research Centre for Department of Transport, Victoria (2009)
6. Wigglesworth, E.C., Uber, C.B.: An evaluation of the railway level crossing boom barrier program in Victoria, Australia. Journal of Safety Research 22, 133–140 (1991)
7. Hollnagel, E.: Extended cognition and the future of ergonomics. Theoretical Issues in Ergonomics Science 2, 309–315 (2001)
8. Walker, G.H., Stanton, N.A., Salmon, P.M., Jenkins, D.P.: Sociotechnical systems. Ashgate, Aldershot (2009)
9. Stanton, N.A.: Hierarchical task analysis: Developments, applications, and extensions. Applied Ergonomics 37, 55–79 (2006)
10. Jenkins, D.P., Stanton, N.A., Walker, G.H., Salmon, P.M.: Cognitive work analysis: coping with complexity. Ashgate, Aldershot (2008)
11. Vicente, K.J.: Cognitive work analysis: Toward safe, productive, and healthy computer-based work. Lawrence Erlbaum Associates, Mahwah (1999)
12. Stanton, N.A., Salmon, P.M., Walker, G.H., Jenkins, D.P.: Human factors and the design and evaluation of central control room operations. Taylor & Francis, Boca Raton (2009)
13. Lenné, M.G., Beanland, V., Salmon, P.M., Filtness, A.J., Stanton, N.A.: Checking for trains: an on-road study of what drivers actually do at level crossings. In: Proceedings of the Fourth International Rail Human Factors Conference, London, UK, March 5-7 (2013)
14. Salmon, P.M., Beanland, V., Lenné, M.G., Filtness, A.J., Stanton, N.A.: Waiting for warning: driver situation awareness at rural rail level crossings. In: Proceeding of the Institute of Ergonomics and Human Factors Annual Meeting, Cambridge, UK, April 15-18 (2013)
15. Lenné, M.G., Rudin-Brown, C.M., Navarro, J., Edquist, J., Trotter, M., Tomasevic, N.: Driver behaviour at rail level crossings: Responses to flashing lights, traffic signals and stop signs in simulated rural driving. Applied Ergonomics (Special Issue on Transportation Safety) 42, 548–554 (2011)

Combined Effect on Accident Risk of a Dual Task and Higher Driving Speed: A Simulator Study

Evangelia Portouli[1], Vassilis Papakostopoulos[2], and Dimitris Nathanael[3]

[1] National Technical University of Athens, School of Mechanical Engineering, Sector of Industrial Management and Operations Research, Ergonomics Unit, Iroon Politechneiou 9, 15773 Zografou, Greece
portouli@mail.ntua.gr
[2] University of the Aegean, Department of Product and Systems Design Engineering Ermoupolis, Syros GR 84100, Greece
papakostopoulos@aegean.gr
[3] National Technical University of Athens, School of Mechanical Engineering, Sector of Industrial Management and Operations Research, Ergonomics Unit, Iroon Politechneiou 9, 15773 Zografou, Greece
dnathan@central.ntua.gr

Abstract. A study was conducted on a dynamic driving simulator aiming to examine whether the effect of mental effort due to an auditory detection task on accident risk is additive to the effect of higher speed. Two levels of the driving task were employed, a low-demanding and a high-demanding one. Twenty drivers were asked to drive two rounds on a rural road with normal traffic, with unexpected traffic events along the second round. In half of the cases an auditory detection task had to be performed in parallel. The analysis of results showed that higher speed or higher mental effort due to the secondary task lead to more accidents and the effects should be considered as additive. These effects should not be considered as the mere effect of attentional resource availability but as depending on the drivers' skill to manage their attentional control.

Keywords: accident risk, driving simulator, mental effort, secondary task.

1 Introduction

Several studies report that an increase in average speed results to an increase in the risk of an accident of any type [1,2]. Other driving studies have focused on the effect of mental workload on driving performance [3,4] often in an attempt to evaluate potential interference of specific in-vehicle systems into the primary driving task. Studies relevant to mobile phones report a decrement in some aspects of driving performance when a mobile phone is used –e.g. in visual search behavior [5,6], braking reaction time [7,8], visual processing and decision making [9-11] that vary according to the level of difficulty or intensity of the primary task. Studies of using visual route guidance systems and performing a secondary auditory task while driving [12,13] report that drivers tend to reduce their driving speed as well as to neglect subsidiary

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 36–42, 2013.
© Springer-Verlag Berlin Heidelberg 2013

tasks (e.g. mirror-checking) when task demand increases but this does not imply that accident risk is equally reduced. However, the linkage that has been made so far between the relation of mental workload and accident risk is rather weak. It was reported [14] that the speed reduction while looking at a visual display can be interpreted as an indication of drivers' inattention to the primary task. On the other hand, a simulator study [15] reported that a primary task of lower demand does not improve the performance on the secondary task.

If task performance was solely dependent on attentional resource availability, then the lower the demand of the primary task, the better the performance on the secondary task should be. Still according to the attentional management theory [16] - describing dual task performance as a skill of attention management -drivers can achieve a high level of performance when performing a number of tasks concurrently because they prioritise their tasks with respect to the main task goal (namely to arrive safely at the destination), in such a way that it is possible to manage to attend different tasks (e.g. by expecting when an event will take place).

The objective of this experiment, performed within the framework of AIDE Integrated Project, was to study whether the effects on accident risk of increased speed and increased mental workload due to the parallel conduct of a secondary task should be considered as additive or interacting.

2 Method

2.1 Participants

20 persons, 13 males and 7 females, participated in this experiment. Participants were recruited via announcements at the announcement board of the Hellenic Institute of Transport. Participants' age ranged from 24-40 years old (mean: 30.8, standard deviation: 4.06 years). They held a driving licence for an average duration of 10.75 years (standard deviation 4.31 years).

2.2 Apparatus

The experiment was performed on the dynamic driving simulator of the Hellenic Institute of Transport, built around a Smart cabin equipped with sensors. The position of all control levers, windshield wipers, blinker, ignition key and light switch is transmitted to the driving computer. All operational elements, steering wheel, accelerator pedal, brake pedal, gearshift lever and handbrake lever, provide nature-true force reactions. The gearshift functions like in the real car either as automatic or "softtip" with incrementing and decrementing the six gears and with reverse gear. The sight system includes five large-screens, each having a width of 2 m. There is on-screen projection with consumer video projectors with 2500 ANSI-lumen. The sound system generates original sounds according to the situation (starter, engine noise, horn, screeching of tires, drive wind, rain, etc.). The vibration device creates natural true vibrations of the car according to the revolvation of the simulated engine.

The secondary task employed was an auditory detection task with beeps randomly generated every 3 to 5 s, each one lasting for 1 s or until the participant had pressed the response button. Signal intensity was adjusted for each participant so that it was easily perceptible. Participants were asked to push the response button, as soon as a signal was noticed by them.

2.3 Experimental Design

Participants were allocated in two groups and two conditions. In the No pressure group, participants were asked to drive normally, and in the Pressure group, participants were given a short time limit within which they should have completed the simulator scenario, so as to urge them to drive at higher speeds. The two conditions were, Driving only (DO) and Driving while conducting in parallel an auditory detection task (DT). The order of presentation of conditions was counter-balanced.

The simulator scenario was built using a circuit rural route with one lane per direction, no central border and a total length of 6.2 km. There were 2 signalised and 1 non-signalised intersection along the route. Each participant was asked to drive two rounds of the circuit continuously. There were oncoming vehicles and lead vehicles in front of the ego vehicle, some of them driving at low speed. Overtaking in general was possible but risky.

2.4 Procedure

Upon arrival, participants were completing a background questionnaire with personal data and then they were asked to drive for 5 minutes the driving simulator in free traffic, so as to get acquainted with it. Each subject then had to drive the whole simulator scenario, consisting of two rounds. At the end participants were asked to complete a mental effort questionnaire.

During the first round, the driving behaviour of the other road users was in compliance with traffic rules, whereas no other unexpected traffic events took place. During the second round, the following unexpected traffic events were scheduled: a sudden deceleration of the lead vehicle (twice), an animal suddenly crossing the road, a parked car sudden entrance into the lane in front of the ego vehicle, the door of a parked car suddenly opening in front of the ego vehicle.

2.5 Measures and Analysis Method

The simulator logged time, number of accidents, ego vehicle speed, distance to lead car, time when the participant started braking or initiated an evasive manoeuvre after the occurrence of a critical event. An evasive manoeuvre was defined as a sudden change in the steering angle, resulting in a change in the ego vehicle lateral position towards the edge of the road. The following indicators were calculated:

- Response time to a critical event (s). This was the time that the participant initiated a braking or evasive manoeuvre in response to one of the critical events minus the time that the critical event was initiated.

- Headway to lead car at the moment when the participant started braking or initiated the evasive manoeuvre in case of the lead vehicle sudden braking event.

As regards the performance in the auditory task the following indicators were calculated: the mean response time in s and the mean hit rate, the latter being the correct hits within the specified time limit per total beeps emitted.

At the end of each scenario, participants were asked to rate their mental effort, using the Rating Scale of Mental Effort (RSME) [17].

ANOVAs and t tests were used for statistical analysis of data.

3 Results

3.1 Participants' Performance on Primary and Secondary Task

There was an effect of Pressure on ego vehicle speed, on lane keeping and on the headway at which the participants initiated the evasive manoeuvre or started braking in the case of the second lead vehicle braking in both the DO and DT conditions. No effect of Pressure was found for the headway at which the participants initiated the evasive manoeuvre or started braking in the case of the first lead vehicle braking, on RSME ratings or on the performance on detection task (hit rate and response time).

There was an effect of Detection task on RSME ratings but on no other indicator.

No interaction effects were found between Pressure and Detection task.

Table 1. Participants‘ performance per condition and group

	No Pressure		Pressure		Effect pressure	Effect detection task
	DO	DT	DO	DT		
Mean ego vehicle speed (km/h)	65.5	61.3	86.6	85.4	< 0.05	ns
Headway start response 1st lead vehicle brake (s)	0.88	1.01	0.68	0.73	ns	ns
Headway start response 2nd lead vehicle brake (s)	1.29	3.13	0.77	0.78	< 0.05	ns
StDev lateral position (m)	0.63	0.59	0.76	0.77	< 0.05	ns
Mean hit rate		84.5%		83.7%	Ns	ns
Mean response time (s)		0.45		0.44	Ns	ns
RSME	38	65	47	59	Ns	< 0.05

3.2 Type of Accidents and Accident Risk

The total number of accidents for all participants was much higher in the Pressure group (19 accidents in condition DO and 17 in DT) than in the No pressure group (11 accidents in condition DO and 10 in condition DT).

In the No pressure group the number of accidents in the second run was always lower or equal to that of the first run while in the Pressure group the number of accidents was in some cases higher in the second run than in the first one, irrespectively of the order of presentation of the experimental conditions.

No effect of Pressure or of detection task was found on response time to critical events, except from the first case of sudden braking of the lead vehicle and this only between the No pressure-DO and the Pressure-DT conditions.

Table 2. Response time (s) for the various events among groups and conditions

	No Pressure		Pressure	
	DO	DT	DO	DT
Response time 1st lead vehicle brake (s)	0,70	0,52	0,42	0,18
Response time 2nd lead vehicle brake (s)	0,54	1,38	0,43	0,40
Response time animal (s)	1.44	1.39	1.52	1.40

3.3 Combined Effect on Number of Accidents

Speed, headway at which a braking or evasive manoeuvre was initiated, standard deviation of lateral position, RSME ratings, accuracy on detection task (*AccuDet*) and mean reaction time on detection task (*RTDet*) were normalised into a range 0 to 1 and the following regression model was used.

$$\text{Number of accidents} = \exp(\beta_o + \beta_1\,\text{Speed} + \beta_2\,\text{RSME} + \beta_3\,\text{Headway} + \beta 4\,\text{LatPos} + \beta_5\,\text{AccuDet} + \beta_6\,\text{RTDet}) \quad (1)$$

According to the data collected, the speed, RSME rating and accuracy on the detection task were found to have an effect on number of accidents. No other effect was found and no interaction was found between speed, RSME ratings and accuracy on detection task.

Table 3. Regression results for the data collected

Variable	**Coefficient**	**StdError**	**T Stat**	**Significance**
Speed	3.788	1.326	2.858	< 0.05
RSME	2.228	0.852	2.615	< 0.05
Headway	-2.107	2.865	-0.736	ns
LatPos	-1.307	0.950	-1.375	ns
AccuDet	-3.576	0.985	-3.631	< 0.05
RTDet	0.217	0.802	0.270	ns

4 Discussion

Although the number of participants in this study was low and interpersonal differences between participants should be kept in mind, results allow to derive some indications, which however should be validated in future studies with higher number of participants using a within-subjects design.

The time pressure was found to increase mean speed, as expected, and to deteriorate the lane keeping performance of participants, the latter may be due to their more frequent attempts to overtake slower lead vehicles.

The time pressure was also found to reduce the time headway at which participants initiated an evasive manouevre but only for the second occurrence of the sudden braking lead vehicle, not for the first occurrence neither for any other event. This may be due to the participants' attempt to finish the scenario within the set time, which urged them to drive more closely to lead vehicles, especially as time passed by.

The concurrent performance of the secondary task was not found to have an effect on the studied indicators of driving performance or on indicators of the secondary task performance, still it resulted in higher RSME ratings by participants. This may explain why performance on the secondary task did not deteriorate with time pressure. Despite the secondary task, participants may have always maintained their focus on driving and managed to adequately perform both the primary and secondary task, although with an increased mental workload.

As regards accident risk, more accidents occurred in the Pressure than in the No pressure group, which indicates that speed increases accident risk, as expected. In the Pressure group a lot of accidents were related to loss of vehicle control, that is road departure, which may be attributed to these participants driving at higher speeds. In the No pressure group accidents in the second run were less or equal than in the first run, which may indicate that these participants traded-off increasing situational demands due to the unexpected events by shifting attention to traffic dangers while driving at lower speed. In the Pressure group accidents in the second run were more than in the first, possibly due to the participants' attempt to finish the scenario within the set time.

No effect of Pressure or detection task on the response time to a critical event was found, that is the higher speed or the concurrent conduct of the detection task did not prohibit participants from paying full attention to traffic risks and timely reacting to them. Although this may be attributed to inter-personal differences, the lower (although not significantly) response times to critical events of the Pressure compared to the No pressure group may indicate that participants in the Pressure group attempted to trade-off increased speed by increasing their alertness to traffic dangers.

Supported by the results of the regression analysis, higher speed was found to lead to more accidents, the same holds true for the RSME ratings, while the increase in accuracy in detection task was found to relate with less accidents. The latter may be due to interpersonal differences, indicating that participants who performed better in driving performed also better in the secondary task. Moreover, these results are in accordance with the attentional management theory [16]; meaning that the effect of speed and mental workload on accident risk should not be considered as the mere effect of attentional resource availability but as depending on the drivers' skill to manage their attentional control.

No interaction effects of speed, RSME rating and accuracy on the detection task on number of accidents were found. This means that if it is found that a new driving support system results in an increase of driving speed and of mental effort, these effects should be clearly considered as additive in order to estimate the total system effect on accident risk.

References

1. Maycock, G., Brocklebanck, P.J., Hall, R.D.: Road layout design standards and driver behaviour. TRL report No. 332 (1998)
2. Taylor, M.C., Lynam, D.A., Baruya, A.: The effects of drivers' speed on the frequency of road accidents. TRL Report No. 421 (2000)
3. Wickens, C.D., Liu, Y.: Codes and modalities in multiple resources: A success and a qualification. Human Factors 30, 599–616 (1988)
4. Kanheman, D.: Attention and effort. Prentice Hall, Englewood Cliffs (1973)
5. Recarte, M.A., Nunes, L.M.: Effects of verbal and spatial-imagery task on eye fixations while driving. Journal of Experimental Psychology: Applied 6(1), 31–43 (2000)
6. McKnight, A.J., McKnight, A.S.: The effect of cellular phone use upon driver attention. Accident Analysis and Prevention 25(3), 259–265 (1993)
7. Alm, H., Nilsson, L.: The effects of a mobile telephone task on driver behaviour in a car following situation. Accident Analysis and Prevention 27(5), 707–715 (1995)
8. Brookhuis, K.A., DeVries, G., DeWaard, D.: The effects of mobile telephoning on driving performance. Accident Analysis and Prevention 23(4), 309–316 (1991)
9. Nunes, L., Recarte, M.A.: Cognitive demands of hands-free-phone conversation while driving. Transportation Research Part F 5, 133-144 (2002)
10. Briem, V., Hedman, L.R.: Behavioural effects of mobile phone use during simulated driving. Ergonomics 38(12), 2536–2562 (1995)
11. Brown, I.D., Tickner, A.H., Simmonds, D.C.V.: Interference between concurrent tasks of driving and telephoning. Journal of Applied Psychology 53(5), 419–424 (1969)
12. Lansdown, T.C., Burns, P.C.: Route guidance information: visual, verbal or both? In: International Ergonomics Association 13th Triennial Congress, Tampere, Finland, June 29-July 4 (1997)
13. Fairclough, S.H., Ashby, M.C., Parkes, A.M.: In-vehicle displays, visual workload and usability evaluation. In: Gale, A.G., Brown, I.D., Haslegrave, C.M., Kruysse, H.W., Taylor, S.P. (eds.) Vision in Vehicles - IV, pp. 245–254. Elsevier, Amsterdam (1993)
14. Dingus, T.A., McGehee, D.V., Mankkal, N., Jahns, S.K., Carney, C., Hankey, J.M.: Human factors field evaluation of automotive headway maintenance/collision warning devices. Human Factors 39, 216–229 (1997)
15. Cnossen, F., Rothengatter, Meigman, T.: Strategic changes in task performance in simulated car driving as an adaptive response to task demands. Transportation Research, Part F 3, 123–140 (2000)
16. Gopher, D.: The skill of attentional control: Acquisition and execution of attentional strategies. In: Kornblum, S., Meyer, D.E. (eds.) Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence and Cognitive Neuroscience. MIT Press, Cambridge (1993)
17. Zijlstra, F.R.H., Van Doorn, L.: The construction of a scale to measure perceived effort. Department of Philosophy and Social Sciences, Delft University of Technology, Delft, The Netherlands (1985)

Development of a Systems-Based Human Factors Design Approach for Road Safety Applications

Gemma J.M. Read[1], Paul M. Salmon[2], and Michael G. Lenné[1]

[1] Human Factors Group, Accident Research Centre, Monash Injury Research Institute, Monash University, Victoria, Australia
[2] University of the Sunshine Coast Accident Research, University of the Sunshine Coast, Maroochydore, Queensland, Australia
{gemma.read,michael.lenne}@monash.edu,
psalmon@usc.edu.au

Abstract. Cognitive work analysis (CWA), a systems-based analysis framework, is intended to inform system design. However, there is little guidance available about how to use the framework in design. This paper identifies desirable methodological attributes for a new design approach for CWA and describes a process of refining these to a core set based on the opinions of CWA practitioners. The new design approach, the CWA Design Tool (CWA-DT), is outlined in terms of how it aligns with these core attributes. Finally, implications of application of the CWA-DT for road safety design will be identified and discussed.

1 Introduction

With increasing calls for a systems approach to road safety [e.g. 1], practitioners require accessible methods and approaches for the analysis and design of road systems. Without such approaches, the much heralded systems approach cannot be realised in road transport. This paper describes a new approach that aims to support practitioners in the application of Cognitive Work Analysis (CWA), a systems-based analysis framework, in road safety design efforts.

CWA is a framework of methods that supports the analysis of complex socio-technical systems with the aim of improving system design [2]. The framework sits within the field of cognitive systems engineering (CSE) [3]. CWA has been widely used to understand complex systems including nuclear power generation, military command and control, air traffic control, healthcare and rail. It is also being increasingly applied within road transport [e.g. 4, 5, 6]. The framework uses five phases of analysis to describe the constraints of the system of interest from different perspectives. This provides a model of the boundaries of possibilities for action in the system, as opposed to specifying the desired behaviour.

As a framework, rather than a prescriptive methodology, skill and expertise is required in selecting the appropriate CWA phases and methods. In addition, a recent review identified a need for additional design methods when designing for intentional systems, such as road transport [7]. CWA design applications have also tended to focus on the development of displays and interfaces in isolation, rather than whole system design [7]. As such, there is a need to extend the design process to support

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 43–52, 2013.
© Springer-Verlag Berlin Heidelberg 2013

wider system elements such as task design, function allocation, physical system design, team-based distributed working, training and procedures. The implication of this is that a new approach for using CWA outputs in design is needed to assist road designs from a systems perspective. Without such an approach, the full potential of CWA in this domain is not being realised.

Based on a review of literature concerned with human factors (HF) methods generally and CSE methods specifically, this paper presents a discussion and synthesis of the various methodological attributes required for a CWA design approach. Following this, the results from a survey study in which CWA practitioners were asked to prioritise these attributes are presented. The findings are used to determine key features of the proposed CWA Design Tool (CWA-DT).

A new approach to designing with CWA should align with accepted HF methodological attributes [8]. As part of the development of a new CWA-based design approach, the literature was reviewed to draw out the methodological attributes for consideration. Each attribute identified is outlined below with a brief discussion of the relevant literature.

Attribute 1: Creativity *(Definition: Facilitates creativity and / or innovation)*

Design is well recognised as a creative process, and this view has been echoed in the CSE literature. For example, Rasmussen and colleagues [9] suggest that design is a creative process that should not be controlled by formal, normative procedures. Similarly, Millitello and colleagues argue that design needs to be an opportunitistic and explorative process, rather than one that is structured and systematic [3]. In the face of calls for more guidance and structure in design processes, the need to maintain creativity in the design process has been stressed [10].

Attribute 2: Efficient *(Definition: Process is efficient and / or cost effective)*

Many CWA researchers and practitioners have noted the time requirements associated with applying the framework [e.g. 11, 12, 13] in particular its reputation for being time consuming and resource intensive [14].

There is a need to improve efficiency in the design process [10] and for design processes in CSE to be practical and cost effective [15]. This reinforces previous calls to ensure that the resources consumed in the analysis and design processes are proportionate to the benefits gained [16]. An efficient method that minimises time and resource allocation, including time for training, is more likely to be applied in practice.

Attribute 3: Holistic *(Definition: Supports coordinated design of all system elements, e.g. interfaces, training, support materials, team structures)*

CWA provides a holistic understanding of system functioning. A design approach following from the analysis should encompass this perspective and support a holistic design process. Socio-technical systems theory recognises the systemic nature of design; that due to the interrelatedness of system components, changes made to one aspect of a system will have effects in other parts of the system [17]. As such, designers need to consider unanticipated effects on the system before implementation, and ensure that the system is designed to cope with emerging effects over the system lifecycle.

The majority of CWA design applications have designed interfaces [7], even though it has been emphasised that an interface should not be implemented independently of other systems aspects such as team design, alarm design, etc [12]. Coherent design, where different aspects of the system are designed so that they are compatible and integrated, has been proposed to promote efficiency and to reduce errors [18].

Attribute 4: Integrated *(Definition: Can integrate with existing systems engineering processes)*

The importance of integrating the results of CSE design processes with system design and development processes has been emphasised [e.g. 15, 19]. Integration can be supported through presenting analysis findings in a manner that links to the wider system development process [19] or through CSE practitioners understanding the systems engineering process and terminology and presenting analysis contributions consistent with this knowledge [3].

Methods should be linked to broader design processes and the products of the design should be integrated into this wider process [16]. Methods should also align with, or be consistent with, existing tools and techniques [20] as a method that integrates with existing processes could be easier to learn and should be easier to apply within a systems design process [21].

Attribute 5: Iterative *(Definition: Facilitates an iterative design process)*

CSE methods are generally intended to facilitate ongoing re-evaluation and reconsideration of the problem being investigated as new information arises, or as the analyst progressively builds their understanding of the system [3]. Either the design problem or the initial design solutions may be re-framed or altered throughout the process. This is consistent with views of design as an iterative, rather than step-by-step methodology.

Iteration is required because of the complexity of the domains being analysed and in recognition of the systemic nature of design [17]. Iteration enables decisions to be amended and re-evaluated as the process proceeds [21].

Attribute 6: Reliable *(Definition: Produces consistent results each time it is applied)*

The literature on HF methods emphasises the importance of demonstrating that ergonomics methods actually work, to provide confidence in the methods [22] and to inform decision making regarding cost-effectiveness of the various methods [23]. As a rule, the focus has been on reliability and validity (see Attribute 13) as the basic objective measures of the success of a method [23, 24].

Reliability is concerned with the repeatability of results obtained from a method either between different analysts [22, 24, 25] or within the same analyst over time [24]. It is concerned with whether measurements are repeatable and accurate [26].

CSE researchers have called for design processes to be more repeatable [15], and for the methods in CWA specifically to be better defined to promote analyst reliability [11]. Developments in the practice of CWA and tools to assist analysts have begun working towards supporting consistency in analysis [e.g. 27].

Attribute 7: Stakeholder involvement *(Definition: Involves project stakeholders (e.g. designers, engineers, management) in the design process)*

Participation of stakeholders ensures that the design meets the needs for which it is required [21]. Stakeholders are those affected by, or with an interest in, the outcome of a design process. This could include people who operate in, and manage the system, as well as those involved in the design process such as analysts, designers and engineers. While users of the system are also stakeholders, for the purposes of this analysis users are incorporated in Attribute 14 (Worker / user involvement).

Stakeholders have different perspectives on the system, and different views of a design problem. Sensitising stakeholders to the different world views and concerns of other stakeholders can assist to reconcile issues arising from this [28]. The involvement of stakeholders with diverse knowledge, skills and expertise can also facilitate multidisciplinary learning and foster creativity and innovation [17]. Stakeholder participation generally encourages ownership of designs by those who will be responsible for the implementation, operation and use of the system [17].

Attribute 8: Structured *(Definition: Provides structure to the design process)*

The importance of structured and systematic methods has a long history in HF. Degree of structure has been used as a criterion to evaluate human error identification methods [29], and there are calls for more systematic methods in the CSE field [27].

A structured approach to design should provide a link between the analysis of the system and the cognitive artifacts produced in the design process. It should provide accountability in the design process and enable the specification of a clear path forward with the ability to trace and understand the reasons for decisions made further back in the analysis [30]. A structured process can improve efficiency, communication between analysts and reduce training time [14].

Attribute 9: Tailorable *(Definition: Can be tailored for different system types (e.g. intentional, causal, first-of-a-kind)*

Human factors methods generally need to be adaptable to different situations and be tailorable for unique applications. They should be applicable to complex systems and able to deal with the growing complexity expected of future systems [20]. They must also support application to specific situations [21]. Methods should be flexible to meet the requirements of the project, and possess contextual validity through being sensitive to contextual factors within the specific application domain [29]. A method that is tailorable could perhaps achieve this contextual validity.

Attribute 10: Theoretical *(Definition: Is consistent with the underpinning theory and principles of CWA)*

A specific aspect of validity (see Attribute 13) is construct validity. This is met when the method is based on an appropriate underlying theory [24], and has an internal structure that aligns with that theory [29]. To fully exploit the benefits of the CWA philosophy, a design method to be used with CWA should align with its underlying theoretical basis.

While CWA may not have a strong theoretical grounding in terms of a single theory upon which it is structured, it does draw from a number of established theoretical areas. These include ecological psychology and systems theory [11] and, in particular, socio-technical systems theory [31]. The underlying principles of CWA include the need to make system constraints visible to workers or users, to support cognitive processing at the appropriate level avoiding unnecessary workload, and to support knowledge-based behaviour in unusual, unanticipated circumstances [2]. These principles should be reflected in the design concepts and solutions developed.

Attribute 11: Traceable *(Definition: Provides a detailed record of design decisions)*

Traceability between design products and the information represented in the analysis enables testing of whether the design adequately addresses what was uncovered by the analysis [30]. Where designers have not been involved in the analysis, a traceable process enables designers to discover the rationale behind, and justification for, decisions that affect the subsequent design process [32]. A traceable process provides auditable documentation [29] enabling updating and supporting communication within the design team [16]. Materials that promote traceability can also assist integration with other system development processes [19].

Attribute 12: Usable *(Definition: Is usable for CWA practitioners, systems designers, engineers, etc)*

To be accepted by its potential users, a method needs to be usable. When considering HF methodologies, usability can be defined as the ease of use of the method [29]. A method that is usable and straightforward to learn is more likely to be selected for use in practice [21]. Usability has previously been applied as a criterion for evaluating human error identification methods [25, 29] and it has been suggested that a usable method promotes better consistency amongst analysts and less errors than one which is demanding or difficult to use [25].

Attribute 13: Valid *(Definition: Does what it says it will do, e.g. produces effective designs)*

As discussed previously, validity is considered one of the cornerstone measures of a robust methodology [23]. Various types of validity are discussed in the literature including face validity, construct validity and predictive validity [24]. In general, validity means that a method does what it claims it will do.

Validity is related to the efficacy of a method for the problem it is intended to solve. It has been suggested that CWA should be able to demonstrate that it can handle new and novel problems and provide better outcomes than other, less resource intensive techniques [11]. A number of the attributes already discussed such as contextual validity (Attribute 9) and theoretical validity (Attribute 10) could contribute to the achievement of a valid design process.

Attribute 14: Worker / user involvement *(Definition: Involves workers / end users in the design process)*

As noted previously in relation to stakeholder involvement, ownership of the system should reside, in part, with those who will operate and use the system as these people will continue to shape the system over time [17]. User participation in design is one of the underpinning principles of the HF discipline. User-centered design has been used in combination with CWA to improve design concepts and solutions in a number of applications [e.g. 4, 33]. Participation of potential end users has also been used as a criterion to evaluate HF methods [e.g. 20].

1.1 Prioritisation of Methodological Attributes

The 14 attributes were presented to CWA practitioners via an electronic survey. Respondents were asked to rank the identified attributes for a CWA-based design approach in order of importance, with a rank of one indicating that this was the most important attribute and a rank of fourteen indicating an attribute was the least important. This was undertaken to gain feedback from the potential user group on the relative priority of the different attributes. This prioritisation process was considered necessary as it can be difficult for a single method to fulfill all attributes [29].

The rankings provided by the survey respondents revealed six key attributes required for a new CWA-based design approach (see Table 1). These principles, highlighted in italics, were that the approach facilitates creativity, is holistic, facilitates iteration, is efficient, integrates with existing system engineering processes and provides a structured approach.

Table 1. Results of ranking of the 14 proposed attributes of a new CWA-based design approach

Attribute	Average ranking	Attribute	Average ranking
1. Creative	*5.65*	*8. Structured*	*7.1*
2. Efficient	*6.5*	9. Tailorable	7.45
3. Holistic	*5.65*	10. Theoretical	8.4
4. Integrated	*6.8*	11. Traceable	9.3
5. Iterative	*6.3*	12. Usable	7.65
6. Reliable	8.05	13. Valid	8.75
7. Stakeholder involvement	9	14. Worker / user involvement	8.4

2 The CWA Design Tool

Based on a review of the literature and the survey findings a prototype CWA design process was developed. The CWA-DT was developed to align with the six more highly endorsed principles from the CWA practitioner survey. Development of the tool drew upon information from the literature on HF methods, systems engineering, and design. Input was also gathered via a workshop of CWA experts, with further discussions held to assist in evaluating options and refining the design approach.

The CWA-DT comprises guidance for analysis and design, acknowledging the interconnectivity of these processes and the artificial nature of a stand-alone design approach with no reference to the analysis. The analysis guidance assists practitioners to define the perceived problem and any known boundaries on the analysis and design process. It provides guidance for selecting the most appropriate phases and tools for exploring that problem and for moving through the analysis stages, referring to other literature and tools already available to CWA users. Key aspects of the analysis process include that it promotes connectivity and iteration between phases, involvement of subject matter experts and stakeholders, and documentation of design insights throughout the analysis process.

The design process begins with a review of the design insights and synthesis of these into a design brief. Next, the outputs from the last phase of the analysis undertaken (as not all phases will be completed for all design applications), are reviewed and a template completed identifying the high level requirements for subsequent design activities. Key aspects of the design process include use of structured documentation such as a design brief and design criteria document, promotion of iteration amongst phases of design (including changes to design documentation where required) and participation of users and other stakeholders.

Table 2 outlines how the CWA-DT aligns with the methodological attributes rated highly in the CWA practitioners' survey.

Table 2. Features of the CWA-DT related to the six highest rated methodological attributes

Attribute	Description of CWA-DT alignment
Creative	- Incorporates techniques to promote creativity during the design process for example: ~ Individual and group brainstorming in a workshop setting ~ Creativity boosting exercises ~ Room set up according to research on creativity boosting environments ~ Selection of diverse design participants ~ Provision of constraints within which creativity can be explored
Holistic	- Incorporates prompts based on the holistic view provided by the CWA analysis - Incorporates prompts for considering which aspects of the system should be considered (e.g. does the proposed design require a change to procedures or rules, or to the way teams coordinate, etc) - Designs evaluated through the analysis models highlighting interactions (potential emergence) and potential inconsistencies between the existing and proposed system - Evaluation of design concepts to identify potential unexpected system functioning from implementation
Iterative	- Iteration encouraged throughout the analysis and design process, particularly as design concepts emerge. For example, the process encourages: ~ Review and revision of the WDA following each later analysis phase ~ Iteration of design concepts and solutions throughout the design phase - Encourages amendments to design templates if necessary due to emerging understanding of the problem as the design process proceeds
Efficient	- Guidance minimises time required to set up a design process for each application - Guidance supports non-CWA trained participants in the design process to more quickly understand the key concepts - Guidance provided for tailoring design process to meet project constraints - Encourages analysts, designers and stakeholders to participate in both analysis and design changes, minimising inefficiencies involved in hand-over activities
Integrated	- Prompts consideration of relevant design standards relating to the system of interest - Analysis brief prompts consideration of the project dependencies such as concurrent system design activities - Encourages participation of system designers in the analysis and design process, facilitating integration with concurrent or subsequent system design activities
Structured	- Use of standardised templates for the analysis brief, design brief, design criteria and design concept documentation - Guidance provides structure to analysis and design processes, acknowledging that users of the methodology may modify the processes to suit their particular purposes

3 Application to Road Safety Issues

While intended that the CWA-DT will be useful for design within a wide range of domains, it will be formally validated in a road safety context. This will involve applying the tool to a road safety issue and then gathering data and feedback on the extent to which it meets the highly rated attributes outlined in Table 2, as well as more traditional reliability and validity measures.

There are numerous applications within road safety for the CWA-DT including areas already explored with CWA such as intersections [6] and in-vehicle interfaces [e.g. 4, 5], as well as emerging areas of concern [34]. These emerging areas might include designing to promote safe cycling and walking and improving the design of road interfaces with public transportation including railway level crossings (grade crossings) and passenger safety at bus and light rail stops. The approach could potentially be applied to new developments at individual locations, to the design of new technologies or stand-alone countermeasures. The CWA-DT could also potentially be used more broadly to re-define standards and policies for road design, as well as training and education programs. Key to realising the potential of CWA in the road transport context is the evaluation, refinement, and communication of the CWA-DT.

4 Conclusion

While the CWA framework has strong support and evidence for its analysis function, it does not currently provide a design process and users of the framework are required to craft their own approach to design, with little direct guidance available within the existing literature. Through sharing the structured approach developed through this work, we hope to assist in the translation of CWA outputs into real world designs in areas such as road transport. It is expected that the availability of this approach will increase the frequency of CWA design applications, which will in turn lead to the realisation of safer, more efficient and effective road systems.

Acknowledgements. Gemma Read's contribution to this article was funded through an Australian Postgraduate Award (Industry) provided by an Australian Research Council Linkage Grant (ARC, LP100200387) to Monash University in partnership with the Victorian Rail Track Corporation, Transport Safety Victoria, Public Transport Victoria, Transport Accident Commission, Roads Corporation (VicRoads) and V/Line Passenger Pty Ltd. Dr Paul Salmon's contribution to this article was funded through an Australian National Health and Medical Research Council post doctoral training fellowship.

References

1. Salmon, P.M., McClure, R., Stanton, N.A.: Road transport in drift: Applying contemporary systems thinking to road safety. Safety Science 50, 1829–1838 (2012)
2. Vicente, K.J.: Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work. Lawrence Erlbaum Associates, Mahwah (1999)

3. Militello, L.G., Dominguez, C.O., Lintern, G., Klein, G.: The role of cognitive systems engineering in the systems engineering design process. Systems Engineering 13, 261–273 (2010)
4. Mendoza, P.A., Angelelli, A., Lindgren, A.: Ecological interface design inspired human machine interface for advanced driver assistance systems. Intelligent Transport Systems, IET 5, 53–59 (2011)
5. Birrell, S.A., Young, M.S., Jenkins, D.P., Stanton, N.A.: Cognitive Work Analysis for safe and efficient driving. Theoretical Issues in Ergonomics Science (2012)
6. Cornelissen, M., Salmon, P.M., Young, K.L.: Same but different? Understanding road user behaviour at intersections using cognitive work analysis. Theoretical Issues in Ergonomics Science, 1–24 (2012)
7. Read, G.J.M., Salmon, P.M., Lenne, M.G.: From work analysis to work design: A review of cognitive work analysis design applications. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting (2012)
8. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical Guide for Engineering and Design. Ashgate, Surrey (2005)
9. Rasmussen, J., Pejtersen, A.M., Schmidt, K.: Taxonomy for cognitive work analysis. Risø National Laboratory (1990)
10. Hajdukiewicz, J.R., Burns, C.M.: Strategies for Bridging the Gap between Analysis and Design for Ecological Interface Design. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 48, pp. 479–483 (2004)
11. Sanderson, P.M.: Cognitive Work Analysis. In: Carroll, J. (ed.) HCI Models, Theories, and Frameworks: Toward an Interdisciplinary Science, Morgan-Kaufmann, New York (2003)
12. Vicente, K.J.: Ecological interface design: Process and challenges. Human Factors 44, 62–78 (2002)
13. Jamieson, G.A.: Bridging the Gap Between Cognitive Work Analysis and Ecological Interface Design. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 47, pp. 273–277 (2003)
14. Rehak, L.A., Lamoureux, T.M., Bos, J.C.: Communication, coordination, and integration of Cognitive Work Analysis outputs. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 50, pp. 515–519 (2006)
15. Gualtieri, J.W., Szymczak, S., Elm, W.C.: Cognitive System Engineering - Based Design: Alchemy or Engineering. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 49, pp. 254–258 (2005)
16. Potter, S.S., Roth, E.M., Woods, D.D., Elm, W.C.: A framework for integrating cognitive task analysis into the system development framework. In: Proceedings of the 42nd Human Factors and Ergonomics Society Annual Meeting, vol. 1, pp. 395–399. HFES, Santa Monica (1998)
17. Clegg, C.W.: Sociotechnical principles for system design. Applied Ergonomics 31, 463–477 (2000)
18. Gonzalez Castro, L.N., Pritchett, A.R., Bruneau, D.P.J., Johnson, E.N.: Applying coherent design to uninhabited aerial vehicle operations and control stations. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 51, pp. 181–185 (2007)
19. Bisantz, A.M., Roth, E., Brickman, B., Gosbee, L.L., Hettinger, L., McKinney, J.: Integrating cognitive analyses in a large-scale system design process. International Journal of Computer-Human Studies 58, 177–206 (2003)
20. Waterson, P.E., Older Gray, M.T., Clegg, C.W.: A sociotechnical method for designing work systems. Human Factors 44, 376–391 (2002)

21. Older, M.T., Waterson, P.E., Clegg, C.W.: A critical assessment of task allocation methods and their applicability. Ergonomics 40, 151–171 (1997)
22. Stanton, N.A., Stevenage, S.V.: Learning to predict human error: issues of acceptability, reliability and validity. Ergonomics 41, 1737–1756 (1998)
23. Stanton, N.A., Young, M.S.: What price ergonomics? Nature 399, 197–198 (1999)
24. Baber, C., Stanton, N.A.: Task analysis for error identification: Theory, method and validation. Theoretical Issues in Ergonomics Science 3, 212–227 (2002)
25. Baysari, M.T., Caponecchia, C., McIntosh, A.S.: A reliability and usability study of TRACEr-RAV: The technique for the retrospective analysis of cognitive errors - For rail, Australian version. Applied Ergonomics 42, 852–859 (2011)
26. Gawron, V.J.: Handbook of human performance measures. Lawrence Erlbaum Associates, Mahwah (2000)
27. Sanderson, P., Eggleston, R., Skilton, W., Cameron, S.: Work Domain Analysis workbench: Supporting Cognitive Work Analysis as a systematic practice. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 43, pp. 323–327 (1999)
28. Baxter, G., Sommerville, I.: Socio-technical systems: From design methods to systems engineering. Interacting with Computers 23, 4–17 (2011)
29. Shorrock, S., Kirwan, B.: Development and application of a human error identification tool for air traffic control. Applied Ergonomics 33, 319–336 (2002)
30. Elm, W., Gualtieri, J., Tittle, J., Potter, S.S., McKenna, B.: Pragmatic use of cognitive work analysis in system design- Extending current thinking by adapting the mapping principle. In: Bisantz, A.M., Burns, C.M. (eds.) Applications of Cognitive Work Analysis. CRC Press (2008)
31. Stanton, N.A., McIlroy, R.C.: Designing mission communication planning: the role of Rich Pictures and Cognitive Work Analysis. Theoretical Issues in Ergonomics Science 13, 146–168 (2012)
32. Kilgore, R., St-Cyr, O., Jamieson, G.A.: From Work Domains to Worker Competencies: A Five-Phase CWA. In: Bisantz, A.M., Burns, C.M. (eds.) Applications of Cognitive Work Analysis, pp. 15–48. CRC Press (2008)
33. McIlroy, R.C., Stanton, N.A.: Specifying the requirements for requirements specification: the case for Work Domain and Worker Competencies Analyses. Theoretical Issues in Ergonomics Science 13, 450–471 (2012)
34. Salmon, P., Stephan, K., Lenné, M., Regan, M.: Cognitive Work Analysis and Road Safety: Potential Applications in Road Transport. In: Australasian Road Safety Research Policing Education Conference (2005)

Awesome Foursome? The Compatibility of Driver, Cyclist, Motorcyclist, and Pedestrian Situation Awareness at Intersections

Paul M. Salmon[1], Michael G. Lenné[2],
Guy H. Walker[3], and Ashleigh Filtness[2]

[1] University of the Sunshine Coast Accident Research (USCAR),
Faculty of Arts and Business, School of Social Sciences, Maroochydore, QLD 4558, Australia
[2] Human Factors Group, Monash University Accident Research Centre,
Building 70, Clayton Campus, Victoria 3800, Australia
[3] School of the Built Environment, Heriot-Watt University,
Edinburgh, EH14 4AS, UK
psalmon@usc.edu.au

Abstract. Collisions between distinct road users (e.g. drivers and motorcyclists) make a substantial contribution to the road trauma burden. Although evidence suggests distinct road users interpret the same road situations differently, it is not clear how road users' situation awareness differs, nor is it clear which differences might lead to conflicts. This article presents the findings from an on-road study which examined driver, cyclist, motorcyclist and pedestrian situation awareness at intersections. The findings suggest that situation awareness at intersection is markedly different across the four road user groups studied, and that some of these differences may create conflicts between the different road users. The findings also suggest that the causes of the differences identified relate to road design and road user experience. In closing, the key role of road design and training in supporting safe interactions between distinct road users is discussed.

1 Introduction

Road transport-related trauma continues to be one of the leading causes of death and disability across the world (World Health Organization, 2009). Although significant reductions in death and injury have been made over the last four decades in most motorized countries (Elvik, 2010) a number of complex intractable issues remain. One of these is collisions between different types of road user (e.g. drivers and motorcyclists, drivers and cyclists). For example, an analysis of UK motorcyclist crashes found that their most common cause was other vehicles entering motorcyclists' path when exiting side roads (Clarke et al, 2007). Similarly, the road safety literature suggests that a high proportion of cyclist crashes involve drivers' failing to detect cyclists and colliding with them (Wood et al, 2009). Elvik (2010) identifies incompatibilities between different road user groups as one of five key road safety issues.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 53–62, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Despite forming a substantial component of the road trauma burden, the causes of collisions between distinct road users remain ambiguous. Moreover, it is not clear what countermeasures are the most appropriate. Recent evidence suggests that the ubiquitous concept of situation awareness has a key role to play in understanding and preventing collisions between different road users. Specifically, studies of road user situation awareness underpinned by Niesser's (1976) seminal perceptual cycle model suggest that differences in road user schema and behavior, driven by experience, transport mode, and road design, may lie at the root of these conflicts (e.g. Salmon et al, 2013; Walker et al, 2011). Low sample sizes have however thus far limited the generalizability of results, and researchers acknowledge the need for further confirmatory research (Salmon et al, 2013; Walker et al, 2011). This paper presents the findings from a large scale on-road investigation of driver, cyclist, motorcyclist and pedestrian situation awareness at intersections. The study involved assessing situation awareness across seventy eight participants whilst they negotiated an urban study route incorporating 3 major intersections requiring a right hand turn. The aim of the study was to identify the key differences in situation awareness between road users, to identify the causes of these differences, and to identify potential conflicts that arise when road users understand the same road situations differently.

2 Assessing Situation Awareness on the Road

Following Salmon et al (2013) and Walker et al (2011), the present study used a network analysis-based approach to describe and assess road user situation awareness. The approach involves constructing situation awareness networks using data derived from the Verbal Protocol Analysis (VPA) method, which involves participants 'thinking aloud' as they perform tasks. Based on content analysis of the VPA transcripts, the situation awareness networks depict the information or concepts underlying situation awareness and the relationships between the different concepts. For example, the concept 'Traffic light' may be related to the concept 'Green' since the traffic light 'is' green. Similarly, the concept 'car' may be related to concepts such as 'speed' (as in car 'has' speed), 'brakes' (as in car 'has' brakes), 'moving' (as in car 'is' moving) and 'driver' (as in car 'has' driver). Mathematical analysis is then used to interrogate the content and structure of the networks. This enables comparison of situation awareness across different actors and scenarios (e.g. Walker et al, 2011).

3 On-Road Study

Situation awareness networks were used to describe road user situation awareness when turning right at three major signalized intersections. A range of quantitative and qualitative network analysis procedures were then used to analyze the content and structure of the networks. In the present paper, the content of the networks is examined as a way of determining what each road user groups' situation awareness comprised when they negotiated the three intersections. Based on previous research (e.g. Salmon et al, 2013; Shahar et al, 2010; Walker et al, 2011), the hypothesis was that the different road users (drivers, cyclists, motorcyclists, pedestrians) would interpret the intersection situations differently. Specifically, there would be differences in the concepts underpinning each road user group's situation awareness. Following this, an investigation into the compatibility between road users' situation awareness and the reasons underpinning the key differences in situation awareness was undertaken.

3.1 Methodology

Design. The study was an on-road study using a semi-naturalistic paradigm whereby participants negotiated a pre-defined route incorporating three major intersections requiring a right hand turn. Drivers drove the Monash University On-Road Test Vehicle (ORTeV), whilst motorcyclists and cyclists completed the route using their own motorcycle or bicycle which was instrumented with video and audio recording equipment. Pedestrians negotiated the three intersections on foot whilst wearing video recording glasses. All participants provided concurrent verbal protocols as they negotiated the study route.

Participants. Seventy eight participants (52 male, 26 female) aged 21 - 64 years (mean = 35.81, SD = 13.03) took part in the study. They comprised 20 car drivers, 18 motorcyclists, 20 cyclists, and 20 pedestrians. An overview of the participants in each group, including gender, mean age and experience is presented in Table 1.

Table 1. Participant demographics

Road user group	Mean age (SD)	Gender	Hours drove/rode/cycled/walked per week
Drivers	34.9yrs (12.53)	10 males 10 females	11.5 hours
Cyclists	32.4yrs (10.42)	15 males 5 females	6.85 hours
Motorcyclists	45.5yrs (12.87)	17 males 1 female	7 hours
Pedestrians	30.5yrs (11.86)	10 males 10 females	8.92 hours

Participants were recruited through a weekly on-line university newsletter and were compensated for their time and expenses. Prior to commencing the study ethics approval was formally granted by the Monash Human Ethics Committee.

Materials. A demographic questionnaire was completed using pen and paper. A desktop driving simulator was used to provide verbal protocol training. A 15km urban route, located in the south-eastern suburbs of Melbourne, was used for the on-road study component.

Drivers drove the route in a 2004 Holden Calais sedan equipped to collect various vehicle, driving scene and driver-related data. A Dictaphone was used to record all participants verbal protocols. Motorcyclists rode the route using their own motorcycle. Each motorcycle was fitted with an Oregon Scientific ATC9K portable camera, which, depending on motorcycle model, was fixed either to the handlebars or front headlight assembly. The ATC9K camera records the visual scene, speed and distance travelled (via GPS). A microphone was fitted inside each motorcyclist's motorcycle helmet and attached to the Dictaphone to record their verbal protocols. Cyclists cycled the route using their own bicycle. To record the cycling visual scene and the cyclist verbal protocols, the ATC9K portable camera was fitted to the cyclists' helmets, and cyclists wore Imging HD video cycling glasses. Pedestrians negotiated the intersections on foot whilst wearing Imging HD video sunglasses and a microphone linked to a Dictaphone. All verbal protocols were transcribed using Microsoft Word.

For data analysis, the Leximancer™ content analysis software was used.

Procedure. In order to control for traffic conditions, all trials took place at the same pre-defined times on weekdays (10am or 2pm Monday to Friday). Participants first completed an informed consent form and demographic questionnaire and were then briefed on the research and its aims (expressed as a study of driver behavior). Following this they were given training in providing verbal protocols which included a desktop driving simulator task where they were asked to complete the drive whilst providing a verbal protocol. An experimenter monitored the drive and provided feedback to the participant regarding the quality of their verbal protocol. Participants were then shown the study route and were given time to memorize it. When comfortable with the verbal protocol technique and route, participants were taken to their vehicle and asked to prepare themselves for the test. They were then given a demonstration of the video and audio recording equipment, which was also set to record at this point. Following this, the experimenter instructed the participant to begin the study route. For the drivers, an experimenter was located in the vehicle and provided route directions if necessary. For the motorcyclists and cyclists, an experimenter followed behind (in a car for the motorcyclists, on a bicycle for the cyclists) ready to intervene if the participants strayed off route. Pedestrians were taken by car to the first intersection and instructed to negotiate the intersection and walk to a set point following the intersection. Once the participant reached this point, they were picked up by the experimenter and driven to the next intersection. This process was repeated until all three intersections had been negotiated.

Participants' verbal protocols were transcribed verbatim using Microsoft Word. For data reduction purposes, extracts of each verbal transcript for each intersection were taken from the overall transcripts. The extracts were taken based on the video data and pre-defined points in the road environment (e.g. beginning and end of intersection). The verbal transcripts were then analyzed using the Leximancer content analysis software in order to create situation awareness networks. Leximancer uses text representations of natural language to interrogate verbal transcripts and identify themes, concepts and the relationships between them. The software does this by using algorithms linked to an in-built thesaurus and by focusing on features within the verbal transcripts such as word proximity, quantity and salience. The output is a network showing concepts and the relationships between them according to the transcript.

4 Results

Leximancer was used to construct overall driver, cyclist, motorcyclist and pedestrian situation awareness networks for each intersection. The four networks for each intersection were then mapped onto one another to produce a multi-road user situation awareness network for that intersection. For example, the multi-road user situation awareness network for intersection 1 is presented in Figure 1. Within Figure 1 the nodes and links are shaded to depict each road user group's situation awareness. Figure 1 shows how situation awareness differed across the distinct road user groups whilst negotiating intersection 1, both in terms of the concepts underpinning situation awareness (i.e. nodes in the network), and also in the way in which the concepts were linked together (i.e. links between the nodes in the network). Moreover the network demonstrates that, even when the different road users were using the same concept, they were doing so in conjunction with other different concepts. This pattern is repeated over the other two networks studied (the networks for intersections 2 and 3 are not presented due to space constraints). The multi-road user situation awareness networks therefore confirm that driver, cyclist, motorcyclist, and pedestrian situation awareness was different when negotiating the three intersections studied.

The differences in situation awareness are explored further in Table 2, which presents a summary of concept usage across the four road user groups for each intersection. Table 2 shows that, at intersection 1, only 19% of all concepts from the multi-road user network were used by all road user groups. Similarly, only 20.9% and 14.6% of concepts were found in all four road user groups situation awareness networks at intersection 2 and 3 respectively. Table 2 also shows that, at intersection 1, around 5% of all concepts were unique to drivers, around 10% were unique to cyclists, around 10% were unique to motorcyclists, and almost 20% were unique to pedestrians. A similar pattern is also found at intersections 2 and 3.

Table 2 also shows how some concepts were common across different combinations of the four road user groups. For example, almost 10% of concepts at the three intersections were found in driver, cyclist, and motorcyclist situation awareness networks, but not in the pedestrian networks. Interestingly, at intersection 3 over 10% of the concepts were found in both cyclist and pedestrian situation awareness networks. This is likely due to the complexity and high risk nature of intersection 3, which meant that many cyclists exploited the pedestrian crossings when negotiating the intersection.

Figure 2 shows the 'common' concepts (i.e. those found in all road user groups situation awareness networks) along with the concepts unique to each road user group at each intersection. Figure 2 shows that the concepts common across all four road user groups at the three intersections were mainly related to the cars, traffic, the road, the lights and their status (e.g. green), the intersection, and the act of turning. This reflects a high focus of all road users on cars, the traffic lights, and the intersection environment itself.

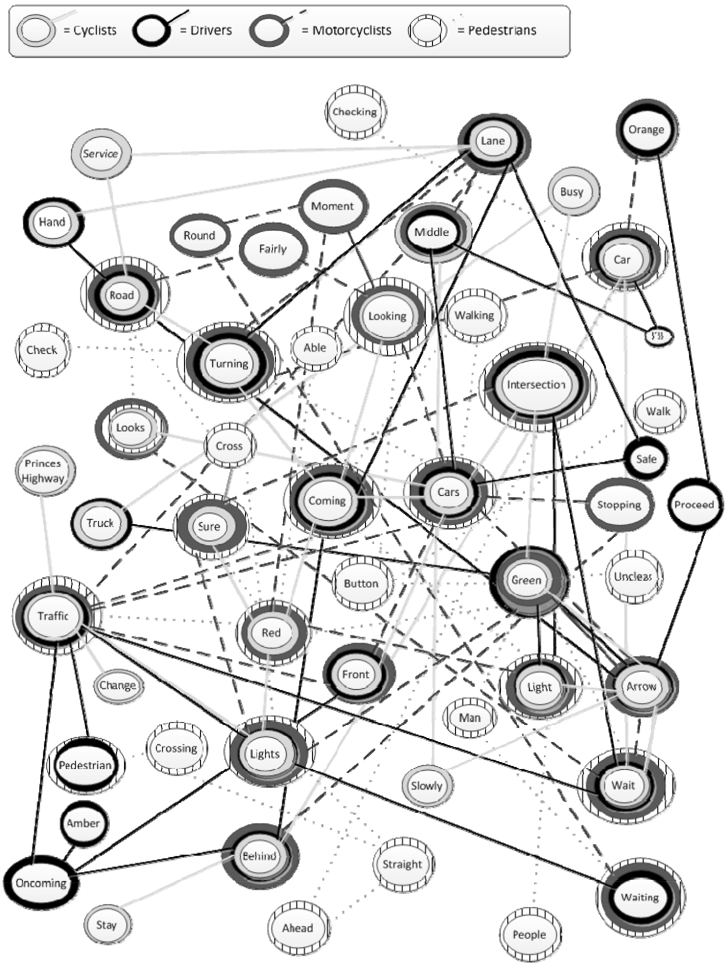


Fig. 1. Intersection 1 situation awareness network showing driver, cyclist, motorcyclist and pedestrian situation awareness networks mapped onto each other

Important differences in concept usage are shown in Figure 2. At intersections 1 and 2, the cyclist networks include the concepts ‘service’, ‘stay’ and ‘route’, all of which reflect a key decision that they face regarding whether or not to use the service lane on approach to the intersection and then cross via the footpath, or to stay on the road and go through the intersection in the normal traffic flow. Moreover, prior to the intersection the cyclists also decide whether they will leave the road and get back into the service lane once they have passed through the intersection. At intersection 3, the cyclists' network included the 'hook' concept, which refers to their decision regarding whether or not to use a hook turn in order to turn right at the intersection. Again this reflects a key decision whereby cyclists try to work out whether it is safe enough to pass through the intersection on the road within the flow of traffic or whether they need to perform a hook turn to avoid conflict with other traffic also turning right.

Table 2. Concept usage across the road user groups

	Int 1	**Int 2**	**Int 3**
Drivers			
Number of concepts	19	22	23
Unique concepts	2 **(4.9%)**	1 **(2.3%)**	6 **(13%)**
Cyclists			
Number of concepts	25	23	23
Unique concepts	4 **(9.8%)**	6 **(14%)**	2 **(4.2%)**
Motorcyclists			
Number of concepts	23	25	24
Unique concepts	4 **(9.8%)**	5 **(11.6%)**	7 **(14.6%)**
Pedestrians			
Number of concepts	22	23	26
Unique concepts	8 **(19%)**	5 **(11.6%)**	9 **(18.8%)**
Common concepts			
Concepts common across all road user groups	8 **(19%)**	9 **(20.9%)**	7 **(14.6%)**
Concepts used by drivers, cyclists, and motorcyclists only	4 **(9.8%)**	3 **(7%)**	4 **(8.3%)**
Concepts used by drivers, cyclists, and peds only	1 **(2.4%)**	2 **(4.7%)**	2 **(4.2%)**
Concepts used by drivers, motorcyclists, and peds only	-	1 **(2.3%)**	1 **(2.1%)**
Concepts used by cyclists, motorcyclists, and peds only	4 **(9.8%)**	1 **(2.3%)**	2 **(4.2%)**
Concepts used by drivers and cyclists only	2 **(4.9%)**	-	1 **(2.1%)**
Concepts used by drivers and motorcyclists only	1 **(2.4%)**	4 **(9.3%)**	1 **(2.1%)**
Concepts used by motorcyclists and cyclists only	2 **(4.9%)**	1 **(2.3%)**	1 **(2.1%)**
Concepts used by drivers and peds only	1 **(2.4%)**	2 **(4.7%)**	-
Concepts used by cyclists and peds only	-	1 **(2.3%)**	5 **(10.4%)**
Concepts used by motorcyclists and peds only	-	2 **(4.7%)**	-

For the motorcyclists, the unique concepts relate primarily to the selection of the left or right hand lane to negotiate the intersection (e.g. 'hand', 'left hand', 'merging'), the motorcycle itself (e.g. 'bike', 'gear') and the 'line' that they should take through the intersection. The 'stopping' concept refers to motorcyclist's own braking behavior, but also to them checking that other traffic approaching from behind are stopping when the traffic lights are on red.

For the pedestrians, the unique concepts were primarily related to the physical acts of walking (e.g. 'walk/walking) and crossing the road (e.g. 'cross/crossing') and also the crossing infrastructure (e.g. 'button', green 'man'). Interestingly, only the pedestrian networks included the concepts 'check/checking' and 'look/looking', which indicate that the other road users placed less emphasis on checking other traffic and the road environment when negotiating the intersections.

	Common across all road user groups	Unique to drivers	Unique to cyclists	Unique to motorcyclists	Unique to pedestrians
Intersection 1	Car(s), Road, Light(s), Green, Intersection, Traffic, Wait/ Waiting, Coming	Safe, I'll	Service, Road name, Slowly, Stay	Stopping, Fairly, Moment, Round	Walk/ Walking, Check/ Checking, Button, Ahead, Man, Cross/ Crossing, Able, Straight
Intersection 2	Car(s), Lane/ Lanes, Light(s), Green, Intersection, Sure, Red, Straight, Turning	Pull	Service, Time, Ready, Stay, Take, Route	Hand, Left Hand, Bike, Gear, Line	Turned, Anyway, Button, Man, Seems
Intersection 3	Car(s), Coming, Light(s), Green, Intersection, Red, Turning	Change/ Changing, Right hand, Forward, Making, Notice, Route	Hook, Doing	I'll, Hand, Moving, Gear, Merging, Assume, Stopping	Middle, Walk/ Walking, Able, Check/ Checking, Clear, Flashing, Look/ Looking, Button, Man

Fig. 2. Common and unique concepts across road user groups at each intersection

5 Discussion

The aim of this study was to confirm previous exploratory study findings which indicate that different road users experience the same road situations differently (e.g. Salmon et al, 2013; Walker et al, 2011) and to explore the causes and effects of these differences. The analysis presented demonstrates that driver, cyclist, motorcyclist, and pedestrian situation awareness was different when negotiating the same three intersections. Specifically, different concepts were found in the distinct road users' situation awareness networks, and even when the same concepts were present, the integration with other concepts in the networks was different across the road user groups studied. These findings point to the conclusion that distinct road users experience the same intersections differently to one another. They possess different intersection schema, perform different tasks, interact with the environment differently, and integrate information regarding the intersection situation differently, all of which culminates in a markedly different understanding of the intersection situation.

The next aim was to examine whether these differences are safe or not. Given the nature of the different road users' tasks (e.g. operating a motorcycle versus walking) it is not surprising that their situation awareness differs in some way; however, it is also apparent that some differences may lead to conflicts. Examination of the unique and common concepts (Figure 2) raises some concerns. The unique cyclist concepts 'service', 'stay', 'route' and 'hook' derive from the key decision on approach concerning whether they should negotiate the intersection on the road within the traffic flow, via the pedestrian crossing, or via a hook turn and then also whether they should stay on the road after the intersection or head into the service lane. From a situation awareness perspective, this decision is informed by awareness of the real-time safety risks associated with each path through the intersection and of the ease of taking one option over the others. The result may be that cyclists' attention is taken away from the traffic surrounding them whilst they focus on working out which route through the intersection is the safest and easiest to access. Moreover, once the route through is decided upon, a major maneuver may be required (for example, moving across three lanes of traffic into the service lane, or from the service lane onto the road and into the right hand lane). When this is considered with the fact that the other road users had a high focus on cars, the traffic lights, and the act of turning right, along with the absence of 'check/checking', 'look/looking', and left or right 'hand' side concepts, a major conflict becomes apparent. That is, cyclists are required to focus their attention on situational features other than the traffic and then potentially make a major maneuver in close proximity to the intersection, whereas drivers are not expecting cyclists to make these maneuvers and may not be checking the environment for them.

A similar issue was found with the motorcyclists, who appeared to have a focus on the 'line' to be taken through the intersection, and on the selection of the most appropriate lane in which to negotiate the intersection. Again this represents a key decision point for motorcyclists and also creates the potential for lane change maneuvers in close proximity to the intersection, whereas the driver networks do not contain concepts related to checking and looking for other road users. This represents a potential conflict in that motorcyclists are making maneuvers just prior to the intersection, but drivers may not expect these maneuvers or be on the lookout for motorcyclists.

The evidence suggests, therefore, that the propensity for cyclists and motorcyclists to seek the safest path through the intersection may be raising potential conflicts with drivers. It is apparent that there are two factors creating these potential conflicts: the way in which the intersection is designed, and driving experience. In the case of road design, cyclists and motorcyclists face a key decision in close proximity to the intersection itself, and intersection 'systems' do not support either decision. Moreover, the intersection system does not make other road users (e.g. drivers) aware of the likelihood that cyclists and motorcyclists could potentially make major maneuvers in close proximity to the intersection itself. In the case of road user experience, it appears that drivers are not expecting cyclists and motorcyclists to be maneuvering in and around the intersection, which in turn means they are not looking for them.

One solution is to design intersections that support situation awareness and decision making across all road users and which increase road users' awareness of how other road users behave. For example, it would be useful to ensure that cyclists and motorcyclists decide on the path through the intersection earlier and away from the complex intersection situation. This could be achieved by constraining behavior; for example, taking cyclists and motorcyclists through the intersection via dedicated bicycle and motorcycle lanes. Road signage encouraging drivers to be on the lookout for motorcyclists and cyclists maneuvering across traffic lanes would also be useful. Another solution is the provision of driver training focused on developing an understanding of other road users' behavior. Research has shown that drivers who are also licensed motorcyclists are involved in fewer car-motorcycle collisions than car drivers who do not hold a motorcycle license (Magazzù et al, 2006). The concept of cross mode training (Maguzzù et al, 2006) where different road users receive training in how other road users interpret the road situation and behave in different situations could be useful for developing anticipatory schema of other road users in drivers. The next phase of this research program will explore new intersections design concepts designed to support situationa awareness across all forms of road user.

References

1. Clarke, D.D., Ward, P., Bartle, C., Truman, W.: The role of motorcyclist and other driver behaviour in two types of serious accident in the UK. Accident Analysis & Prevention 39(5), 974–981 (2007)
2. Elvik, R.: Why some road safety problems are more difficult to solve than others. Accident Analysis & Prevention 42(4), 1089–1096 (2010)
3. Magazzù, D., Comelli, M., Marinoni, A.: Are car drivers holding a motorcycle license less responsible for motorcycle-car crash occurrence? A non-parametric approach. Accident Analysis & Prevention 38(2), 365–370 (2006)
4. Neisser, U.: Cognition and reality: principles and implications of cognitive psychology. Freeman, San Francisco (1976)
5. Salmon, P.M., Young, K.L., Cornelissen, M.: Compatible cognition amongst road users: the compatibility of driver, motorcyclist, and cyclist situation awareness. Safety Science 56, 6–17 (2013)
6. Shahar, A., Poulter, D., Clarke, D., Crundall, D.: Motorcyclists' and car drivers' responses to hazards. Transportation Research Part F: Traffic Psychology and Behaviour 13, 243–254 (2010)
7. Walker, G.H., Stanton, N.A., Salmon, P.M.: Cognitive compatibility of motorcyclists and car drivers. Accident Analysis & Prevention 43(3), 878–888 (2011)
8. Wood, J.M., Lacherez, P.F., Marszalek, R.P., King, M.J.: Drivers' and cyclists' experiences of sharing the road: Incidents, attitudes and perceptions of visibility. Accident Analysis & Prevention 41(4), 772–776 (2009)
9. World Health Organisation. Global status report on road safety: time for action. Geneva, World Health Organization (2009)

Part II

Cognitive Issues in Aviation

How Can a Future Safety Net Successfully Detect Conflicting ATC Clearances – Yet Remain Inconspicuous to the Tower Runway Controller? First Results from a SESAR Exercise at Hamburg Airport

Marcus Biella[1], Karsten Straube[1], Marcus Helms[1], Stephen Straub[2], Benjamin Weiß[2], Felix Schmitt[2], Heribert Lafferton[2], Stéphane Dubuisson[3], and Roger Lane[3]

[1] DLR Institute of Flight Guidance, Braunschweig, Germany
{marcus.biella,karsten.straube,marcus.helms}@dlr.de
[2] DFS Deutsche Flugsicherung GmbH, Langen, Germany
{stephen.straub,benjamin.weiss,felix.schmitt,
heribert.lafferton}@dfs.de
[3] EUROCONTROL Experimental Centre, Brétigny, France
{stephane.dubuisson,roger.lane}@eurocontrol.int

Abstract. To increase runway safety a new safety net for Tower Runway Controllers was developed which detects if controllers give a clearance to an aircraft or vehicle contradictory to another clearance already given to another mobile. In a shadow mode validation exercise with eleven controllers at the operational environment of the airport Hamburg (Germany) operational feasibility was tested in order to clarify if operational requirements in terms of usability are fulfilled. At the same time operational improvements regarding safety were studied e.g. if the new safety net detects all conflicts and if nuisance alerts are suppressed.

Keywords: Safety, Air Traffic Control, Airport Operations, Runway Incursions, SESAR.

1 Introduction

On 31st July 2008 the flight crew of a Fairchild SA227 Metroliner III contacted the Zurich Tower Runway Controller to inquire if it is an option to land on runway 16 instead of runway 14. This is a frequent request by pilots because landing on runway 16 allows a shorter way to the terminal; in doing so runway 28 has to be crossed. In this case, the controller cleared the aircraft to land on runway 16. Shortly after that an Airbus A319-100 taxied to runway 28 for departure and was cleared to line-up and hold. Then the tower controller ordered the flight crew of a helicopter to hold position and not to cross the departure path of runway 28 because of the intended departure of the Airbus. Subsequently the controller cleared the Airbus for take-off. Suddenly the controller noticed that the landed Metroliner was still rolling on runway 16 while the Airbus was starting to accelerate down runway 28. This situation was obviously

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 65–74, 2013.
© Springer-Verlag Berlin Heidelberg 2013

dangerous because both aircraft approached the intersection. Upon notice the Airbus crew was ordered to abort take-off [1] [2].

This incident illustrates the risk of a conflicting air traffic control (ATC) clearance and its potential consequence. In this case a fatal collision could be prevented. However a conflicting ATC clearance given on 1st February 1991 in Los Angeles led to a collision between two aircraft where 34 people lost their lives [3].

In 2011, altogether 66 runway incursions - not leading to an accident - have been reported in Germany. Only 12% of these rare events were caused by controllers [4] but it can be presumed that conflicting clearances were given before. In order to prevent this unique cause for a potentially dangerous situation, an additional "Conflicting ATC Clearances safety net" was created. This safety net detects if clearances given to aircraft or vehicles could lead to an unsafe situation.

2 Concept

2.1 Background

The "Single European Sky Air Traffic Management Research" (SESAR) programme is one of the most ambitious research and development projects ever launched by the European Union. The programme is the technological and operational dimension of the Single European Sky (SES) initiative to meet future capacity and air safety needs, i.e. an improvement of safety by a factor of 10 [5]. In this context runway incursions shall be reduced. They are defined by International Civil Aviation Organization (ICAO) as "any occurrences at an aerodrome involving the incorrect presence of an aircraft, vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft" [6].

The "Conflicting ATC Clearances safety net" concept as well as the prototypes used for validation were developed under the SESAR programme and co-financed by the European Community and EUROCONTROL. Work on the report of the final validation exercise finished at the end of February 2013, therefore it should be stressed that the source material for the results published in this paper has not been approved by the SESAR Joint Undertaking (SJU) yet. The sole responsibility of this paper lies with the authors. The SJU and its founding members are not responsible for any use that may be made of the information contained herein.

2.2 The "Conflicting ATC Clearances Safety Net" Concept

Currently the only safety net available to Tower Runway Controllers is the Runway Incursion Monitoring System (RIMS). It uses Advanced Surface Movement Guidance and Control System (A-SMGCS) Surveillance data to detect dangerous situations within the Runway Protection Area. Detections and subsequent alerts to controllers are provided at the very last moment and require immediate reaction.

The new "Conflicting ATC Clearances safety net" will not replace the existing RIMS but is intended as an additional layer of safety. It will detect conflicting ATC clearances much earlier – when the controller inputs clearances into the Electronic Flight Strips (EFS), which are already in operational use in many control towers. To do so, it will perform crosschecks with the clearances input on the EFS, and in most cases the aircraft position, to see if the given inputs violate the rules and procedures at the concerned airport, which could lead to a hazardous situation [7]. In the introductory example at Zurich airport, a conflicting "land vs. take-off" alert would have been given.

2.3 Recommendations from Real Time Simulation

A first prototype had already been successfully tested in a SESAR real time simulation exercise [8] with three air traffic control officers (ATCOs) in 2011. It was recommended that the detection of take-off versus line-up and line-up versus take-off should be fine-tuned so that the system takes into account the line-up point of the taxiing aircraft and not the actual position of the aircraft. This would prevent a so-called nuisance alert that is triggered when the aircraft that was due to line-up would be still taxiing on the taxiway parallel to the runway but is in front of the aircraft taking off, but the line-up point is behind the aircraft taking off.

Furthermore it was recommended to make the safety net more proactive instead of reactive. A "what-if tool" would be capable to highlight potential conflicting ATC clearances before these clearances are actually given. This would eliminate alerts and therefore the need for the ATCO to revise clearances.

2.4 Description of DFS's Prototype

The prototype to support the final validation was developed by DFS based on the flight data processing system (FDPS) *SHOWTIME* including electronic flight strips, and on the surveillance data processing system (SDPS) *PHOENIX*. For a detailed description of the prototype's detection logic for conflicting ATC clearances, please refer to [9]. The present section briefly summarizes some aspects of the prototype, focusing mainly on its human machine interface (HMI).

Conflicting ATC clearance alerts are displayed both in the FDPS HMI (Figure 1) and the SDPS HMI (Figure 2, left) for both the tower runway and ground controller. As can be seen in the figures, the type of conflict is displayed both on the flight strip and on the SDPS target label. An alert may be acknowledged by clicking on the "ACK" part of the strip on the right; this makes the alert display less obnoxious, but does not suppress it completely.

Clearances are entered into the electronic flight strips using a mouse. In particular, the next clearance (according to standard procedure) can be entered by clicking on the part of the strip that displays the currently active clearance (the square symbols on the very left in Figure 1). Taking back an entered clearance is possible via a menu, or – more quickly – using a special undo button.

Fig. 1. Conflicting take-off vs. land Clearance ("TOF/LND") in SHOWTIME

In developing the prototype, recommendations from the previous real time simulation (cf. section 2.3) were taken into account. In particular, the route-based conflicting clearance detection mechanism [9] helps in avoiding certain nuisance that occurred in the real time simulation (e.g. clearances that would be identified as conflicting clearances although the trajectories would never cross each other due to the positions or cleared routes). The prototype generates the needed ground routes automatically for all aircraft with a flight plan. If necessary this plan can be changed manually by the ATCO.

The core detection logic uses routes as inputs and – roughly speaking – checks whether the cleared routes of two mobiles overlap somewhere on a runway. An example is shown in Figure 2: The system identified the clearances line-up vs. land as conflicting, because the aircraft's routes overlap on the runway. The conflict disappears as soon as the landing aircraft passes the runway entry of the second aircraft.

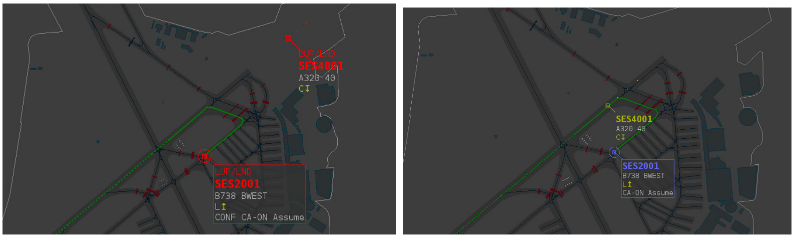


Fig. 2. Conflicting line-up vs. land ("LUP/LND") clearance in PHOENIX (left), neutralized after SES4001 passed the entryway of SES2001

Fig. 3. Predictive conflict indication: two possible conflicting clearances indicated by a red dot in the flight strips UAE25 and DLH1MA

Furthermore a **predictive indication** is introduced as well. It is integrated on the left side of the flight strip. If the next clearance (according to standard procedures) would currently cause a clearance conflict, this is indicated by a little red dot. In contrast, a green dot indicates that giving this next clearance would currently not cause a clearance conflict. For example as shown in Figure 3, giving a line-up clearance to UAE25 or to DLH1MA would create a clearance conflict, whereas giving the clearance to GEC9834 would not.

2.5 Validation Objectives for Shadow Mode Trials

First of all, the operational feasibility in terms of fulfillment of operational requirements (as stated in the Operational Services and Environmental Description (OSED)

[7]) had to be checked, mainly by controllers' feedback on the usability of the different alerts and the HMI design.

Secondly operational improvements in terms of safety had to be studied. It was crucial that the new safety net detected all conflicting situations. Furthermore the safety net should allow the controller to solve detected situations timely. In addition to that the false alert rate had to be acceptable for the controller. Finally it should be tested if some detections were considered as nuisance alerts by the controllers.

3 Method

The shadow mode trials were performed with different controller teams each day at the airport environment in Hamburg between the 26th and 30th November 2012. A controller team consisted of a Ground and Runway controller.

3.1 Sample

In total eleven Tower Controllers took part in the study. Six were active Hamburg controllers, one was recently retired in 2011. Additional one ATCO each came from the airports in Hamburg Finkenwerder, Leipzig (both Germany), Klagenfurt (Austria) and Lamezia Terme (Italy). Eight of them were male, three were female. Their average age was 35.5 years (standard deviation: 7.3 years). For the six active Hamburg controllers the mean reported experience was 6.3 years (standard deviation: 4.7 years).

3.2 Shadow Mode Environment

The exercise was located outside the control tower environment to not interfere or disturb the active controllers and pilots at the time. All data was copied and re-routed to a separate, temporary control room set up for the duration of the exercise.

3.3 Traffic

Real life traffic of the Hamburg Airport was used. Additional synthetic traffic was produced to create pre-conditions for conflicting clearances. ATCOs were informed that these synthetic targets could be injected to increase the number of sufficient critical situations in the trials.

3.4 Task

Due to the nature of a shadow mode trial both ATCOs of a team had to act as if they were in charge but without any intervention to the real traffic. One of the two ATCOs started as tower runway controller, assisted by a technical supporter from DFS on his left, and the validation supervisor on his right. The following clearances of the tower controller were part of the exercise: line-up, take-off, land and cross runway.

Any combination of these clearances as defined in [7] shall trigger an alarm. The other ATCO had to act as a ground controller, dealing with all other clearances. Together with the co-validation supervisor they created potential conflicting situations for the Runway Controller.

The ATCO was briefed to make an input to the electronic flight strip (EFS) for an aircraft in accordance to a clearance by the real operational ATCO in the Hamburg control tower. The validation supervisor identified a second aircraft and asked the ATCO in the validation scenario to give now a pre-defined conflicting ATC clearance. For example, the ATCO made a take-off clearance input on the EFS for an aircraft. After that he gave – on order of the validation supervisor – a cross clearance to another aircraft. This resulted in a "take-off vs. cross" conflict.

The first part of each day was dedicated to brief both ATCOs on the scope and objectives of the shadow mode trials and to train them on the equipment and environment. Most of them already had a pre-training on DFS's FDPS PHOENIX and DFS's SDPS SHOWTIME the week before as SHOWTIME was not in operational use at Hamburg tower resp. PHOENIX is used in another version and configuration.

3.5 Scenarios

Three shadow mode trials lasting seventy minutes each were performed during the day. After 35 minutes ATCOs were told to switch roles (from tower to ground controller and vice versa). The first of the three shadow mode trials focused on scenarios with the first clearance being given was "land". The second shadow mode trial took into account scenarios with the first clearance being given was "line-up" or "take-off". The third and final shadow mode trial dealt mainly with cross scenarios and any other conflicting clearance which had not been tested before or which was regarded as particularly interesting.

3.6 Measurements

To check if the operational requirements of the OSED [7] were fulfilled a Post Trials Questionnaire was tailor-made [10] to capture ATCOs' feedback and comments. Each ATCO had to complete the questionnaire in an excel spreadsheet after the last of the three shadow mode trials. Controllers were asked how far they could agree to or not by choosing answers amongst six categories ranging from 1 (strongly disagree) to 6 (strongly agree).

Mean values (M) and standard deviations (SD) were calculated to describe the result. Furthermore, by use of a binomial test [11] for a single sample size, each item was proven for its statistical significance with an expected mean value = 3.5, test ratio: .50 and alpha = 0.05.

Furthermore the triggering of the correct type of alert and the amount of false alerts was measured by observations of experts.

4 Results

4.1 Operational Feasibility

ATCOs agreed in the Post Trials Questionnaire that they appreciate the conflict information (M=4.7 on a six point Likert scale, SD=0.9, N=10, p=0.02). In the following further results from the questionnaire are reported [10], arranged for different types of alerts:

Detailed feedback for conflicting clearances alerts including a **landing aircraft** was given by the ATCOs. The alert for conflicting clearances with two landing aircraft was rated as usable (M=4.9, SD= 0.7, N=10, p=0.00), especially when two aircraft received landing clearances on the same runway. The alert in case of an aircraft cleared to land and another aircraft being cleared for line-up was rated as usable as well (M=5.3, SD=0.6, N=11, p=0.00), especially when the aircraft receiving the line-up clearance was in front of the aircraft receiving the landing clearance on the same runway. The alert was rated as particularly usable when the aircraft receiving the clearances were on the opposite ends of the same runway. The alert for the conflicting clearances “land vs. take-off” (M=5.1, SD=0.6, N=10, p=0.00) was usable for the ATCOs as well. Notable situations to be mentioned here are two aircraft on the same runway; respectively two aircraft at opposite ends of the same runway. ATCOs also agreed that “land vs. cross” alerts were usable (M=4.7, SD=1.2, N=11, p=0.01). Of particular importance are situations when an aircraft receiving the cross clearance was in front of the aircraft receiving the landing clearance on the same runway.

Further positive results were gained for alerts in case of conflicting clearances including a **line-up clearance**. Alerts for two aircraft being cleared for line-up were usable according to the ATCOs (M=5.3, SD=0.5, N=6, p=0.03) especially when both aircraft were on the same or adjacent holding points on the same runway when multiple line-up was not authorized. Furthermore the alert was rated as usable when holding points were opposite on the same runway. An alert for “line-up vs. take-off” was also rated as usable (M=4.5, SD=1.0, N=10, p=0.02), especially when the aircraft receiving the line-up clearance was in front of the aircraft receiving the take-off clearance on the same runway. Furthermore the alert is usable when both aircraft were on the opposite ends of the same runway. The alert “line-up vs. cross” was usable as well (M=4.7, SD=1.1, N=10, p=0.02), for example when holding points were opposing on the same runway.

ATCOs also agreed on the usability of alerts in case of conflicting clearances including an **aircraft being cleared for take-off**. An alert for two aircraft receiving take-off clearances was rated as usable (M=4.9, SD=0.7, N=10, p=0.02), especially for two aircraft on the same runway or at opposite ends of the runway. Usable were also alerts with cleared aircraft on different but intersecting runways when aircraft trajectories were converging. The alert “take-off vs. cross” was usable as well (M=5.2, SD=0.4, N=10, p=0.00), especially when an aircraft receiving the cross clearance was in front of the aircraft receiving the take-off clearance on the same runway.

In conclusion ATCOs agreed that the alerts for conflicting clearances with two aircraft **cleared for cross** were usable as well (M=4.8, SD=1.1, N=11, p=0.01) especially when holding points were opposing on the same runway.

ATCOs gave positive **feedback for the HMI** design aspects. They agreed that the configuration of the alert window was fine with them regarding size (M=4.7, SD=0.6, N=11, p=0.01), the use of the alert color "red" (M=4.9, SD=0.8, N=11, p=0.01), and contrast (M=4.8, SD=0.4, N=11, p=0.00). Furthermore audio alarms were rated as usable (M=4.8, SD=0.4, N=10, p=0.00).

4.2 Operational Improvements in Terms of Safety

Detection of Conflicting Situations. Based on observation by experts the correct type of alert was triggered in each case. In detail, the following alerts were triggered successfully during the week of shadow mode testing: 55 land vs. land; 55 land vs. line-up; 96 land vs. take-off; 25 land vs. cross; 35 line-up vs. line-up; 27 line-up vs. take-off; 18 line-up vs. cross; 39 take-off vs. take-off; 25 take-off vs. cross; and 4 cross vs. cross [10].

In addition all ATCOs emphasized that no alerts were missing in the different trials. It could be shown that multiple alerts with more than two aircraft can be displayed comprehensibly. ATCOs state that the alert indication for these rare events can be improved [10].

Timely Detection of Alerts. There is no doubt among the ATCOs that the alerts are generally displayed in time (M=5.0 on a six point Likert scale, SD=0.5, N=9, p=0.00) [10].

Acceptability of False Alert Rate. Based on observation by experts no alerts were given by the system in case that no conflict existed. Therefore no false alerts can be reported [10].

Absence of Nuisance Alerts. ATCOs were asked if alerts were given in situations where the alert is not necessary according to (local) procedures. ATCOs agreed in the Post Trials Questionnaire that the number of nuisance alerts was acceptable (M=4.8 on a six point Likert scale, SD=1.2, N=8, p=0.07 indicating a statistically significant trend). Furthermore the number of alerts that were displayed "too early" was sufficiently low (M=5.3, SD=0.5, N=6, p=0.03) [10].

ATCOs reported that two "line-up vs. cross" alerts were not necessary because the breath of these particular two taxiways allows a simultaneous line-up and cross of two aircraft.

5 Discussion

Overall the validation can be considered as very successful. DFS had provided a well working safety net [9] which was updated and fine-tuned according to the previous validation results of a real time simulation [8] and the updated Operational Services and Environment Description documentation for the safety net [7]. The technical feasibility of the safety net within a real airport environment could be shown. Response of the new safety net was faster than required. The display of alerts simultaneously on SDPS and FDPS and the use of an audio alert were appreciated by the ATCOs [10].

ATCOs' feedback as given in the questionnaires and debriefing was very positive regarding the new safety net. According to the ATCOs and the observers every expected alert was generated and displayed on time by the system. No false alerts were observed during the trials. Operational improvements in terms of safety were indicated by the ATCOs in their comments and in the questionnaire results. Especially the added value of a predictive information tool contributed to this result. It was especially appreciated by the ATCOs because it reduces the number of situations in which the controller would have to *react* on an already given conflicting clearance to a minimum. The implementation of the safety net is capable to assist the ATCOs to perform their tasks more safely while maintaining the efficiency of the airport operations.

The concept in general was considered to be a useful predictive safety support tool that would work in conjunction with additional safety nets (e.g. RIMS).

In the next step the use of the underlying routing function as part of the concept will be discussed because its added value to suppress nuisance alerts was shown in the Hamburg shadow mode trials.

Moreover the interaction of different safety nets should be studied, namely the new developments for Conflicting ATC Clearances plus an additional Conformance Monitoring tool and RIMS which is already in operational use at several airports [10]. Firstly the priority of alerts has to be identified. Secondly it has to be clarified which type of alert should be triggered at which time. In this context it is necessary to discuss if a simultaneous display of different alerts is required or if one safety net should be capable to overwrite alerts given by another safety net. For example a RIMS alert should be given more importance than a conflicting clearance alert. Results from exercises with the simultaneous use of these three safety nets do not exist to give indications in this context by now [10].

Furthermore the necessity of additional real time simulations was stressed by the validation team, ATCOs and observers. They should involve the above mentioned safety nets, and include visual flight rules traffic and helicopters to test more complex situations (e.g. traffic without flight plans). This will certainly increase workload for the controller and probably create more safety critical situations. Conflicting *taxi* clearances could be tested in this environment as well [10].

In the validation exercise conflicting ATC clearances were provoked on purpose to test the concept. However, in the real operational environment the new safety net acts as a kind of watchdog in the background, visible only in the rare occasion of a clearance conflict. It would be a revealing test to let the system run silently and unattendedly in shadow mode linked to the EFS inputs of the real operational tower controllers. This would allow one to measure how often conflicting clearance alerts occur in practice with real controllers acting normally (the goal being that this happens almost never).

References

1. Hradecky, S.: Report: Air Berlin A319 and OLT SW4 at Zurich on Jul 31st 2008, conflicting ATC clearances (2009), http://avherald.com/h?article=42246863 (February 12, 2013)
2. Schlussbericht Nr. 2047 des Büros für Flugunfalluntersuchungen, http://www.bfu.admin.ch/common/pdf/airprox/2047_d.pdf (January 22, 2013)
3. National Transport Safety Board. Aircraft Accident Report Runway. PB91-910409 NTSB/AAR-91/08
4. Deutsche Flugsicherung: Luftverkehr in Deutschland. Mobilitätsbericht (2011), http://www.dfs.de/dfs/internet_2008/module/presse/dfs_mobilitaetsbericht_2011_de.pdf
5. SESAR: The Roadmap for sustainable Air Traffic Management: European ATM Master Plan (2nd edn.) (2012)
6. ICAO Document 4444 - Procedures for Air Navigation Services -Air Traffic Management (14th edn.) (2001)
7. SESAR P06.07.01 D16 Updated Operational Service and Environment Definition (OSED) for Conflicting ATC Clearances, Version 1.0 (2012)
8. SESAR P06.07.01 D15 V2 VALR V2 ATC Conflicting Clearances Validation Report (VALR) – Version 1.0 (2012)
9. Weiß, B., Centarti, F., Schmitt, F., Straub, S.: Route-Based Detection of Conflicting ATC Clearances on Airports. In: International Symposium on Enhanced Solutions for Aircraft and Vehicle Surveillance Applications (Proceedings ESAVS 2013) (2013)
10. SESAR P06.07.01 D19 V2 VALR V3 ATC Conflicting Clearances Validation Report (VALR) – Version 00.01.00 (2013)
11. Handbook of Biological Statistics: Exact test for goodness-of-fit, http://udel.edu/~mcdonald/statexactbin.html (January 18, 2013)

The Analysis of Safety Recommendation and Human Error Prevention Strategies in Flight Operations

Jeng-Chung Chen[1], Chia-Fen Chi[1], and Wen-Chin Li[2]

[1] Department of Industrial Management, National Taiwan University of Science & Technology
43, Keelung Road Section 4, Taipei 106, Taiwan
[2] Psychology Department, National Defense University,
Beitou District, Taipei City 112, Taiwan
D10101005@mail.ntust.edu.tw

Abstract. This study applied Human Factors Intervention Matrix (HFIX) framework described by Wiegmann and Shappell. METHOD: The data set comprised of 31 incident investigation reports taken place between 2009 and 2011 and included 182 unique safety recommendations to reduce human errors. The results indicated that major recommendations were directed at organizational/ administrative and human/ crew approaches and the most effective interventions concentrated on decision errors and violations. This study has demonstrated that the HFIX framework can be applied to improve human errors by five different approaches. It also has suggested that decision error and violations are critical issues of flight safety and these can be improved by training and organizational administration.

Keywords: Accident Prevention, Human Errors, Human Factors Analysis and Classification System, Human Factors Intervention Matrix.

1 Introduction

According to an analysis of global safety data involving commercial air transport aircraft with a maximum certificated take-off mass of more than 2,250 kilograms, the accident rate are between 3.5ppm and 4.5ppm during 2002 and 2011[1], but human factors doesn't only decent but also become the major cause of aviation mishaps[2]. Due to understand the fundamental nature and underlying causes of aircrew error and unsafe acts, aviation psychologists proposed varied perspectives on pilots' performance in the cockpit[3]. The cognitive perspective asserts the pilot's mind can be conceptualized as an information-processing system. When information from the environment makes contact with one of the senses, it progresses through a series of stage or mental operations to produce a response or action[4]. Therefore pilot errors are believed to occur when one or more of these mental operations fail to process information appropriately[5]. The ergonomics and systems perspective asserts pilot performance is the result of a complex interaction among several factors [7]. Flying an airplane is a very complicated and dynamic task and pilots interact with high-tech airplanes and cockpit equipment, as well as must safely operate their aircrafts in all

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 75–84, 2013.
© Springer-Verlag Berlin Heidelberg 2013

types of weather conditions. Therefore pilot error is believed to occur when there is a mismatch or breakdown in the interface between the aircrew and the technology. The psychosocial perspective asserts flight operations are best viewed as a social endeavor that requires aircrew to interact with other flight support personnel. Therefore, aircrew performance is directly influenced by the nature or quality of these interactions [8]. The organizational behavior perspective asserts that pilot performance must be viewed in term of the organizational context in which it takes place [9]. Aviation organizations are responsible for ensuring that only those pilots with the 'right stuff' are hired to fly their aircraft. In addition, these organizations are also responsible for instituting appropriate procedures that ensure safe operations of the aircraft [10]. Therefore, aircrew errors and subsequent accident are believed to occur when managers and supervisors fail to set up basic conditions within the organization that promote flight safety [11].

Regarding with analysis framework and strategy of human factors, the Human Factors Analysis and Classification System (HFACS) is the most popular framework for investigating human errors in flight operations. HFACS proposed by Wiegmann and Shappell [12] were originally designed and developed as a generic human error framework for investigating and analyzing human error accidents in US military aviation operations. HFACS addresses human errors at four levels. Level 1 (unsafe acts of operators - active failures) is the level at which the majority of accident investigations are focused. Failures at this level can be classified into two categories; errors and violations. Level 2 (preconditions for unsafe acts -latent/active failures) addresses the latent failures within the causal sequence of events as well as the more obvious active failures. It also describes the substandard conditions of operators and the substandard practices that they perform. Level 3 (unsafe supervision - latent failures) traces the causal chain of events producing unsafe acts up to the level of the front-line supervisors. Level 4 (organizational influences -latent failures) describes the contributions of the most elusive of these latent failures, fallible decisions of upper levels of management which directly affect supervisory practices, as well as the conditions and actions of front-line operators.

Due to develop efficient intervention strategy of human errors, Wiegmann and Shappell [13]proposed Human Factors Intervention Matrix (HFIX) framework. HFIX matrix pits the level-1 of HFACS (unsafe acts) against the five different safety approaches. The unsafe acts were described as operators commit errors or violation which includes decision errors, skill-based errors, perceptual errors and violations. On the other hand, improving these unsafe acts can used following approach, organizational/administrative, human/crew, technology/engineering, task/mission and operational/ physical environment. The HFIX framework is described diagrammatically in Figure 1.

Wiegmann & Shappell [13]has applied HFIX framework to analyze safety recommendations of 147 commercial aviation accident reports which were adopted from National Transportation Safety Board (NTSB) accident databases and taken between 1998 and 2004. These investigators' recommendations can be clustered on category of organizational/administrative, human/crew, technology/engineering, task/mission. In addition, these recommendations in aviation tend to focus more on improving the design of system or some manner of organizational change rather than focusing on operational personnel.

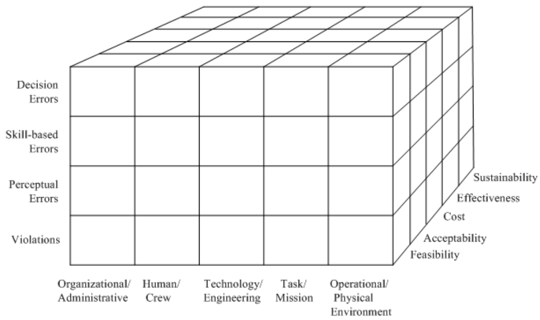


Fig. 1. Human Factors Intervention Matrix (HFIX) framework[13]

National culture has been implicated as a factor in many aircraft accidents. Merritt [14]called the failure to take into account the effects of national culture "cultural imperialism." Hofstede [15] pointed out that national cultures vary on dimensions such as power-distance, uncertainty avoidance, individualism, masculinity femininity, and long-term vs. short-term five areas which can affect interactions in the cockpit and which are known to have an impact on safety. Power Distance has been defined as the extent to which the less powerful members of organizations and institutions (like the family) accept and expect that power is distributed unequally. In Hofstede et al. [16] Power Distance Index scores are listed for 76 countries; they tend to be higher for East European, Latin, Asian and African countries and lower for Germanic and English-speaking Western countries. Uncertainty avoidance is not the same as risk avoidance; it deals with a society's tolerance for ambiguity. It indicates to what extent a culture programs its members to feel either uncomfortable or comfortable in unstructured situations. In Hofstede Uncertainty Avoidance Index scores, they tend to be higher in East and Central European countries, in Latin countries, in Japan and in German speaking countries, lower in English speaking, Nordic and Chinese culture countries. Individualism on the one side versus its opposite, Collectivism, as a societal, not an individual characteristic, is the degree to which people in a society are integrated into groups. In Hofstede Individualism Index scores, individualism tends to prevail in developed and Western countries, while collectivism prevails in less developed and Eastern countries; Japan takes a middle position on this dimension. The dimension of masculinity-femininity evaluates the delegation of responsibilities between males and females in societies. In masculine societies, the focus is on competition and hardship in ideas and materiality, in feminine societies, a person who is treated with injustice are paid attention to. In Hofstede Masculinity versus Femininity Index scores, masculinity is high in Japan, in German speaking countries, and in some Latin countries like Italy and Mexico; it is moderately high in English speaking Western countries; it is low in Nordic countries and in the Netherlands and moderately low in some Latin and Asian countries like France, Spain, Portugal, Chile, Korea and Thailand. Long term perspective are societies which encourage righteous (stamina and economics) for future rewards. Short term perspective describes societies that encourage righteous regarding now and the past like as respect regarding customs and following society's necessities. Both two positive and negative perspective of this dimension are found in Confucius teachings and are applicable for all companies. Long-term oriented are East Asian countries, followed by Eastern- and Central Europe. A medium term orientation is found in South- and North-European and South Asian countries. Short-term oriented are U.S.A. and Australia, Latin American, African and Muslim countries.

2 Methods

Data. The data were derived from the narrative description of incident occurring in commercial airlines between 2009 and 2011. The data set comprised of 31 incident reports, these report included specific causes and 72 safety recommendations which was suggested by investigators to improve specific human error. After compiling, the final list was consisted of 182 unique recommendations which preserve the intent of original investigators and readiness for future analysis.

Classification Framework. This empirical study used the HFIX framework as described in Wiegmann and Shappell [13]. The front side of HFIX pits the unsafe acts against the five different approaches. The unsafe acts were described as operators commit active failure which includes decision errors, skill-based errors, perceptual errors and violation. On the other hand, reducing these unsafe acts can adopt following approach, organizational/administrative, human/crew, technology/engineering, task/mission and operational/ physical environment.

Coding Process. Each incident report was coded by two participants(one with doctoral-level background in aviation psychology, the other one with a graduate background in management). Both participants were trained together on the HFIX framework for 10 hours to ensure that they achieved a detailed and accurate understanding to the categories of the HFIX. During classification, each incident and its cause were presented at first. Then each recommendation was independently code into categories by two participants based on their similarities. Moreover, participants were instructed to pair any HFIX intervention approach with HFACS unsafe acts categories which can be remediated by specific intervention. After the initial rating, two participants were permitted to discuss their classification to resolve any differences. A final consensus classification for each recommendation was then provided for further analysis.

3 Results

Sample Characteristic. According to these 31 incident reports, there is no fatality and also aircrafts doesn't damage. Among these 31 incidents, 18(58%) occurrences were observed by manager/ supervisor , the other 13(42%) were caught by feedback of information system. Aircrew were involved 14 incidents, flight attendants involved 4 incidents, and the other flight support personnel involved 13 incidents. From the pur pose of 182 unique recommendations, 52(29%) of recommendations can apply with pilots, the other 32(18%) and 98(53%) can implement with flight attendants and the other flight support personnel respectively.

The Frequency and Percentage of Category in HFIX Framework. A total of 182 recommendations were analyzed and recorded within the HFIX framework and results were showed in Table1. The organizational/administrative approach was involved 76 (41.1%) recommendations; 63 (35.2%) of recommendations was directed at human/crew approach and followed by approach of task/ mission 21 (11.4%), the others 22 (12.1%) recommendations involved technologic/ engineering, and operational/ physical environment. Initial results can explain consideration of these investigators, which the first step of reducing unsafe acts should deal with organizational/administrative issues which include categories of level-4(organizational influence) and level-3(unsafe supervision) within HFACS framework. From effectiveness of mitigating unsafe acts which is y-axis of HFIX framework, majority concentrates on decision errors (83%) and violations (80.2%). In contrast, skill-based errors were associated with 57% of the recommendations followed by perceptual errors (3%).

Table 1.

Category of HFIX		Frequency Of recommendation	Percentage Of recommendation	Inter-rater Reliability	
				Cohen's Kappa	Percentage agreement
x-axis	Organizational/Administrative	76	41.1%	0.673	84.1%
	Human/ Crew	64	35.2%	0.569	81.3%
	Technology/Engineering	20	11.0%	0.897	97.8%
	Task/Mission	21	11.4%	0.509	87.9%
	Operational/Physical Environment	2	1.1%	0.330	98.9%
y-axis	Decision error	151	83%	0.656	90.7%
	Skill-based error	103	56.6%	0.815	90.7%
	Perceptual error	7	3.8%	0.608	96.7%
	Violation	146	80.2%	0.512	78.0%

Inter-rater Reliability of HFIX Classification. The inter-rater reliabilities assessed using Cohen's Kappa ranged between 0.33 and 0.897, a range of values spanning between moderate agreement and substantial agreement. Five HFIX categories exceeded a Kappa of 0.60, which indicated substantial agreement. Three categories had Kappa value 0.41 and 0.6 indicating moderate levels of agreement [17]. Inter-rater reliabilities calculated as a simple percentage rate of agreement obtained reliability figures between 78% and 98.9%, also indicating acceptable reliability.

4 Discussion

How does Active Failure Impact with flight Safety. The y-axis of HFIX framework comprised decision errors, skill-base errors, perceptual errors and violations, which are identified with active failures from Reason's [11]system-wide model of human error. Active failures are associated with the performance of first-line operators in complex systems and latent failures are characterized as inadequacies or misspecifications which might lie dormant within a system for a long time and are only triggered when combined with other factors to breach the system's defenses.

From the material of this study, 182 recommendations had the highest ratio(83%) at mitigating decision errors. This result can explained that decision errors are principal cause of aviation accident and it usually attribute a fault to flight safety [18][19]. This study found that aircrew and other flight support personnel commit decision errors including inadequate risk assessment, task misprioritization, ignored caution/warning sign, selecting wrong course of action in a time-constrained environment. Moreover, recommendations of reducing violation is rated 80% in this study. Violations are factors in a incident when the actions of the operator represent willful disregard for rule and instructions and lead to an unsafe situation, such as aircrew didn't obeyed SOP to reconfirm the notice of dispatcher before take-off and piloted wrong aircraft for flight mission, cargo forwarders used inadequate check-list to load wrong container and made aircraft flied with unbalance weight and etc. Violation can be caught by supervision and information system in organization, but these failures could become disaster if upper level managers would not deal with it in time.

Table 2.

Intervention Approaches of HFIX	Recommendation's Percentage of current study	Recommendation's Percentage of NTSB database
Organizational/Administrative	41.1%	34.18%
Human/ Crew	35.2%	11.47%
Technology/Engineering	11.0%	31.20%
Task/Mission	11.4%	23.16%
Operational/physical environment	1.1%	N/A

Note: The recommendation's percentage of USA National Transportation Safety Board(NTSB) is taken from Shappell and Wiegmann ' A methodology for assessing safety program targeting human error in aviation'[13]

Results of comparing this study data with National Transport Safe Board U.S.(NTSB) was showed at Table 2. Organizational/administrational approaches get the highest ratio from both investigators' recommendations to mitigate human errors, but rank of other approaches are difference obviously. This phenomenon may occasion with national culture [16]. Asian countries tent to higher power distance and collectivism that subordinates have high dependency to superiors and don't disagree with superiors directly. There is a real instance of this study, which can explain effect of national culture. At taxing operations in unfamiliar airport, the pilot officer paid concentration with circumstance of runway and neglected aircraft approaching the stop line gradually. But the co-pilot may be affected by deference of hierarchy and experience and didn't remind pilot officer. The aircraft passed stop line over 3 inches finally and made the other aircraft which is preparing for landing has to go around. This study found that implementing the recommendations of "human/crew" approaches can improve reaction and communication of aircrew affected national culture.

Analysis of Recommendation under HFIX Framework. Depend on past studies, the objective was to provide probabilities for the co-occurrence of categories across adjacent levels of the HFACS to establish how factors in the upper (organizational) levels in the framework affect categories in lower (operational) levels [12]. That is fallible decisions in upper management levels affect supervisory practices directly(level-3), thereby creating the psychological preconditions for unsafe acts and hence indirectly impairing the performance of pilots, ultimately leading to accidents. This study found that organizational/ administrative and human/crew approaches has the highest ratio. Obviously, investigators assert remedial solution for unsafe acts by way of organizational/ administrative approaches fall in categories of level-4 organizational influences, level-3 unsafe supervision and level-2 preconditions of unsafe acts.

This study has demonstrated that the HFIX framework can be applied to mitigating human errors with five different approaches and diversify remedial solutions were showed in Table 3.

Table 3.

Intervention Approach of HFIX		Frequency	Percentage
Organizational/ Administrative	Rule, regulations and policies	15	8.2%
	Information management and communication	59	32.0%
	Research and special study	0	0.0%
	Human resource management	1	0.5%
Human/ Crew	Selection	1	0.5%
	Training	27	14.8%
	Job assignment	2	1.1%
	Motivation	19	10.4%
	Crew interaction	3	1.6%
Technology/ Engineering	Design/Repair	21	11.5%
	Inspection	3	1.6%
Task/Mission	Procedure	28	15.2%
	Manual	1	0.5%
Operational/Physical Environment	Physical environment	2	1.1%
	Technological environment	0	0.0%
	Total	182	100%

Organizational/Administrative Approaches. This approach focus on the way in which the organization and supervisors may need to change to improve safety [20]. Among 182 unique recommendations of this study, there are 59(32%) recommendations belong to category of “information management and communication” and investigators suggested that improving managers/supervisors involvement and oversight/monitor can change safety culture of organization, enhance awareness of hazard and risk of front-line operators. Moreover, 15(8.2%) recommendations were classify into ”Rule, regulations and policies" category and these investigators suggested that managers/supervisors should pay more concentration on establishing, issuing, modifying and revising the navigation route map, standard operation process(SOP) and regulations to ensure those items not out-of-date. Subsequently, the recommendations of “human resource management” were to apply policies of human resource management, such as selection, incentive, and promotion, to built upon safety culture of organization.

Based on the assertion of organization behavior perspective [9], first-line operators of flight mission occur errors and subsequent accident when managers and/or supervisors fail to set up basic conditions within the organization that promote flight safety [11]. Therefore these recommendations of organizational/administrative approach can be implemented to create restricted mechanism and the upper level managers/supervisors involvement and incentive can reinforce safety culture of organization.

Human/crew Approaches. Flight operations requires aircrew to interact with other flight support personnel, such as air traffic controllers, dispatchers, ground crew, maintenance personnel, and flight attendants. Therefore, 64 (35.2%) recommendations were classified in these approaches and these recommendations focus on change or improve the individual worker or work team to enhance situation awareness and job satisfaction. This study found that “training” has the highest rate(14.8%) of recommendations in human/crew approaches and investigators had suggestions such as holding training course with qualified trainer periodically, enhancing front-line operators situation awareness and professional skill with scenario-base training. Other recommendations of this approach has second highest rate in category of “incentive” (10.4%) and implementable suggestions include building objective and workable apprising system and reward for safe behavior. On the other hand, “crew pairing” should consider with items of experience, personality, language, or other critical factors; “job assignment” should consider with operators' circadian rhythms, especially arrange work break to reduce operator fatigue and maintain alertness in long-distance flight. Result of this study match with assertion of psychosocial perspective that aircrew pairing has to base on their level of experience, flight skills and personalities and develop crew resource management(CRM) training to improve interaction and communication between partners [9]. Additionally, the recommendations of " human/crew" approach would benefit teamwork and create harmonious collaboration in developed and Western countries which tend to prevail individualism.

Technology/Engineering Approaches. This approach focuses on change or improvement in tools, technology, and job aids to remediate human errors. This study found that 21 (11.5%) recommendation, which come from investigators and are classified into “technology/engineering” approaches, belonging to design/maintain. Implementable suggestions include improving warning or alarms to increase operators' awareness of abnormal conditions, developing new system t enter into “failsafe” mode problems occur, and scheduling survey new technologies or products in market. Moreover, the equipment which possess critical function should prove adequate spare parts or redundancy and SOP to prevent breakdown or interference during operation. According to the ergonomics and systems perspective, pilot performance is the result of a complex interaction among several factors [8]. Flying an aircraft is a very complicated and dynamic task, as well as pilots interact with high-tech airplanes and cockpit equipment and must safely operate their aircrafts in all type of weather conditions. Therefore, recommendations of "technology/Engineering" are improving man-machine interface, altimeter and airspeed indicator instrument, autopilot system and flight management system(FMS) to reduce the opportunity for pilot committing error and even these errors can be caught by airplane computer.

Task/ Mission Approaches. This approaches focus on way of changing operators' task to reduce errors and improve flight safety. This study found that recommendations of “Procedure” has the highest rate(15.2%) of this approaches and investigators' recommendations include using checklist or automatic facilities to reduce requirement for human memory, performing double-check with team member to avoid errors occurring in important steps, developing reward system to reinforce the behaviors of compliance with safe work practices. Regarding with Recommendations of “manual”, investigators suggest that redesigning procedure and checklist to be clearer or more user-friendly; rewriting procedure to delete ambiguous or inapplicable safety criteria. The assumptions of cognitive perspective that pilot's mental operation include things such as information recognition, problem diagnosis, goal setting, and strategy selection [5] and pilot errors are believed to occur when one or more of these mental operations fail to process information appropriately [6]. The recommendations of “task/mission” approaches can applied to discipline pilots processing information with efficient method, modify task to reduce aircrew's work-load and opportunity for error.

5 Conclusions

The Human Factors Analysis and Classification System (HFACS) is based on Reason’s model of human error, it described that active failures associated with the performance of front-line operators in complex systems, and latent failures which lie dormant within the system, combine with other local factors to breach a system’s defaces. Active failures of operators have a direct impact on safety. However, latent failures are spawned in the upper levels of the organization and are related to management and regulatory structures. On the other hand, the Human Factors Intervention Matrix (HFIX) is applied to develop intervention strategy of human errors with 5 difference approaches. This study found major recommendations were directed at organizational/ administrative and human/crew approaches respectively. Moreover, the most effective interventions focus on decision errors and violations. Human error is the principal threat to flight safety. In a worldwide survey of causal factors in commercial aviation accidents, in 88% of cases the crew was identified as a causal factor; in 76% of instances the crew were implicated as the primary causal factor. This research has demonstrated that the HFIX framework can be applied to improve human errors by way of five different approaches. It also has suggested that decision error and violations are critical issues of aviation safety and these can be improve by training and organizational administration for developing safety approaches.

References

1. International Civil Aviation Organization, Convention on International Civil Aviation: Annual Report of the Council-2011 (2012), http://www.icao.int/pubilcations/Docummnents/9975_en.pdf
2. Federal aviation administration. Aviation Maintenance Handbook-General-addendum/Human Factor, Washington, D.C. (2011)
3. Hollnagel, E.: Cognitive Reliability and Error Analysis Method – CREAM. Elsevier Science, Oxford (1998)
4. Wickens, C., Flach, J.: Information process. In: Wiener, E., Nagel, D. (eds.) Human Factors in Aviation, pp. 111–115. Academic Press, San Diego (1988)
5. Wiegmann, D.A., Shappell, S.A.: Human factors analysis of post-accident data: Applying theoretical taxonomies of human error. The international Journal of Aviation Psychology 7, 67–81 (1997)
6. Edwards, E.: Introductory overview. In: Wiener, E., Nagel, D. (eds.) Human Factors in Aviation, pp. 3–25. Academic Press, San Diego (1988)
7. Helmreich, R.L., Foushee, H.C.: Why crew resource management? Empirical and theoretical bases of human factors training in aviation. In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 3–45. Academic Press, San Diego (1993)
8. Heinrich, H.W., Petersen, D., Roose, N.: Industrial Accident Prevention: A Safety Management Approach, 5th edn. McGraw-Hill, New York (1980)
9. Shappell, S.A., Wiegmann, D.A.: The human factors analysis and classification system (HFACS). (Report Number DOT/FAA/AM-00/7). Federal Aviation Administration, Washington, DC (2000)
10. Reason, J.: Human Error. Cambridge University, New York (1990)
11. Wiegmann, D.A., Shappell, S.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Ashgate, Aldershot (2003)
12. Shappell, S.A., Wiegmann, D.A.: A methodology for assessing safety programs targeting human error in aviation. The International Journal of Aviation Psychology 7(1), 252–269 (2009)
13. Merritt, A.C.: Australian Aviation Psychology Industry Seminar. Sydney Australian Aviation Psychology Association, Sydney (1993)
14. Hofstede, G.: Dimensionalizing Cultures: The Hofstede Model in Context. Online Readings in Psychology and Culture, Unit 2 (2011), http://scholarworks.gvsu.edu/orpc/vol2/iss1/8
15. Hofstede, G., Hofstede, G.J., Minkov, M.: Cultures and Organizations: Software of the Mind, Rev. 3rd edn. McGraw-Hill, New York (2010)
16. Landis, J., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics (33), 159–174 (1977)
17. Diehl, A.: The effectiveness of training programs for preventing aircrew error. In: Sixth International Symposium on Aviation Psychology, pp. 640–655 (1991)
18. Jensen, R.S.: The boundaries of aviation psychology, human factors, aeronautical decision making, situation awareness, and crew resource management. The International Journal of Aviation Psychology 7(4), 259–268 (1997)
19. Department of Defense of United States: Department of Defense Human Factors Analysis and Classification System (DoD HFACS) A mishap investigation and data analysis tool (2005)

The Application of Human Factors Analysis and Classification System (HFACS) to Investigate Human Errors in Helicopter Accidents

Shao-Yu Liu[1,*], Chia-Fen Chi[1], and Wen-Chin Li[2]

[1] Dept. of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan, R.O.C.
[2] Graduate School of Psychology, National Defense University, Taipei, Taiwan, R.O.C.
davidliu735@gmail.com

Abstract. current study investigates 83 civil aviation and military services helicopter accidents in Taiwan between 1970 and 2010. The probable and latent causes of those accidents are clearly defined, and statistically analyzed by error related paths and Human Factors Analysis and Classification System (HFACS). Results indicate that categories of the higher level have better predicted power (between 4.25% and 24.9%) than categories of the lower levels (with odd ratios between 0.19 and 8.67). Fallible decisions in upper command levels directly affect supervisory practices which create pre-conditions for unsafe acts, impair performance of pilots, and lead to unexpected accidents. By identifying the higher level human errors leading to low level helicopter mishaps, HFACS is useful a tool for accident investigations and accident prevention strategies. Current study provides a practical suggestion to top managers for a better helicopter operational safety environment.

Keywords: Human Factors Analysis and Classification System (HFACS), Human Errors, Helicopter Flight Operations.

1 Introduction

Taiwan is a mountainous island surrounded by Pacific Ocean and Taiwan Strait. Natural disasters such as typhoons and earthquakes have constantly led to catastrophic damage in human livies and property. Helicopter, due to its maneuverability and operational flexibility, is very adapted for emergent rescue missions such as ambulance, observation on disastrous landscape, material transportation, and reconnaissance patrol over disastrous regions in Taiwan remote villages. The average flight hours are, therefore, increased from 1,000 hours to more than 10,000 hours per year since 1999 to 2008. (Aviation Safety Council, ASC, 2010). As a result, the average helicopter accident rate in the past ten years had dramatically increased to 10.24 accidents per 100,000 flight hours, which is 38.06 higher than the accident rate in the United States of American (USA). This astonishing accident record has casted

[*] Corresponding author.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 85–94, 2013.
© Springer-Verlag Berlin Heidelberg 2013

the public the image that the helicopter is one of the most unsafe transporter in Taiwan. The motivation of current study is to investigate the root causes of those accidents through the application of Human Factors Analysis and Classification System (HAFCS) and propose suggestions for top managers to improve organization aviation safety by reducing those potential hazards. Among those categories of root causes, human errors take part in 70% to 80% of civil and military aviation accidents. (O'Hare, Wiggins, Batt, & Morrison, 1994; Wiegmann and Shappell, 1999). The root causes of helicopter accidents in Taiwan are similar to scenarios worldwide that 57.14% were due to human factors, 28.50% were environmental factors, and 14.28% were contributed as maintenance factors. Human-error related accidents are still relatively high and stable over the last several years worldwide. (Shappell & Wiegmann, 1996). Therefore, Shappell and Wiegmann constructed HFACS as an analytical technique to look into the human error related incidents accidents qualitatively. The HAFCS is based on Reason's model of latent and active failures of an accident. In addition, FAA and NASA, USA have also applied HFACS, as a complement tool to pre-existing system with civil aviation in an attempt to capitalize on gains realized by the military (Ford, Jack, Crisp, & Sandusky, 1999). The HFACS framework bridges the gap between theory and practice by providing safety professionals with a theoretical-based tool for identifying and classifying the causes of human error aviation accidents. Because the framework focuses on both latent and active failures and their interrelationships, it facilitates the identification of the underlying causes of human error. HFACS has been proven to be an useful tool within the context of military aviation. This research examines the applicability of HFACS framework for the analysis of helicopter accidents investigation.

2 Literature Review

Many human factors accident analysis frameworks, taxonomies and analysis strategies have been devised over the years (e.g. Diehl, 1989; Feggetter, 1991). In recent years, accident investigation the scientific focus has shifted away from psychomotor skill deficiencies and emphasis is now more placed upon inadequacies in decision-making, attitude, supervisory factors and organizational culture as being the primary causal factors (Diehl, 1991; Jensen, 1997, and Klein, 2000). Dekker (2001) has proposed that human errors are systematically connected to features of operators' tools and tasks, and error has its roots in the surrounding system: the question of human or system failure alone demonstrates an oversimplified belief in the roots of failure. By examining and correlating information across a number of accidents, predictors may be identified which may then be applied to individual crews or situations in order to develop effective prevention strategies. Wiegmann & Shappell (2001) claim that the HFACS framework bridges the gap between theory and practice by providing safety professionals with a theoretically based tool for identifying and classifying the human errors in aviation mishaps. Since its It is based on a sequential or chain-of-events theory of accident causation and was derived from Reason's (1990), the classification system was originally developed for use within the

US military both to guide investigations and to analyse accident data (Shappell & Wiegmann, 2000b). Development has been used in a variety of transport and occupational settings including aviation, road and rail transport (Shappell & Wiegmann, 2000a; Federal Railroad Administration, 2005; Gaur, 2005; Li & Harris, 2005). It has also been used by the medical, oil and mining industries (Reinach and Viale, 2006). HFACS has also been used to analyse major flying operations (ex, commercial) and specific accident types, such as controlled flight into terrain (CFIT). Within the US aviation studies, the results have been consistent over time, with only small changes in the percentage of accidents associated with unsafe acts observed between earlier and later studies (Wiegmann & Shappell,2001;2005).The application of HFACS has also been effective for conducting comparisons between countries. Studies comparing US aviation accidents and those of other countries including China, Greece and India have been consistent. Their results indicated that while there were differences in the contributory factors between the countries, skill-based errors were associated with the greatest number of accidents in each of the countries followed by decision errors, violations and perceptual errors respectively (Gaur, 2005; Li & Harris, 2005).The system focuses on both latent and active failures and their inter-relationships, it facilitates the identification of the underlying causes of human error. However, as aviation accidents are the result of a number of causes, the challenge for accident investigators is how best to identify and mitigate the causal sequence of events leading up to an accident. whether this framework is suitable to meet needs of aviation accident's classification and investigation.

3 Method

Data: A total of 83 helicopter accidents and reported incidents, from 1970 to 2010 in Taiwan, is investigated. The aviation accident reports were obtained from Aviation Safety Council(ASC), Civil Aeronautical Administration(CAA), and Ministry of Defense(MOD) of Taiwan, R.O.C. There were same types of helicopter involved in the accidents, including commercial and Military aviation. All accidents and serious incidents conformed to the definition within the 9th edition of the Convention on International Civil Aviation, Annex 13 (International Civil Aviation Organisation, 2006).

Classification Framework: HFACS framework proposed by Wiegmann and Shappell (2003). HFACS Level-1: 'unsafe acts of operators' is the probable cause that directly lead to an accident. This Level-1 comprises four categories which are 'decision errors'; 'skill-based errors'; 'perceptual errors' and 'violations'. HFACS Level-2 is concerned with 'preconditions for unsafe acts'. This Level-2 has seven categories including : 'adverse mental states'; 'adverse physiological states'; 'physical /mental limitations';' crew resource management'; 'personal readiness';' physical environment' and 'technological environment'. HFACS Level-3 is concerned with 'unsafe supervision' which includes the four categories 'inadequate supervision'; 'planned inappropriate operation'; 'failure to correct known problem' and 'supervisory violation'. Level-4,the highest level in the framework is labelled 'organizational influences' and comprises of three sub-categories: 'resource management'; 'organizational climate' and 'organizational process'.

Coding Process: Each accident report was scrutinized and coded by two senior aviation investigators whose expertises are instructor pilot and aviation psychologist, respectively. Qualified investigators should possess at least 12-hour HFACS training. The presence (code 1) or the absence (code 0) of each HFACS category was carefully assessed in each accident report, narrative. To avoid over-representation from any single accident, each HFACS category was counted a maximum of only once per accident. The count acted simply as an indicator of presence or absence of each of the 18 categories in a given accident..

4 Analysis

In total instances of 626 category assignments were made to described the causal factors underlying the 83 accidents. The inter-rater reliabilities calculated on a category-by-category basis were assessed using Cohen's Kappa. The values obtained ranged between 0.62 and 1.0 (see table 1). Fourteen HFACS categories exceeded a Kappa of 0.60 indicating substantial agreement (Landis & Koch, 1977). As Cohen's Kappa can produce misleadingly low figures for inter-rater reliability where the sample size is small or where there is very high agreement between raters associated with a large proportion of cases falling into one category (Huddlestone, 2003), inter-rater reliabilities were also calculated as a simple percentage rate of agreement. These showed reliability figures between 89.2% to 100%, further indicating acceptable reliability between the raters. See Li & Harris (Li & Harris, 2006) for further details. Relatively few categories had exceptionally low counts. Only the categories of 'Failed to correct known problem';'Personal readiness' and 'adverse physiological state' failed to achieve double figures. The results reported only to the instances where the PRE was in excess of 5%. The data were cross tabulated to describe the association between the categories at adjacent levels in the HFACS analytical framework. Goodman and Kruskall's lambda (λ) was used to calculate the proportional reduction in error (PRE) (Goodman, 1954). The Lambda statistic is analogous to the R squared statistic for continuous data. For categorical data (such as that found in contingency tables), its value reflects the PRE when predicting the outcome category from simply the baseline prevalence as compared to using information from the predictive category. For the purposes of this study the lower level categories in the HFACS were designated as being dependent upon the categories at the immediately higher level in the framework, which is congruent with the theoretical assumptions underlying HFACS: from this standpoint, lower levels in the HFACS cannot affect higher levels. Finally, odds ratios were also calculated which provide an estimate of the likelihood of the presence of a contributory factor in one HFACS category being associated concomitantly with the presence of a factor in another category. However, it must be noted that odds ratios is an asymmetric measure and so are only theoretically meaningful when associated with a non-zero value for lambda.

Table 1. The frequency and percentage of 83 accidents by HFACS categories

HFACS category		Frequency	Ordinal	Percentage	Inter-rater reliability	Cohen's Kappa
Level 1	Decision errors	68	2	81.9%	92.8%	0.788
	Skill-based Errors	57	4	68.7%	90.4%	0.792
	Perceptual Errors	24	11	28.9%	98.8%	0.970
	Violations	17	14	20.5%	98.8%	0.962
Level 2	Adverse mental State	55	5	66.3%	100.0%	1.000
	Adverse physical state	8	16	9.6%	98.8%	0.927
	Physical/Mental limitation	47	7	56.6%	98.8%	0.976
	Crew resource management	73	1	88.0%	89.2%	0.624
	Personal readiness	2	18	2.4%	100.0%	1.000
	Physical environment	67	3	80.7%	100.0%	1.000
	Technical environment	20	13	24.1%	100.0%	1.000
Level 3	Inadequate supervision	49	6	59.0%	91.6%	0.830
	Planned inappropriate Operation	25	10	30.1%	96.4%	0.911
	Failed to correct known problem	5	17	6.0%	100.0%	1.000
	Supervisory violation	10	15	12.0%	100.0%	1.000
Level 4	Resource management	32	9	38.6%	96.4%	0.922
	Organizational climate	24	11	28.9%	95.2%	0.877
	Organizational processes	43	8	51.8%	94.0%	0.880

5 Result

All these relationships were also associated with high odds ratios, suggesting that inadequate performance in the higher level HFACS categories was associated with much increased levels of poor performance at the lower levels. The strength analysis on HAFCS level-4 'organizational influences' associated with adjacent HFACS level-3 'unsafe supervision' indicates that, among possible 12 pairs of relationships, 3 associations are significant ($p<0.05$), From statistical analysis, 'Inadequate supervision' is 5.28 times more likely to occur when organizational level associates with poor 'Resource management'; 'Supervisory violation' is 0.83 times more likely to occur when organizational level associates with poor organizational climate; Similarly, 'inadequate supervision' is 2.55 times more likely to occur in the presence of poor organizational processes. The strength analysis on HAFCS level-3 'unsafe supervision' associated with adjacent HFACS level-2 'pre-conditions for unsafe acts' indicates that, among possible 28 relationships, 6 associations are significant. ($p<0.05$).'CRM' is 0.91 times more likely to occur when supervision level associates with poor 'inadequate supervision' and 'Adverse mental state' is 0.07 times more likely to occur when supervision level associates with poor 'supervisory violation'; An issue also is associated with 'Adverse physical state' is 4.58 times more likely to occur when supervision level associates with poor 'planned inappropriate operations' and is 8 times more likely to occur when supervision level associates with poor 'failed to correct a known problem'; 'Personal readiness' is 1.25 times more likely to occur when supervision level associates with poor 'Supervisory violation', and 'CRM' is 0.71 times more

Table 2. Significant association between upper level and adjacent downward level categories in the HFACS framework

	χ^2 test		τ (PRE)		Odds rati
	Value	p-level	Value	p-level	
Organizational process vs Inadequate supervision	4.25	.004	.005	.040	5.27
Organizational climate vs supervisory violation	4.63	.005	.056	.033	0.83
Resource management vs Inadequate supervision	10.62	.001	.128	.001	2.55
Inadequate supervision vs Crew resource management	9.33	.002	.112	.002	0.19
Inadequate supervision vs Adverse mental state	16.38	.000	.197	.000	0.70
Planned inappropriate operations Vs adverse Physical state	4.41	.050	.053	.037	4.58
failed to correct a known problem vs adverse Physical state	5.63	.071	.068	.018	8.00
Supervisory violation vs Crew resource management	15.45	.002	.186	.000	0.07
Supervisory violation vs Personal readiness	14.96	.013	.180	.000	1.25
Adverse mental state vs Perceptual errors	9.74	.002	.117	.002	8.66
Physical/mental limitation vs Decision errors	6.35	.001	.077	.050	0.27
Crew resource m nagement vs Violations	24.92	.300	.039	.000	4.56
Crew resource management vs Decision errors	10.90	.004	.131	.001	0.11
Personal readiness vs Decision errors	4.49	.096	.054	.035	0.29
Physical environment vs Decision errors	5.79	.017	.070	.017	0.11
Technological environment vs Perceptual errors	5.57	.025	.067	.019	5.53
Technological environment vs Decision errors	4.58	.046	.055	.033	0.20

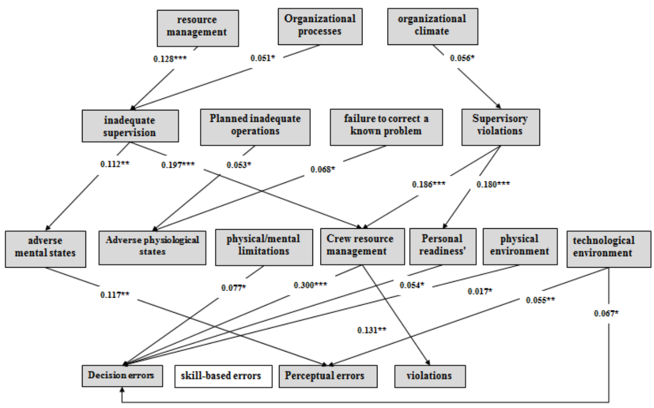


Fig. 1. The significant association of Chi-square (χ2) and Tau (τ)

likely to occur when supervision level associates with poor 'Supervisory violation'. The strength analysis on HAFCS level-2 'pre-conditions for unsafe acts' associated with adjacent HFACS level-1 'unsafe acts of operators' indicates that a possible 28 relationships, 8 pairs of associations are significant ($p<0.05$)

'Decision errors' is 0.27 times more likely to occur when pre-condition level associates with poor 'Physical/mental limitation'; Similarly, 'Decision errors' is 0.12

times more likely to occur in the presence of poor ‘CRM’; ‘Decision errors’ is 0.29 times associated with poor ‘Personal readiness’ ; ‘Decision errors’ is 0.11 times associated with poor Physical environment, and ‘Decision errors’ is 0.21 times associated with poor technological environment. Perceptual errors are over 8.67 times more likely to occur when there are pre-conditions level issues associated with poor ‘adverse mental state’ and over 5.54 times associated with poor ‘Technological environment’. Similarly, Violations is 4.56 times more likely to occur in the presence of poor ‘CRM’.

6 Discussion

In application of HFACS framework, inadequacies in the following categories of Level-4: ‘Organizational processes’ particularly result from excessive time pressures, poor mission scheduling, poor risk management programs, inadequate management checks for safety, failing to establish safety programs and ‘Resource management’ which involves the staff selection, training of human resources at an organizational level, excessive cost cutting, unsuitable equipment, and failure to remedy-design flaws also shows strong correlation with the level-3 categories of ‘inadequate supervision’. Untrained supervisors and general loss of situation awareness at the supervisory level(Li et al,2006a,b;2008). ‘Organizational climate’ including inadequacies in chain of command, poor delegation of authority, inappropriate organizational customs and beliefs. however, it is strongly correlated with the 'Supervisory violations'. hypothesize that inappropriate decision-making by upper-level management can adversely influence the personnel and practices at the supervisory level (Reason, 1990; Wiegmann & Shappell, 2003). It is strongly correlation with two categories of ‘CRM’, such as Poor ‘CRM’ including ‘Lack of teamwork’ and ‘Adverse mental states' such as ‘Failure to provide proper training’ or ’Adequate rest periods’(Li and Harris, 2006a,b). In addition, Adverse mental state includes ‘Lack of mental fatigue’ and ‘stress’. Inadequacies in ‘adverse mental practices’ are particularly influenced by the level-3 category of ‘inadequate supervision’. This category encompasses issues such as ‘failure to provide adequate rest periods’ and performance of personnel’ (Reason, 1990). These erroneous actions need not be confined to either 'Inadequate supervision’ or ‘Adverse mental’. Poor 'Supervisory violations' includes ‘authorizing an unqualified crew for flight’ and ‘Supervisors violating procedures’, it also shows strong relationships with the level-2 two categories of ‘CRM’ and ‘Personal readiness’. The Poor CRM includes lack of teamwork and ‘failures of leadership’. This category encompasses issues such as ‘failure to provide authorizing an unqualified crew for flight’. Moreover, poor ‘Personal readiness’ includes lack of selfmedication’and overexertion while off duty. Inadequacies in ‘Personal readiness’ practices are particularly influenced by the category of ‘Supervisory violations'. This category encompasses issues such as failure to provide ‘supervisors violating’ inadequate documentation and willful disregard of authority by the supervisor (Li and Harris’, 2008). The ‘preconditions for unsafe acts’ category poor ‘Adverse physiological states’ encompasses issues associated with inadequate training, self-medication and overexertion while off duty. Inadequacies in 'Adverse physiological states' practices are particularly influenced by the level-3 category of' 'Planned inadequate operations' and 'Failures to correct inappropriate behavior'. The Poor 'Planned inadequate operations' includes poor crew pairings and excessive task/workload, 'Failures to correct inappropriate behavior', failing to remove a known safety hazard, failing to report unsafe tendencies. This category in HFACS framework encompasses issues such as failure to provide poor crew pairings, failure to establish if risk outweighed benefit. Current study clearly provides evidence that inadequacies at HFACS level-2 'preconditions for unsafe acts' has associations with further inadequacies at HFACS level-1 'unsafe acts of operators'(Table2). The most frequently occurring category is 'Decision errors' which is also a particularly important factor at this 'unsafe acts of operators'. 'Decision errors' encompasses issues associated with failure of selecting inappropriate strategies during mission. The next most frequent category are lack of teamwork, poor communication, failure of leadership and inadequate briefing. The 'Technological environment' category covers issues such as equipment design, cockpit display interfaces, automation and checklist layout (Li el al, 2006a；2008).The level-1'unsafe acts of operators' category of 'Perceptual errors'. Poor 'Perceptual errors' included encompassed issues associated with experiencing spatial disorientation and descent rate during IMC. 'Perceptual errors' practices were particularly influenced by the level-2 category of' 'adverse mental states' and 'Technological environment'. This category in the HFACS encompasses issues such as a failure to provide, 'adverse mental states' included issues such as over-confidence, stress, distraction, and task saturation (Li and Harris 2008). Another 'Technological environment' by the effects of the lower.This accident involves 'unsafe acts of operators' category of 'CRM' which includes Poor 'CRM' such as lack of teamwork, poor communication and inadequate briefing. 'CRM' practices are particularly influenced by the level-2 category of' 'violations'. 'Violations' in HFACS framework encompasses issues such as pilots fail to provide or follow standard operation procedures (SOPs) (Li et al, 2008).

7 Conclusion

The Human Factors Analysis and Classification System (HFACS) was developed as an analytical framework for the investigation of the role of human factors in aviation accidents, becoming one of the most commonly used and is the one used herein as a basis for the current work. Strategies in application of HFACS on accident investigations have been successfully verified by many aviation psychological scholars. (Diehl, 1989; Wiegmann & Shappell, 2003). The benefit of HFACS is that the contributing human errors in any single accident can be properly categorized regardless the aircraft type of helicopter accident, and provide a preventive strategy for safety assurance. In most cases, space Disorientation and CFIT are two major probably causes for helicopter accidents. Consequently, strategy for helicopter safety promotion suggested from current study to the top management of helicopter organization is as follows. First, supervision on flight plan and pre-condition briefing requires reinforced compliment, particular on violators. Second, safety equipment promotion such as all weather radar and terrain detection radar increases the quality on decision-making of crewmembers.

References

1. Aviation Safety Council Report no. Taiwan's aviation safety statistics 2000-2010, `http://www.asc.gov.tw/author_files/statistics02-11.pdf`
2. Diehl, A.: Human performance/system safety issues in aircraft accident investigation and prevention. In: Proceedings of 11th International Symposium on Aviation Psychology, Columbus, OH (1989)
3. Ford, C., Jack, T., Crisp, V., Sandusky, R.: Aviation accident causal analysis. In: Advances in Aviation Safety Conference Proceedings, p. 343. Society of Automotive Engineers Inc., Warrendale (1999)
4. Gaur, D.: Human factors analysis and classification system applied to civil aircraft accidents in India. Aviat. Space Env. Med. 76, 501–505 (2005)
5. Goodman, L., Kruskal, W.H.: Measures of association for crossclassifications. J. Am. Stat. Assoc. 49, 732–764 (1954)
6. Reason, J.T.: Human Error. Cambridge University Press, Cambridge (1990)
7. Reason, J.T.: Managing the Risks of Organizational Accidents. Ashgate, UK (1997)
8. Reinach, S., Viale, A.: Application of a human error framework to conduct train accident/incident investigations. Accident Analysis and Prevention 38, 396–406 (2006)
9. Hunter, D.R., Baker, R.M.: Reducing Accidents among General Aviation Pilots through a National Aviation Safety Program. In: The Fourth Australian Aviation Psychology Symposium Aldershot, Ashgate, England (2000)
10. Huddlestone, J.A.: An evaluation of the training effectiveness of a lowfidelity, multi-player simulator for Air Combat Training. Unpublished Ph.D.Thesis. College of Aeronautics, Cranfield University, England (2003)
11. International Civil Aviation Organisation. Convention on International Civil Aviation ICAO. Document 7300/9, ninth ed. ICAO, Monteal, Canada (2006)
12. Landis, J.R.: Koch GGThe Measurement of Observer Agreement for Categorical Data. Biometrics 33, 159–174 (1977)
13. Li, W.C., Harris, D.: HFACS analysis of ROC air force aviation accidents: reliability analysis and cross-cultural comparison. Int. J. Appl. Aviat. Stud. 5, 65–81 (2005a)
14. Li, W.C., Harris, D.: Where safety culture meets national culture: the how and why of the China Airlines CI-611 accident. Hum. Factors Aerospace Saf. 5, 345–353 (2005b)
15. Li, W.C., Harris, D.: Pilot error and its relationship with higher organizational levels: HFACS analysis of 523 accidents. Aviat. Space Env. Med. 77, 1056–1061 (2006a)
16. Li, W.C., Harris, D.: Breaking the chain: an empirical analysis of accident causal factors by human factors analysis and classification system (HFACS). In: Proceedings of International Society of Air Safety Investigators Seminar Cancun, Mexico, September 11-14 (2006b)
17. Li, W.-C., Harris, D., et al.: Routes to failure: Analysis of 41 civil aviation accidents from the Republic of China using the human factors analysis and classification system. Accident Analysis and Prevention 40(2), 426–434 (2008)
18. O'Hare, D., Wiggins, M., Batt, R., Morrison, D.: Cognitive failure analysis for aircraft accident investigation. Ergonomics 37, 1855–1869 (1994)

19. O'Hare, D.: The 'Wheel of Misfortune': A taxonomic approach to human factors in accident investigation and analysis in aviation and other complex systems. Ergonomics 43, 2001–2019 (2000)
20. Reinach, S., Viale, A.: Application of a human error framework to conduct train accident/incident investigations. Aviat. Space Env. Med. 30, 396–406 (2006)
21. Shappell, S., Wiegmann, D.: U. S. Naval Aviation mishaps 1977-92: Differences between single- and dual-piloted aircraft. Aviation, Space, and Environmental Medicine 67, 65–69 (1996)
22. Shappell, S., Wiegmann, D.: The Human Factors Analysis and Classification System (HFACS). (Report Number DOT/FAA/AM-00/7). Federal Aviation Administration, Washington, DC (2000a)
23. Shappell, S., Wiegmann, D.: Is proficiency eroding among U.S. Naval aircrews? A quantitative analysis using the Human Factors Analysis and Classification System (HFACS). In: Proceedings of the 44th Meeting of the Human Factors and Ergonomics Society (2000b)
24. Shappell, S.A., Wiegmann, D.A.: Applying Reason: The Human Factors Analysis and Classification System (HFACS). Human Factors and Aerospace Safety 1(1), 59–86 (2001)
25. Shappell, S.A., Wiegmann, D.A.: A human error analysis of general aviation controlled flight into terrain accidents occurring between 1990 and 1998. Report no. DOT/FAA/AM-03/4. Federal Aviation Administration, Washington, DC (2003)
26. Shappell, S.A., Wiegmann, D.A.: HFACS analysis of military and civilian aviation accidents: a North American comparison. In: Proceedings of International Society of Air Safety Investigators, Australia, Queensland, November 2-8 (2004)
27. Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., Wiegmann, D.: Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System. Human Factors 49, 227–242 (2007)
28. Thomas, M.J.W.: Improving organisational safety through the integrated evaluation of operational and training performance: an adaptation of the Line Operations Safety Audit (LOSA) methodology. Hum. Factors Aerospace Saf. 3, 25–45 (2003)
29. Wiegmann, D.A., Shappell, S.A.: Human error and crew resource management failures in Naval aviation mishaps: A review of U.S. Naval Safety Center data, 1990-96. Aviation, Space, and Environmental Medicine 70, 1147–1151 (1999)
30. Wiegmann, D.A., Shappell, S.A.: Human error analysis of commercial aviation accidents: application of the human factors analysis and classification system. Aviat. Space Env. Med. 72, 1006–1016 (2001a)
31. Wiegmann, D.A., Shappell, S.A.: A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System. Ashgate, UK (2003)
32. Wiegmann, D.A., Esa, M.: Rantanen. Defining the Relationship Between Human Error Classes & Technology Intervention Strategies (AHFD-03-15, NASA-02-1) (2003)
33. Wiegmann, D.A., Shappell, S.A.: (Report No. DOT/FAA/AM-05/24). Federal Aviation Administration, Washington, DC (2005)

A Fixed-Based Flight Simulator Study: The Interdependence of Flight Control Performance and Gaze Efficiency

Lewis L. Chuang[1], Frank M. Nieuwenhuizen[1], and Heinrich H. Bülthoff[1,2,⋆]

[1] Department of Perception, Cognition and Action,
Max Planck Institute for Biological Cybernetics, Tübingen
[2] Department of Cognitive and Brain Engineering, Korea University
{lewis.chuang,frank.nieuwenhuizen,
heinrich.buelthoff}@tuebingen.mpg.de

Abstract. Here, a descriptive study is reported that addresses the relationship between flight control performance and instrument scanning behavior. This work was performed in a fixed-based flight simulator. It targets the ability of untrained novices to pilot a lightweight rotorcraft in a flight scenario that consisted of fundamental mission task elements such as speed and altitude changes. The results indicate that better control performance occurs when gaze is more selective for and focused on key instruments. Ideal instrument scanning behavior is proposed and its relevance for training instructions and visual instrument design is discussed.

1 Introduction

Like driving a car, piloting a rotorcraft is a closed-loop motor control task that relies heavily on visual feedback. Visual feedback informs us on the outcome of our control inputs and whether it is necessary to change these inputs in order to arrive at our desired performance levels. Information such as velocity is available from the outside world, it can also be provided more precisely via instruments (eg., airspeed indicator).

Piloting a rotorcraft is, arguably, more difficult than driving a car for several reasons. First, it offers movement in more degrees of freedom. A rotorcraft pilot is required to control not only the vehicle's airspeed and heading, but also its altitude. Second, the control devices of a rotorcraft are not directly mapped to the direction of the vehicle's motion. In a car, accelerations and decelerations can be achieved by exerting pressure on the accelerator and braking pedal, respectively. Maintaining a constant velocity is achieved by determining the appropriate pressure to consistently apply to the accelerator. This connection between control

⋆ The work in this paper was supported by the myCopter project, funded by the European Commission under the 7th Framework Program.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 95–104, 2013.
© Springer-Verlag Berlin Heidelberg 2013

device and vehicle control is direct and intuitive. In contrast, rotorcraft acceleration is achieved by tilting the rotor disc (and thus, the helicopter) forward. Lateral decelerations are effected by tilting the rotor disc in the opposite direction. Thus, flying at a constant airspeed requires continuous adjustments of the rotor disc's tilt. Last but not least, rotorcraft controls are coupled. Tilting the rotor disc forward will not only accelerate the rotorcraft in the same direction but will also reduce its upward thrust, hence causing a loss in altitude. Conversely, decelerating the rotorcraft by reducing the forward tilt of the rotor disc will increase the upward thrust, causing an increase in altitude.

Efficient monitoring of available visual feedback information can be expected to be relevant to effective rotorcraft control. By moving his eyes, the pilot is able to switch between monitoring several visual feedback cues (eg., airspeed, altitude, heading, etc.) and checking for any discrepancies between desired and actual values. This underlies his ability to manipulate the rotorcraft's controls appropriately, such as to reduce detected discrepancies. Scanning strategies have also been shown to differ across levels of control expertise. Recently, navigation accuracy was found to negatively correlate with the amount of time that pilots looked at the "outside" world [1]. Thus, expertise might be reflected in a decreasing reliance on outside world cues.

Gazetracking methods allow researchers to determine how visual feedback information is sought after during flight control. In the context of driving and aviation, gazetracking data can identify which visual cues and instruments are relied on the most. For example, the primary flight display was the cockpit instrument that pilots looked at the most during the approach and landing of a A330 fixed wing aircraft — namely, 40% of looking time [2]. Within this, attention was particularly dedicated to information relating to the attitude of the plane, its airspeed and altitude. Findings such as these can improve cockpit designs, by tailoring the visualization of information to fit the pilot's mission objectives.

To date, relevant eye-tracking research has predominantly addressed expert flight performance. Relatively little is known about the relationship between flight control performance and instrument gaze behavior in novices. Our research seeks to identify methods and technologies that will enable personal aviation transport, by allowing flying a personal aerial vehicle to be as easy as driving a car [3][4]. To do so, it is necessary to measure and model the behavior of untrained individuals that underlie their ability to control an aircraft.

In this study, we designed a flight scenario that reflected a daily commute from a suburban area to a city region. Novices were required to pilot a lightweight rotorcraft model in a helicopter flight simulator that was equipped with a remote eye-tracking system. The motivation of this work was to explore the relationship between flight control performance and scanning strategies of novices. Specifically, we investigated how novices with varying levels of control performance might differ in terms of how they looked at their flight instruments for control information. It should be mentioned that flight trainees are explicitly taught strategies on how to scan their flight instruments for flight control.

In this paper, we investigate how novices are likely to monitor flight instruments without having received such instructions.

The remainder of this paper is organized as follows. Section 2 provides a detailed description of the flight mission and the fixed-base simulator environment. It also describes how data was collected for subsequent analyses. Section 3 presents the descriptive statistics of our current work and our interpretations. Section 4 discusses the implications of the current results and how that could influence training instructions and the interface design of visual instrumentation.

2 Methods

2.1 Participants

Four male participants were recruited for this study (age range: 29–34 years). They possessed normal or corrected-to-normal vision. None of the participants had formal flight training for rotorcraft. More importantly, they have never received any formal instructions for instrument scanning strategies.

2.2 Apparatus and Flight Model

The fixed-based flight simulator was based on a simulated model of a light-weight helicopter with a large bandwidth for control inputs [5] (Figure 1). Visualizations of the outside world were based on the topography of San Francisco, USA. These were presented via a large multi-panelled display that rendered a world environment (field-of-view: 105° x 100°).

A heads-down display consisted of eight standard flight instruments that indicated the rotor speed (RS), airspeed (ASI), attitude (AI; also referred to as the artificial horizon), altitude (AT), torque (TQ), compass (CO), heading (HI), vertical speed (VSI) — see Figure 1 (inset). These are arranged in the standard T-arrangement of ASI, AI, AT on the second to four position in the top row, and HI in the third position of the bottom row. Eye-movements on the instrument panel were recorded using a 60 Hz remote eyetracker (faceLAB; SeeingMachines). The users gaze vector was estimated up to an accuracy of 1.5°, which was sufficient for determining the specific flight instrument that was fixated by the user.

A cyclic stick, collective lever and foot-pedals comprised the control devices. The cyclic stick controls the tilt of the rotorcraft. Namely, the rotorcraft will tilt in the same direction as the cyclic stick. If the cyclic is moved forward, the rotorcraft tilts forward; if the cyclic is moved aft, the rotorcraft tilts aft, and so on. These changes are reflected in the attitude indicator (AI), which indicates the rotorcraft's relation to the horizon — that is, whether it is pointing below or above and whether it is level with the horizon. Tilting the rotorcraft such that it points below the horizon will increase airspeed (ASI). This forward tilt will induce a loss in altitude (AT) unless accompanied by an increase in thrust power (RS). Conversely, reducing airspeed (ASI) by decreasing the rotorcraft's forward

Fig. 1. Flight simulator with remote eye-tracking cameras (indicated in the white ellipse). Instrument panel is featured inset (From left to right. Top panel: RS, ASI, AI, AT. Bottom panel: TQ, CO, HI, VSI.)

tilt will induce an elevation in altitude if the thrust (RS) is not simultaneously reduced. Control of thrust is determined by the collective lever. The yaw of the rotorcraft can be controlled by the foot-pedals. This compensates for any undesired torque forces (TQ) on the fuselage that result from changes in the rotor speed (RS). The global bearing of the rotorcraft is indicated by the compass (CO) and its current heading relative to its destination by a heading indicator (HI).

2.3 Flight Scenario

Prior to experimentation, participants had at least 10 hours of experience in the flight simulator, which familiarized them with the dynamics of the simulated vehicle and the instrument layout. The participants were aware of how the controls were coupled and how to use them in combination to achieve simple maneuvers — such as to accelerate, decelerate, ascend and descend — without creating large instabilities.

Four experimental sessions were conducted per participant with a common flight mission. Our participants were to fly from a suburban airfield to the city area of San Francisco, USA. This reflects a daily personal aerial vehicle commuting scenario. Three classes of mission task elements (MTEs) comprised the entire flight scenario: straight and level flight, altitude change, and airspeed change.

This flight scenario was designed, with the assistance of an experienced helicopter pilot (approx. 650 flight hours), to be within the control capabilities of a beginning trainee.

2.4 Data Collection

The airspeed and altitude of the rotorcraft were recorded during each flight session, in tandem to our participants' gaze. This allowed us to characterize flight control performance (see Subsection 3.1)

Nine regions of interest (ROI) were designated. Namely, the eight instruments and the outside world. Dwells were defined as periods when the participants' gaze intersected with one of these ROIs. Gaze data across the flight mission were first classified into those that belonged to the heads-down instrument panel or the world environment. Dwells belonging to the instrument panel were further classified for the specific instruments using a k-nearest-neighbor filter. Thus, it was possible to derive the number of dwells per ROI as well as the likelihood that a ROI dwell follows from another ROI dwell (see Subsection 3.2).

3 Results and Discussion

3.1 Flight Control

The current flight scenario required our participants to demonstrate control over their airspeed and altitude. Figure 2 (upper half) illustrates the desired trajectory for these two attributes and how our participants deviated from them across their final flight mission. Performance in keeping altitude and airspeed was determined by calculating the root-mean-squared error (RMSE) between desired and actual value. For MTEs in which a change in altitude or airspeed was performed by the participants, the RMSE for altitude error or velocity error, respectively, was not determined.

All participants showed increased instabilities in keeping altitude with respect to straight-and-level flight after they performed changes in airspeed (acceleration and deceleration). Conversely, changes in altitude did not result in more airspeed instability with respect to straight-and-level flight.

Across the whole flight mission, our participants differed primarily on their ability to control airspeed, rather than altitude (see Fig. 2, lower half). Across all participants, a prominent increase in altitude control difficulty is observed in the later stages of the flight mission, especially when decelerating and descending.

We used the RMSE metric to rank-order the performance of our participants, such that we could evaluate their gaze behaviour for corresponding trends. Participant S4 performed worst in keeping airspeed and altitude, whereas participants S1, S2 and S3 had similar performance in keeping altitude, but showed decreasing performance in keeping airspeed, respectively. Henceforth, they will be referred to as S1, S2, S3 and S4 in increasing order of RMSE, and color-coded in the figures as dark-green, green, red and dark-red respectively.

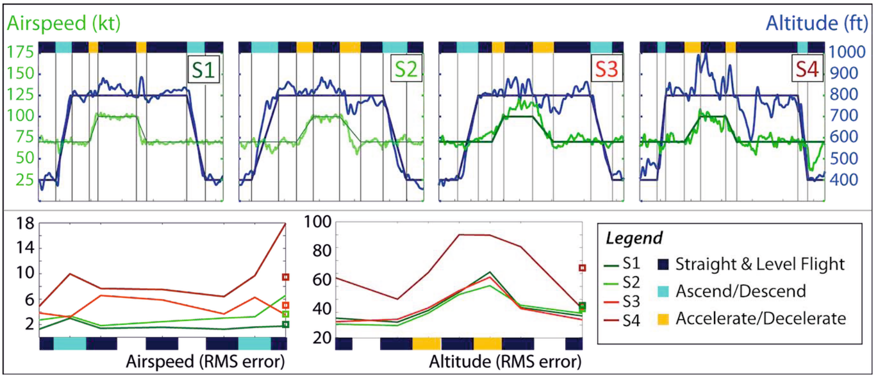


Fig. 2. Upper half: Control performance during a standard flight scenario. Participants are ranked according to decreasing control performance, from left to right. Nine mission task elements comprise the flight scenario: straight and level flight (dark blue), altitude change (light blue), airspeed change (yellow). Desired values for airspeed (green) and altitude (blue) across the flight scenario are indicated by the underlying straight lines and actual performance is overlaid. **Lower half:** Mean RMSEs of flight control performance of each participant. The mean RMSEs across flight mission are indicated on the right vertical axis.

3.2 Gaze Behavior

Gaze Selectivity across Instruments. Figure 3 illustrates the mean number of dwells on each instrument and the outside world (rows) for each MTE (columns) across the flight mission. Each grayscale histogram reflects the number of times each instrument was looked at; white denotes the maximum number of dwells within each individual histogram and black indicates that the instrument was never looked at.

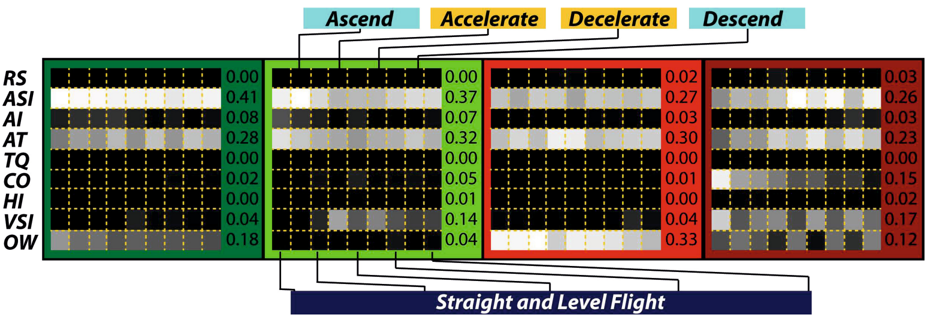


Fig. 3. Grayscale histogram for the mean number of dwells on each instrument and outside world (rows), for each maneuver (columns). The mean proportions of instrument dwells are reported on each histogram's right.

Every participant consistently demonstrated a gaze preference for the airspeed indicator (ASI) and altimeter (AT). This corresponds with the main determinants for flight control performance in our flight scenario. This instrument preference did not vary according to the changing mission task elements.

To reiterate, our participants primarily differed in terms of their ability to control airspeed, rather than altitude. A corresponding trend is noted in the gaze allocation to instruments. Mean proportion of dwells on ASI decreased from S1 to S4. Mean proportion of dwells on AT did not vary much between S1, S2 and S3 but was noticeably less for S4.

More than 85% of S1's dwells can be attributed to ASI, AT and the outside world (OW), in decreasing order of preference. This can be contrasted with S4's gaze, whereby the same three instruments only accounted for 61% of dwells. In comparison, S4 was less selective than S1 and looked at the compass (CO) and vertical speed indicator (VSI) more often than S1, at the expense of his monitoring levels of ASI and AT.

An interesting contrast can be observed between S2 and S3. S2 does not look at the OW as often as S3. In fact, S3 looks at the OW most frequently, relative to the other participants. Instead, S2 demonstrates a gaze preference for the VSI, especially during the middle of the flight mission from acceleration to deceleration, when control performance is most vulnerable across all participants (see Fig. 2).

Looking at Instruments versus Outside World. Control performance in novice pilots is unlikely to be characterized by the relative distribution between looking at the instruments and the OW. About 90% of the dwells on instruments prior to looking at the OW could be accounted for by 8 or fewer dwells in S1's flight missions. And in S2, S3 and S4s' flight missions by 22, 3 and 15 dwells or fewer.

Instrument Scanning. Given that a particular instrument is being looked at, what are the other instruments that will be looked at next? Figure 4 illustrates these gaze relationships between instruments. Each column indexes an instrument and reflects the number of subsequent dwells on other instruments (rows).

The obvious relationship is that between ASI and AT. For S1, he is equally likely to look at AT and OW after looking at ASI when ascending and accelerating. When his control performance decreases during decelerating and descending, he looks at the OW less frequently and switches gaze primarily between AT and OW. Notably, S1 treats ASI as the primary reference point in instrument scanning. That is, S1's gaze consistently returns to this instrument after referencing AT and OW. Dwells in AT and OW tend to occur after ASI dwells (see column 2 for S1) and, reciprocally, dwells in AT and OW tend to be followed predominantly by ASI dwells (see columns 4 and 9 for S1).

A similar pattern is observed for S2. Unlike S1, however, S2 either switches gaze between ASI and AT almost exclusively (ascend, descend) or between ASI, AT and VSI (accelerate, decelerate). He does not exhibit a primary instrument that he returns to preferentially after referencing other instruments, like ASI in S1's case.

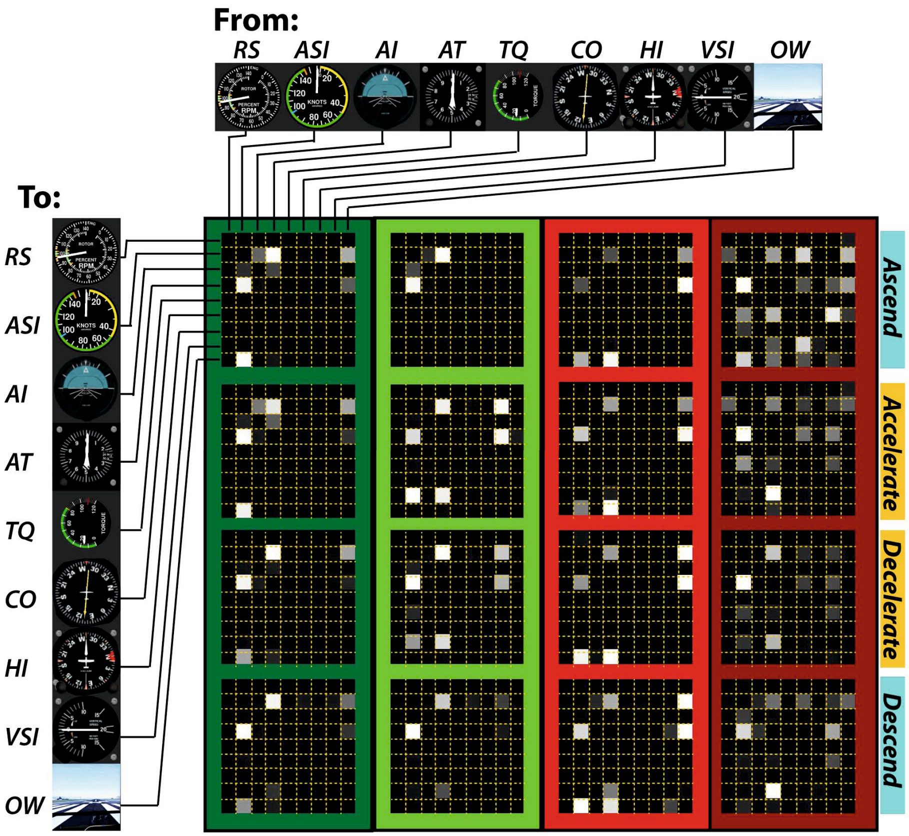


Fig. 4. Transition matrix of instrument dwells in grayscale, for the key maneuvers. Each column is indexed by an instrument. White denotes the maximum number of subsequent dwells within each matrix.

S3's gaze alternates between ASI, AT and OW. Unlike S1, S3 relies on the OW instead of ASI as the primary reference point. In contrast to both S1 and S2, S3 demonstrates less turn-taking between ASI and AT. It is also interesting to note that this reliance on OW as a primary reference point increases when control performance deteriorates (ie., decelerate, descend).

There is no clear transitional pattern of gaze turn-taking across the instruments for S4. The lack of a consistent gaze strategy could underlie S4's poor control performance.

4 Conclusions

The current paper presents a descriptive study that was motivated to demonstrate and understand the relationship between flight control performance and gaze behavior. Our findings show that gaze behavior varies with control performance. In particular, participants with better control performance exhibited gaze behavior that was more selective for key instruments. Our best performer demonstrated a selective and ranked preference for different visual information, whilst our second and third performer distributed gaze equally across three different sources of visual information. The worst performer distributed his gaze across five sources of visual information.

The current findings do not allow us to conclude that selective gaze patterns generate good flight control performance (or vice versa). Nonetheless, we can recommend training instructions for novice pilots with regards to instrument scanning behavior. This can be expected to yield improvements given that gaze strategies are easier to be voluntarily controlled by novice pilots than flight control expertise. Fundamentally, instrument scanning should always be centered on the instrument that is key to control performance. This means that eye-movements should ideally return to one key instrument after referencing secondary sources of information. Distributing gaze evenly and across many instruments without such an underlying strategy is ill-advised. Indeed, such instructions are provided in formal flight training; the T-arrangement is intended to encourage such instrument scanning behavior.

The current work was intended to understand how novices without formal flight instructions would scan instruments. It is interesting to note that they generally prioritized the airspeed indicator, followed by the altitude indicator. This suggests that novice pilots intuitively react to changes in their airspeed and altitude and compensate accordingly. In formal training, pilots are instructed to treat the attitude indicator as the key instrument, given that it is the rotorcraft's attitude that dictates current airspeed and the need to adjust for changes in the altitude. In this regard, formal training serves the function of training novices to behave in a more anticipatory and less reactive fashion. This could account for better control performance and reduced workload.

A principled approach for diagnosing gaze selectivity in trainees could be useful in determining the skill levels of novices (cf. [1]). In fact, flight instructors sometimes rely on playbacks of eye-movements to identify inappropriate scanning strategy in trainee pilots [6]. The current analytical approach is one possible approach for quantifying gaze selectivity.

Finally, the spatial layout of visual instruments could be designed to encourage optimal scanning strategies. First, selective attention to relevant information could also be encouraged by removing the clutter of non-relevant instruments. Second, the spatial layout of instruments could emphasize the importance of a primary instrument by placing it in a central position. This could encourage users to return to it after referencing other instruments (cf., S1's scan pattern), rather than to treat every instrument equally.

Future work should be directed towards identifying effective scanning strategies for different maneuvers and key objectives. In the current work, constant airspeed and altitude control were emphasized above all other attributes (eg., flight path trajectory). It will be useful to determine whether the current principles for ideal instrument scanning are generalizable to different flight scenarios and maneuvers. A general model for ideal instrument scanning would be useful for training purposes as well as interface design.

References

1. Yang, J.H., Kennedy, Q., Sullivan, J., Fricker, R.D.: Pilot performance: Assessing how scan patterns navigational assessments vary by flight expertise. Aviation, Space, and Environmental Medicine 84(2), 116–124 (2013)
2. Anders, G.: Pilot's attention allocation during approach and landing- Eye-and head-tracking research in an A 330 full flight simulator. Focusing Attention on Aviation Safety (2001)
3. Nieuwenhuizen, F., Jump, M., Perfect, P., White, M., Padfield, G., Floreano, D., Schill, F., Zufferey, J., Fua, P., Bouabdallah, S., et al.: mycopter–enabling technologies for personal aerial transportation systems. In: Proceeding of the 3rd International HELI World Conference 2011, pp. 1–8 (2011)
4. MyCopter: Enabling Technologies for Personal Aerial Transport Systems, `http://www.mycopter.eu`
5. Perry, A.: The flightgear flight simulator. In: 2004 USENIX Annual Technical Conference, Boston, MA (2004)
6. Wetzel, P.A., Anderson, G.M., Barelka, B.A.: Instructor use of eye position based feedback for pilot training. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 42, pp. 1388–1392. SAGE Publications (1998)

Discriminability of Flight Maneuvers and Risk of False Decisions Derived from Dual Choice Decision Errors in a Videopanorama-Based Remote Tower Work Position

Norbert Fürstenau, Maik Friedrich, Monika Mittendorf,
Markus Schmidt, and Michael Rudolph

German Aerospace Center, Inst. of Flight Guidance, Braunschweig, Germany
Norbert.fuerstenau@dlr.de

Abstract. Future remote control of small low traffic airports (Remote Tower Operation, RTO) will rely on the replacement of the conventional control tower out-of-windows view by a panoramic digital reconstruction with high resolution and pan-tilt zoom (PTZ) video cameras as basic sensor system. This provides the required visual cues for aerodrome traffic control without a local control tower. Here we show that with a 2 arcmin-per-pixel resolution panorama system even with the use of a manually controlled (analog) PTZ camera (with PAL TV-resolution and selectable zoom factor setting) experiments under operational conditions indicate a significant increase of decision errors under RTO as compared to the conventional out-of-windows view. We quantify the corresponding discrimination difference by means of detection theory (discriminability, decision criteria) and Bayes inference (risk of false decisions) using the response errors of tower controllers with regard to dual choice decision tasks. The results extend the performance and subjective data analysis of safety related maneuvers in 11.

Keywords: Remote Tower Operation, visual cues, videopanorama, field testing, aerodrome circling, flight maneuvers, two-alternative decisions, signal detection theory, Bayes inference.

1 Introduction

Since about ten years remote control of low traffic airports (Remote Tower Operation, RTO) has evolved as a new paradigm to reduce cost of air traffic control 1. It was suggested that technology may remove the need for local control towers. Controllers could visually supervise airports from remote locations by videolinks, allowing them to monitor many airports from a remote tower center (RTC) 234. It is clear from controller interviews that usually numerous out-the-window visual features are used for control purposes 567. In fact, these visual features go beyond those required by regulators and ANSP's (air navigation service providers) which typically include only aircraft detection, recognition, and identification 7. Potentially important additional visual features identified by controllers in interviews involve subtle aircraft motion. In

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 105–114, 2013.
© Springer-Verlag Berlin Heidelberg 2013

fact, the dynamic visual requirements for many aerospace and armed forces tasks have been studied, but most attention has been paid to pilot vision (e.g. 8). A degradation of dynamic visual cues due to limited video framerates and its effect on decision errors during the landing phase was reported 910. Another group of visual cues are derived from flight maneuvers within the range of observability in the control zone. They might be indicative of aircraft status and pilots situational awareness which is important with the higher volume of VFR traffic in the vicinity of small airports.

These considerations led to the design of the present validation experiment within the DLR project RAiCe (Remote Airport traffic Control Center, 2008 – 2012). The field test was realized within a DLR - DFS (German ANSP) Remote Airport Cooperation (RAiCon). Specifically dual-choice decision tasks (the subset of "Safety related maneuvers" in 11) are used for quantifying the performance difference between the standard control tower work environment (TWR-CWP) and the new RTO controller working position (RTO-CWP). The present data analysis is based on objective measures from signal detection theory (SDT) and Bayes inference utilized in previous simulation experiments for deriving video frame rate requirements 910.

Experimental methods and results are provided in sections 2 and 3. In section 4 alternative methods (SDT and Bayes inference) are used for deriving estimates of discriminability, decision bias, and risk of false decisions, based on the measured response matrices. We finish with a conclusion and outlook in section 5.

2 Methods

In what follows only those aspects of the remote tower validation trial relevant for analysis of the two-alternative decision tasks are presented and discussed. Further details of the full passive shadow mode validation trial are reported in 11. The focus here is on quantifying the difference of discriminability and risc of false decisions under TWR-CWP and RTO-CWP conditions derived by objective measures via SDT 12 and Bayes inference and based on the analysis of dual choice decision errors.

2.1 Participants

Eight tower control officers (ATCO's) from DFS were recruited as volunteer participants for the experiment. The average age was 30 (stdev 12) years with 10 (stdev. 10) years of work experience, and they came from different small and medium airports. They took part at the experiment during normal working hours and received no extra payment. They were divided into 4 experimental pairs for simultaneously staffing the control tower (TWR-CWP) and the RTO-CWP.

2.2 Experimental Environment and Conditions

The experiment was performed as passive shadow mode test under quasi operational conditions on the four days July 17 – 20 2012. The remote tower system used in the present experiment was located at the DFS-operated Erfurt-Weimar (EDDE) control tower. It was an improved version of the RTO-experimental testbed at Braunschweig airport which was in use since 2004 for initial verification and validation trials 123. Figures 1 show the sensor system and the RTO-CWP with 200° - videopanorama and operator console based on a reconstructed far view with five HD-format 40”-displays (892 x 504 mm, 1920x1080 pixel, pixel distance = 0.47 mm). A separate monitor (left) displays the pan-tilt zoom camera which is controlled via a pen-input interaction display with virtual joystick, 12 preset viewing directions and selectable zoom factors Z = 2, 4, 8, 16 (viewing angles 26° - 3°). Additional monitors include standard information sources, (middle row from right to left) flight plan data, approach radar and weather display. In contrast to the experimental TWR-CWP the RTO position (RTO-CWP) was on the TWR ground floor in a separate room without visual contact to the airfield. The TWR-CWP was located close to the operational ATCO, but they were instructed not to communicate with each other.

Fig. 1. Remote tower installation with 200°-panorama and pan-tilt zoom camera sensor system at the tower roof (left photo), and operators workplace with 40”-HD-format displays (right)

Pre-defined flight maneuvers were generated with a DLR DO228 twin turboprop engine test aircraft (D-CODE, length 15.03 m, body height x width 1.8 x 1.6 m, wing span 16.97 m, wheel diam. 0.65 m) performing aerodrome circling. The set of well defined maneuvers as decision tasks at given positions within the EDDE control zone is indicated in Fig.2. During the experiment sometimes additional low volume normal traffic took place which now and then lead to delays in the traffic circle. Average duration of a full circle (= one run) was ca. 10 min yielding typically 140 min of experiment duration per participant pair for the nominally 14 full circles.

An example traffic circle is shown in Figure 2 with a 3D plot of the logged on board GPS trajectory. Radio communication between D-CODE pilots and flight engineer and the experimenter at the tower was realized with a separate research frequency in addition to the standard A/C – TWR radio channel. The available time for participant’s responses to decision tasks was limited so that correct, incorrect and non-answers were possible. At position D an altitude variation is observed. Trajectory minimum represents a runway overflight at about 30 m above ground.

2.3 Experimental Design and Task

Based on the fixed set of events during a single circle (A, .., H, I in Fig. 2) the concrete dual choice event (situation) alternative (S1, S2) for decision making switched statistically between runs according to two mirrored scenarios with different task sequences. In this way during the 14 circles per experiment / participant for each event about 7 answers per event alternative and per participant were obtained for the analysis (TWR and RTO CWP condition as independent variables). The within subjects design (TWR vs. RTO-CWP) yields about N = 56 answers for averaging with the 8 participants.

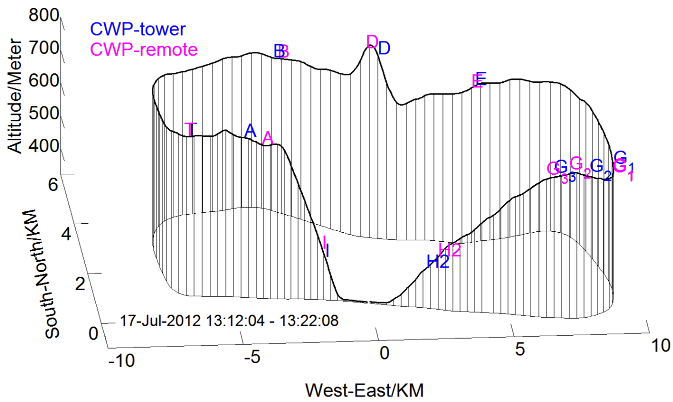


Fig. 2. DCODE trajectory measured with on-board satellite navigation. Letters indicate event positions with logged answers of TWR-CWP (blue) and RTO-CWP (red) operators to simultaneous task oriented on-line questioning. Distance between vertical lines = 5 s, projected to trajectory altitude minimum at ca. 350 m (sea level). Tower position (0, 0), height = 30 m.

During one aerodrome circling the two participants at the TWR and RTO-CWP had to simultaneously answer 19 different types of questions concerning the D-CODE maneuvers (events), object detection, and weather status. The following subset of 4 of the 9 event related questions were used for further evaluation by established objective methods (SDT and Bayes inference) for analysis of dual choice decision tasks (in brackets: maximum response time = interval until next task / question): 1. Does the aircraft perform repeated bank angle changes? (event position A; 30 s), 2. altitude variation ? (by 300 ft, event pos. D; 30 s), 3. landing light-off? Report status, event pos. G_2; 60 sec: switching-off not observable), 4. Landing gear-down? (Report during final approach (event pos. H_2, variable between 1.5 – 0.5 nm; 30 s). A subjective certainty rating on a 5-point scale was not evaluated for the present analysis.

Every pair of participants had to complete two experimental trials. For the first trial (duration approximately 140 min) in the morning till noon, the participants were randomly assigned to one of the two CWP's. Positions were changed for the second trial in the afternoon. The number of airport circles corresponding to the number of decisions for each specific event was between 10 and 14. With two alternative stimuli (S_1, S_2) per task each pair of participants was exposed to 5 – 7 stimuli of the same kind in random succession for each event under TWR and the same number under RTO conditions. The present data analysis is focused on deriving objective measures for the dual-choice decision tasks. Additional data evaluation is presented in a parallel paper 11 addressing performance (answers given, response times, and sources of information) and subjective measures (debriefing, questionnaires).

3 Results

The results can be organized in the form of response matrices which structure the measured estimates of conditional probabilities $p(y|S_1)$ = hit rate H, $p(y|S_2)$ = false alarms FA, $p(n|S_1)$ = misses M, $p(n|S_2)$ = correct rejections CR. Because participant's responses to event related questions were allowed to be positive, negative, and non-answers (no decision during the available time), we analyse two types of response matrices: a) (optimistic) neglecting non-answers, b) (pessimistic) interpreting non-answers as false decisions (M or FA). In this way we obtain for each of the decision tasks an optimistic and a pessimistic estimate with regard to decision errors which translates into corresponding risk, discriminability and subjective criteria (decision bias) values under analysis with signal detection theory (SDT) and Bayes inference.

Within the theoretical framework of detection theory the two alternative stimuli S_1, S_2 for each event define independent statistical variables. Each set of observations / decisions of a single subject for the 14 aerodrome circles with one of the events A, D, G_2, H_2 represents a sample of the randomly presented S_1- and S_2-alternatives, with the subjective responses assumed to be drawn from independent equal variance Gaussian densities ($\mu_{1,2}$, σ) for S_1 and S_2. Any quantitative discriminability difference between TWR and RTO is quantified by their corresponding coefficient $d' = \mu_1 - \mu_2 = z(H) - z(FA)$, and subjective decision bias (criterion) $c = 0.5(z(H) + z(FA))$, with z() = z-score as calculated from the inverse cumulative densities (see 4.2). Probabilities of alternative responses add to 1 (H + M = 1, CR + FA = 1) so that for the determination of d' and c, and graphical presentation in ROC space (receiver operating characteristic) the (H, FA) data set is sufficient.

Table 1 lists the corresponding hit and false alarm rates (± standard errors of mean ≈ 0.5 (95% conf. intervals)) for the four events to be analysed with respect to d' and c. In addition to H and FA the complementary rates CR and M are required for calculating the inverse conditional probabilities using Bayes inference (section 4.1).

Comparing the measured hit and false alarm rates for all four events under TWR and RTO conditions with non-answers not considered (optimistic case a): left two data columns), the RTO-CWP exhibits no significant difference as compared to the TWR-CWP. If however, the non-answers are interpreted as erroneous responses and correspondingly attributed to rates FA and M (pessimistic case b): right two data columns), significant differences TWR vs. RTO are obtained (smaller H(RTO), larger FA(RTO)) for event/task A (bank angle variation?), H (gear down?), G (lights off?), whereas for event/task D responses again exhibit not significant difference. The interpretation of the non-answers as erroneous responses appears to be justified due to increased uncertainty about the correct answer resulting in hesitation to respond at all

Table 1. Measured hit and false alarm rates ± stderr. of mean, for the four events and two conditions (TWR, RTO-CWP) with a) non-answers excluded (left two data columns) and b) non-answers attributed to error responses i.e. error rates FA and M (right two data columns)

Event with Alternatives S_1 or S_2	Distance ± stdev / km	CWP	a) Non-answers excluded		b) Non-answers included	
			$p(y\|S_1)$	$p(y\|S_2)$	$p(y\|S_1)$	$p(y\|S_2)$
A: bank angle var./ no var.	4.3 ± 0.4	TWR	0.92±.04	0.08±.04	0.81±.06	0.20±.05
	4.4 ± 0.4	RTO	0.93±.05	0.11±.05	0.60±.07	0.39±.07
D: Altitude var. / no var.	3.6 ± 0.6	TWR	0.80±.06	0.03±.03	0.77±.06	0.12±.06
	3.6 ± 0.6	RTO	0.73±.07	0.03±.03	0.70±.07	0.06±.04
G: lights off: y / no	7.0 ± 1.2	TWR	0.94±.04	0.25±.07	0.94±.04	0.28±.07
	4.0 ± 2.3	RTO	0.92±.06	0.63±.08	0.65±.08	0.72±.07
H: gear-down / up	2.2 ± 0.5	TWR	0.98±.02	0.06±.04	0.91±.04	0.22±.06
	2.1 ± 0.6	RTO	0.98±.02	0.07±.05	0.77±.06	0.37±.08

because tower controllers work ethics requires decision making with high certainty. This finding will be analysed and discussed in the next section. An extremely high FA difference TWR vs. RTO is observed both for case a) and b) for the "lights-off" event which is reflected also in a large difference of decision distance (correlated with response time).

4 Data Analysis and Discussion

There are of course several physical and psychophysical differences between the real out-of-tower view by human operators and the reconstructed far view with videopanorama and PTZ which leads to predictions concerning performance differences under the two conditions TWR and RTO-CWP due to technical limitations of the state-of the-art RTO technology 123. The measured performance, however depends on the usage of the different available information sources, e.g. videopanorama and PTZ. In the present work the measured performance difference is quantified in terms of SDT-discriminability difference d' and in terms of risk difference for detection errors as quantified by Bayes inference. The technical limitations leading to a prediction on RTO-performance are given by the systems modulation transfer characteristic (MTF), with the digital (pixel) camera resolution providing the basic limit (Nyquist criterion) for detectable objects (minimum resolvable visual angle $\delta\alpha \approx 2$ arc min $\approx$ 1/30° $\approx$ 0.6 m object size / km distance per pixel under maximum visibility. This is about half as good as the human eye (1 arcmin). Reduced visibility and contrast of course reduces the discriminability according to the MTF and the question arises how the discriminability difference TWR vs. RTO-CWP is affected. The gear-down situation with wheel diameter 0.65 m, e.g. can certainly not be detected before the wheel occupies, say, 4 pixels which for the 40" display (0.55 mm pixel size) means a viewing angle of ca 1 mm/2 m $\approx$ 0. 5 mrad corresponding to the visual resolution of the eye (1 arcmin) under optimum contrast. This estimate results in a panorama based gear-down detectability distance of < 500 m, leading to the conclusion that under RTO conditions this task requires usage of PTZ in any case for guaranteeing a high decision certainty. The same argument is valid for the other discrimination tasks.

4.1 Bayes Inference Analysis: Risk of False Decision

H, M, CR, FA are conditional probabilities which after the measurement are used together with the a-priori knowledge $p(S_i)$ as information for calculating via the Bayes theorem the risk of false decisions d_j, i.e. the (inverse) probability for occurrence of an actual situation conditional on the perceived one different from the actual one ($i \neq j$:

$$p(S_i|d_j) = p(d_j|S_i)\, p(S_i) / p(d_j) \tag{1}$$

with responses d_i, $i = 1, 2$, d_1 = yes, d_2 = no, and $p(y) + p(n) = 1$, $p(S_1) + p(S_2) = 1$, under TWR and RTO conditions. Of particular interest are the two probabilities for the risk of a situation contradicting the decision on the nature of the observed event. $p(S_1|n)$ is the probability (risk, likelyhood) of the aircraft with bank angle variation (situation S1, e.g. signaling some special situation during radio interruption) conditional the case that no variation is perceived. $P(S_2|y)$ is the probability for a situation with a/c not performing bank angle variation conditional on the false response "variation perceived". The following Table 2 contains the Bayes inference (risk) results for the four events. It clearly shows that for analysis b) these risks are on average at least two times as high for the RTO-CWP as compared to TWR- CWP, with the exception of event D.:

Table 2. Bayes inference for TWR and RTO-CWP from response data, for cases a), b)

Event with S_1 or S_2	CWP	a) Non-answers excluded		b)Non-answers included	
		$p(S_1\|n)$	$p(S_2\|y)$	$p(S_1\|n)$	$p(S_2\|y)$
A: bank angle var.	TWR	0.06	0.10	0.15	0.24
	RTO	0.06	0.13	0.33	0.46
D: Altitude var.	TWR	0.22	0.03	0.26	0.11
	RTO	0.26	0.03	0.30	0.06
G: lights off	TWR	0.06	0.26	0.06	0.29
	RTO	0.13	0.50	0.48	0.60
H: gear-down	TWR	0.03	0.04	0.14	0.15
	RTO	0.04	0.04	0.33	0.26

Important differences are observed for the analysis a) without non-answers considered (left two inference columns) as compared to the case b) with non-answers included (right two inference columns). The calculated risk for a situation occurring in contradiction to the perceived situation is very low for non-answers excluded and in most cases they increase significantly with non-answers included, which in fact is not surprising. Not expected was the result that in the RTO-CWP the risk in most cases at least doubles as compared to TWR-CWP. The altitude variation (event D) in contrast exhibits no significant difference which can easily be explained by the fact that due to the practically non-existent visual panorama visibility of the D-CODE at event position D the majority of decisions were made based on radar information.

4.2 Signal Detection Theory: Discriminability and Decision Bias

The Bayes inference analysis is supported by a more sophisticated evaluation of data from table 1 using SDT. In contrast to percentage correct (p_c) evaluation of subjects decisions on dual choice tasks it separates the decision maker's discriminability d' from the subjective decision bias c (= individual tendency to more conservative or more liberal decisions) 12.

Figure 3 depicts for analysis of case b) the (H, FA) data of the three tasks at H, A, G in ROC space together with two sets of pairwise ROC curves (one pair for TWR and RTO conditions each). One set (solid lines) is parametrized by discriminability d', the other (dashed) by the subjective decision bias c. E.g. d' = 3 means that the Gaussian densities mean values of perceived situations S_1, S_2 differ by 3 normalized stddev ($\sigma = 1$). Under the above mentioned conditions each (d', c)-ROC curve-pair is unambiguously determined by the single average (H, FA)-point. The d' and c values are calculated via standard procedures (inverse cumulative densities from the (H, FA) data). Dotted lines indicate estimates of standard deviations s(d') as described in 12, based on the binomial variation of measured proportions from sample to sample.

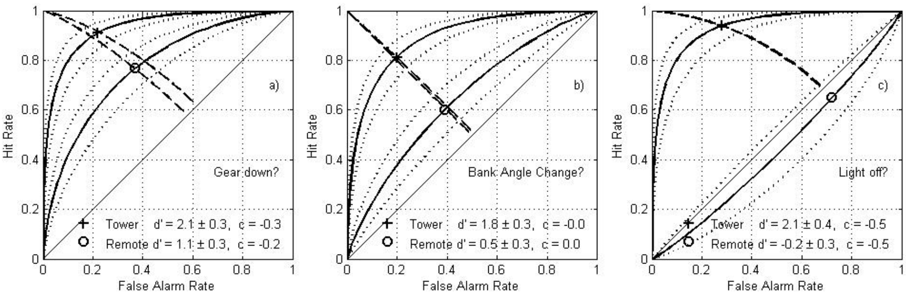


Fig. 3. Measured case b) data points in ROC space of average hit and false alarm rates for three of the four analysed events / tasks, in each case for the two TWR, RTO – conditions, together with the isosensitivity and isobias curves parametrized by discriminability d' and criteria c respectively. Dotted lines are standard deviations based on procedures described in 12

The following table 3 summarizes the discriminability and criteria (decision bias) data corresponding to Figure 3 a), b), c), and like table 2 includes event/task D (altitude variation) with both data analysis cases: optimistic a) and pessimistic b). Again, like with the Bayes inference the case a) analysis shows no significant difference between TWR and RTO-CWP conditions if the non-answers are not considered in the data analysis, with the exception of task G where even with non-answers not considered RTO exhibits a significant decrease of discriminability.

Also for case b) analysis the Bayes inference results are confirmed: again with the exception of task D (altitude change) significant decrease of discriminability is

Table 3. Discriminability d‘ and certainty c as obtained from z-scores based on response matrices (hit and false alarm rates). For uncertainties see Fig. 3.

Event	CWP	a)Non-answers excluded		b) Non-answers included	
		d’	c	d’	c
A	TWR	2.81	-0.01	1.75	-0.02
	RTO	2.72	-0.11	0.54	0.00
D	TWR	2.69	0.49	1.90	0.22
	RTO	2.48	0.62	2.07	0.52
G	TWR	2.24	-0.45	2.14	-0.49
	RTO	1.05	-0.86	-0.20	-0.48
H	TWR	3.63	-0.30	2.12	-0.30
	RTO	3.47	-0.30	1.07	-0.20

observed if non-answers are attributed to erroneous decision (M, FA), for task G (landing lights off) even zero detectability. Decision bias in most cases does not exhibit significant differences between TWR and RTO-CWP.

5 Conclusion

The present detailed analysis of two-alternative decision making with safety related aircraft maneuvers provides an answer to the observed discrepancy in the percentage correct (p_c) analysis of the corresponding observation data in 11 as compared to the subjective success criteria. The perceived safety was rated as insufficient [11] which agrees with the objective data of the present analysis. 11Nevertheless it only represents a first step towards quantification of RTO performance and safety compared to TWR-CWP. Neglecting non-decisions during simultaneous decision making at TWR- and RTO-CWP shows mostly no significant difference of error risk and discriminability d’ whereas the interpretation of non-decisions as false responses (misses or false alarms) leads to significant error increase under RTO as compared to TWR conditions and correspondingly to reduced d’. This (worst case) result comes not unexpected for the state-of-the-art technology (2 arcmin visual resolution (Nyquist limit) of panorama = half the human eye performance, and manually controlled PTZ). It may be expected that increased automation (e.g. automatic PTZ-object tracking based on movement detection and data fusion with Mode-S transponder and approach radar) and future 1-arcmin panorama resolution with improved contrast will increase discriminability. A confirmation of the initial results presented in the present paper and in 11 requires more field trials under comparable conditions with improved statistics and subjects instructions requesting forced decisions for avoiding non-answers.

Acknowledgement. We are indepted to DFS personnel N. Becker, T. Heeb, P. Distelkamp, and S. Axt for support during preparation of the experiment. Many thanks are due to C. Möhlenbrink and A. Papenfuß for performing part of the online interviews during the exercises. A. Grüttemann was responsible for the onboard data acquisition and served as flight engineer. We acknowledge the support of the DLR flight experiments department and in particular cooperation with pilots G. Mitscher and P. Bergmann.

References

1. Schmidt, M., Rudolph, M., Werther, B., Möhlenbrink, C., Fürstenau, N.: Development of an Augmented Vision Video Panorama Human-Machine Interface for Remote Airport Tower Operation. In: Smith, M.J., Salvendy, G. (eds.) Human Interface, Part II, HCII 2007. LNCS, vol. 4558, pp. 1119–1128. Springer, Heidelberg (2007)
2. Fürstenau, N., Schmidt, M., Rudolph, M., Möhlenbrink, C., Papenfuß, A., Kaltenhäuser, S.: Steps towards the Virtual Tower: Remote Airport Traffic Control Center (RAiCe). In: Proc. EIWAC 2009, ENRI Int. Workshop on ATM & CNS, Tokyo, March 5-6, pp. 67–76 (2009)
3. Schmidt, M., Rudolph, M., Papenfuß, A., Friedrich, M., Möhlenbrink, C., Kaltenhäuser, S., Fürstenau, N.: Remote airport traffic control center with augmented vision videopanorama. In: Proc. 28th IEEE-DASC 2009, Orlando, pp. 4.E.2-1–4.E.2-15 (2009)
4. Hannon, D., Lee, J., Geyer, T.M., Sheridan, T., Francis, M., Woods, S., Malonson, M.: Feasibility evaluation of a staffed virtual tower. The Journal of Air Traffic Control, 27–39 (Winter 2008)
5. Ellis, S.R., Liston, D.B.: Visual features involving motion seen from airport control towers. In: Proc.11th IFAC/-IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France, September 31-October 3 (2010)
6. Ellis, S.R., Liston, D.B.: Static and motion-based visual features used by airport tower controllers. NASA TM-2011-216427, Ames Research Center, Moffett Field, CA (2011)
7. Van Schaik, F.J., Lindqvist, G., Roessingh, H.J.M.: Assessment of visual cues by tower controllers. In: Proc. 11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Valenciennes, France, August 31-September 3 (2010)
8. Watson, A.B., Ramirez, C.V., Salud, E.: Predicting visibility of aircraft. PLoS ONE 4(5), e5594 (2009), doi: 10.1371/journal.pone.0005594 (Published online May 20, 2009)
9. Fürstenau, N., Mittendorf, M., Ellis, S.R.: Remote Towers: Videopanorama frame rate reqirements derived from visual discrimination during simulated aircraft landing. In: Schaefer, D. (ed.) Proc. SESAR Innovation Days, Eurocontrol (2012) ISBN 978-2-87497-024-5
10. Ellis, S.R., Fürstenau, N., Mittendorf, M.: Frame Rate Effects on Visual Discrimination of Landing Aircraft Deceleration: Implications for Virtual Tower Design and Speed Perception. In: Proceedings Human Factors and Ergonomics Society, 55th Annual Meeting, Las Vegas, NV, USA, September 19-23, pp. 71–75 (2011)
11. Friedrich, M., Möhlenbrink, C.: How realistic can a concept be tested? A field trial for validating a remote tower operation concept. In: 10th USA/Europe Air Traffic Management Research and Development Seminar (ATM 2013) (submitted, 2013)
12. MacMillan, N.A., Creelman, C.D.: Detection Theory. Psychology Press, Taylor and Francis (2005)

Single-Seat Cockpit-Based Management of Multiple UCAVs Using On-Board Cognitive Agents for Coordination in Manned-Unmanned Fighter Missions

Stefan Gangl, Benjamin Lettl, and Axel Schulte

Universität der Bundeswehr München (UBM), Department of Aerospace Engineering
Institute of Flight Systems (LRT-13), 85577 Neubiberg, Germany
{stefan.gangl,benjamin.lettl,axel.schulte}@unibw.de

Abstract. This article describes an automation concept, which enables the pilot of a single-seat fighter aircraft to manage more than one unmanned combat aerial vehicle (UCAV). The presented concept bases on the theory of cognitive and cooperative automation and suggests that unmanned aircraft are equipped with on-board artificial cognitive units (ACUs). By this, unmanned platforms are enabled to exhibit cooperative capabilities and rational behavior in the context of the mission. To accomplish efficient manned-unmanned cooperation the concept additionally proposes to support the pilot with an assistant system module for team coordination tasks and to provide a self-explanation capability to the unmanned aircraft. This concept has been realized as laboratory prototype and already been tested with operational personnel in our human-in-the-loop full scenario simulation environment. For the further evaluation of the concept an experimental design has been worked out.

Keywords: multi-UAV, multi-UCAV, mission planning, assistant system, manned-unmanned teaming, human-machine interaction, cognitive automation, cooperative automation, artificial intelligence, human-automation integration.

1 Introduction

Over the past years, unmanned combat aerial vehicles (UCAVs) have become increasingly important in the military field. Up till now they are mainly in single vehicle missions operated from ground control stations in use. At the same time, many multi-ship missions are still conducted by solely manned fighter teams. It is reasonable that in the upcoming years manned and unmanned forces become more and more integrated. This inevitably leads to increasingly complex information flows between manned and unmanned units. One possibility to encounter this concern on an organizational level is to guide UCAVs airborne on a high Level of Interoperability (LOI) [4], like LOI 4 or 5, as according to STANAG 4586. This means, that there are one or more highly automated unmanned aircraft under control of a manned platform including aircraft guidance and mission payload.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 115–124, 2013.
© Springer-Verlag Berlin Heidelberg 2013

The introduction of highly complex automation aboard the UAV poses the serious challenge of supervising and monitoring of the automation to the human operator. Experiences with automation systems for manned aircraft show that complex automation can lead to automation induced errors [3][10] and may raise the human workload in situations where the workload is already at a critical level.

An automation concept to accomplish such a team configuration while avoiding potential issues in cooperation with complex automation has been developed at the Institute of Flight Systems at the University of the Bundeswehr Munich (UBM) and is presented in this article. Our work is based upon previous results [9][8] concerning unmanned cooperating fighters performing a SEAD mission. A major add-on in this study is the investigation of the role of the human operator.

Related work in this field has been reported in [1] and [2]. Those approaches are based on the guidance of intelligent unmanned capabilities on high level mission commands [2]. The operator, the pilot of the manned aircraft, passes specific mission tasks to a "pool" of co-operating UAVs. The unmanned systems self-organize and perform all of these tasks independently. Therefore, the unmanned vehicles are furnished with a multi-agent system (MAS), consisting of different types of software agents with various abilities. This concept was successfully tested in real flight tests (one manned fighter, one real and three simulated UAVs) in spring 2007 [1].

2 Cognitive and Cooperative Automation

2.1 Cognitive Automation

The automation design approach of cognitive automation was developed at the Institute of Flight Systems and is described in detail by Schulte & Onken [10]. In the following subsection a summary of the most important aspects with respect to this article is given.

In the domain of flight guidance up to now mainly what we call conventional automation is in operation. Operators supervising complex automated systems are exposed to several problems like "opacity", "literalism" or "brittleness" as stated by the well-known critics on aircraft automation by Billings [3]. These problems are even increased in the case of UAV flight guidance due to even more complex missions and automation. To overcome these shortcomings of conventional automation the cognitive automation approach proposes the usage of automation with cognitive capabilities like decision making, problem solving and planning.

Suchlike automation will, in our case, be implemented by so-called Artificial Cognitive Units (ACUs). These are basically artificial, knowledge-based agents, which are able to develop rational goal-directed behavior by processing knowledge.

The approach of cognitive automation suggests two possibilities for implementing this kind of automation in form of ACUs in a human-machine system design. These are called the "dual-modes" of cognitive automation [10] and will be explained by use of the work system framework. A work system consists of two basic components, there is an Operating force (OF) and there are Operating Supporting Means (OSMs). The OF, which is always represented by at least one human operator, understands and pursues the overall work objective. It generates a comprehensive understanding of the actual situation and knows the capabilities of the available OSMs. Based on this knowledge an OF is able to decide for the course of action to comply with the work objective by using the OSMs. In contrast to this, the OSMs do not know the overall mission objective, they only accomplish sub-tasks assigned by the OF. In the dual-mode cognitive automation concept an ACU can be either part of the OF or of the OSMs (Fig. 1). These ACUs are called Operating Cognitive Unit (OCU) and Supporting Cognitive Unit (SCU), respectively. From a human operator's point of view, an SCU is a subordinate system deployed in a supervisory control relationship whereas an OCU teams up with the human operator, both of which composing the OF. The relationship between the human operator and an OCU is of special interest in the context of this article and is therefore presented in more detail. This relationship can show different characteristics mainly depending on qualification and capabilities of the cooperating entities.

For example in a pilot/co-pilot relationship, both team mates have the same expertise. Therefore, takeover of tasks is possible and mutual understanding of each other behavior/misbehavior can be supposed. In a human-machine team with similar cooperative characteristics, the machine teammate acts as an OCU and is also called an assistant system in our notion. Such an assistant system can, as according to Onken & Schulte [10], appear in three different roles as cooperating partner in the OF, i.e. associative assistance, alerting assistance and/or (temporarily/permanently) substituting assistance. The micro-behavior of the assistant system is defined by the three basic requirements for human-automation co-action [10].

Like in human teams a specific kind of cooperative relationship can be found, which is of particular relevance for the concept, which will be described in chapter 3. Given e.g. a medical team consisting of a surgeon and an anesthetist doing a surgery. The team members of such a team have different abilities and qualifications. Therefore, they cannot be replaced by each other and takeover of tasks is only possible to a limited extend. Work-share and responsibilities are defined prior to assignment of a work objective. Both team members know the overall work objective and pursue it in cooperation with each other. The contribution of both team members is needed to comply with the work objective.

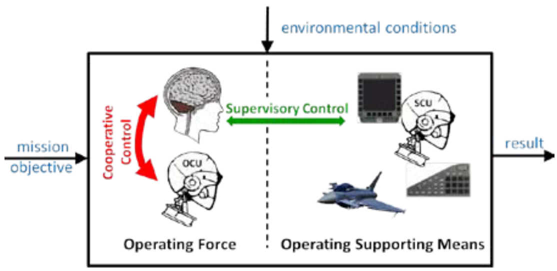


Fig. 1. Concept of Dual-Mode Cognitive Automation

Therefore, we propose to define a similar human-cognitive agent team in analogy to a solely human team. If we want to integrate a cognitive agent being part of such a team in a work system, as according to the dual-mode concept, this agent acts as an OCU in the permanently substituting role. To clearly distinguish this kind of OCU for the purpose of our work, it will be called "complementing OCU" in contrast to an "assisting OCU". Under consideration of a more general background of cooperation, the cooperative relation of such a man-machine team is discussed in the next section.

2.2 Cooperation

In this subsection guidelines are derived for the design of machine team members, which are supposed to cooperate with humans.
A general definition of cooperation on which is widely referred to origins from Deutsch [5] and states: *"Cooperation is the situation where the movement of one member towards the goal will to some extent facilitate the movement of other members towards the goal"*. According to this definition a cooperative situation is basically characterized by "the goal" and by "the movements" on which cooperating individuals interfere. Consequently two necessary conditions can be derived, which individuals have to fulfill to be part of a cooperative situation:

1. ***Strive towards complementary or identical goals***, *if this is not the case team members are not in a cooperative but in a competitive situation.*
2. ***Capable to coordinate***, *which means to be able to manage interdependencies or to do rational "movements".*

A human-machine team in which individuals comply with these conditions is able to process assigned tasks in a rational and effective way, but not mandatory. It can be assumed, that a human who cooperates with machine team members will find himself/herself in situations where he/she not understands what the teammates are doing or why they are doing what they do. This lack in understanding leads to instantaneous confusion, as well as lasting decline of trust in machine agents by the human. Moreover, even if the human team member is able to understand the observed behavior the process of reach understanding by interpretation of available information is a rather cognitively demanding one. The issues of lack of understanding of behavior, a low level of trust and the demanding process of generating understanding lead to potential rejection of cooperation by humans.

Consequently, the above-mentioned necessary conditions are extended by two sufficient conditions, which machine team members should fulfill at their best to finally yield good cooperative performance:

3. *Support the* ***generation of mutual understanding and acceptance*** *of behavioral pattern of the other.*
4. *Support the* ***development of trust*** *by team members toward own behavior.*

3 Automation Concept for a Manned-Unmanned Fighter Team

In this chapter the proposed automation design concept for a manned-unmanned fighter team is developed for a given application. In a first step a manned fighter team will be analyzed. Based on this analysis and under consideration of the theory of cognitive automation and cooperation as described in chapter 2, the concept is developed.

Finally, in the following subsections the most important parts of the concept are discussed in more detail.

Assuming a Close Air Support (CAS) mission with the objective to engage a by SAM-sites protected target with a laser guided bomb (LGB) is conducted by a manned team. One team member would have the expertise and resources to attack the target, maybe it is a pilot trained for interdiction strike missions (IDS). The other three team members form a sub team with the supporting ability to take photos, suppress SAM-sites or designate the target. They are maybe qualified for reconnaissance (RECCE) or suppression of enemy air defense missions (SEAD). Analyzing such a team set-up by the work system framework, results in a work system configuration as shown in Fig. 2 left. Every pilot is in a supervisory control relationship to his own aircraft and systems, as shown by the green arrows, and in a cooperative relationship to his team, as shown by the red arrow. In a next step we have a closer look to the cooperative relationship between the attack pilot and the supporting sub-team. Both cooperating individuals in this relationship know the overall work objective, workshare and responsibilities between them are defined with the assignment of a work objective and takeover of tasks is not possible because they are specialized experts. Consequently this relationship has similar characteristics as the one in the surgery team exemplarily mentioned in section 2.1.

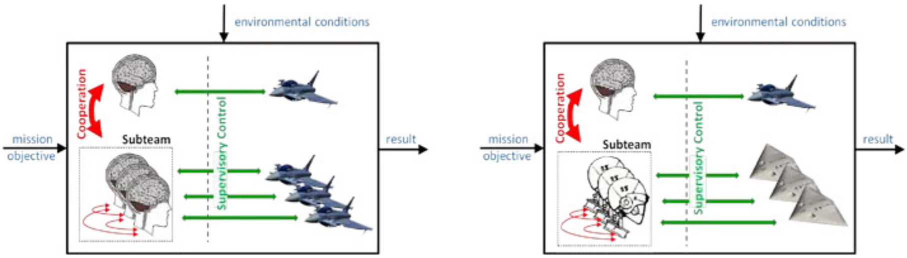


Fig. 2. Work systems of a manned (left) and a manned-unmanned fighter team (right)

Transferring such a team set-up to a human-machine team by use of cognitive agents, we end up in a work system configuration as shown in Fig. 2 right. The supporting pilots are substituted by complementing OCUs and consequently the manned platforms by unmanned ones. The control relationships from the attack aircraft pilot's point of view remain the same. This means the pilot supervises the own aircraft and systems and cooperates with the unmanned sub-team. In our case the unmanned sub-team members will not be tasked individually, but on the level of the whole team.

These OCUs must have the capability to utilize their individual platforms to implement decisions, to cooperate with the unmanned teammates, and last but not least to cooperate with the human in the context of the mission. These agents are implemented by use of the Cognitive System Architecture (COSA) [11]. The COSA framework offers the possibility to model domain knowledge subdivided into knowledge-packages with a rule-based syntax. To provide the required capabilities they are given knowledge on system management, the environment, the mission and cooperation, all of which implemented as so-called "knowledge packages" [8][9].

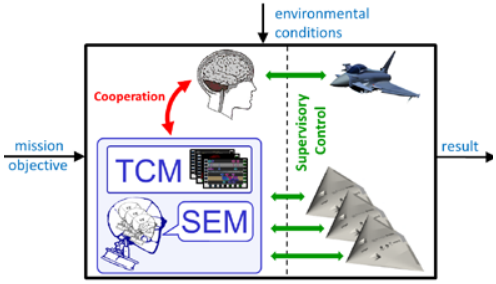


Fig. 3. Work system of proposed concept

Assuming good team performance of this OCU-Team as such, the cooperative relationship between the unmanned team and the manned platform can be identified as the most important parameter with respect to overall performance of the work system. Therefore, an additional focus of this concept is to ensure efficient cooperation between the pilot and the unmanned team in all mission phases. To accomplish this, two additional system components are proposed. First of all we provide a Team-Coordination-Module (TCM) to the pilot. This module comprises different assisting system components with the goal to support the pilot by team coordination tasks. In addition it offers a communication interface to inform the team about mission objective and constraints. In subsection 3.2 the components of this module are described in detail. Secondly, we propose to expand the abilities of the OCUs with a Self-Explanation-Module (SEM), which enables them to provide explanation about own actions and decisions to the pilot. The purpose of this measure is to support understanding of behavioral patterns and to support development of trust towards machine behavior in line with the sufficient conditions for cooperation as derived in chapter 2.2. More specific aspects on this module are discussed in subsection 3.1. The whole assistant system composed of the core elements OCU-Team, Team-Coordination-Module and Self-Explanation-Capability is visualized in the work system in Fig. 3.

3.1 Self-Explanation-Module

Due to communicational shortcomings, machine teammates accomplish the sufficient conditions for cooperation inadequately. To overcome this deficiency we suggest a self-explanation-module for cognitive agents, which enables them to provide causal information about own actions and decisions to human team mates.

The concept for this module proposes a self-explanation capability based on two modes of explanation as shown in Fig. 4. The first mode is called “continuous self-explanation” (SEC-Mode) and the second one “self-explanation on demand” (SED-Mode). By the SEC-Mode we understand the advanced capability of an agent to provide relevant situation adjusted explanations on own initiative. In contrast the SED-Mode provides the possibility to the operator to request an explanation. The SED-Mode is designed to complement the SEC-Mode. All explanations are presented to the pilot by verbal output generated by a speech synthesizer.

In order to enable the ACU to exhibit self-explaining behaviors, its knowledge base is extended by specialized knowledge packages (see dashed box of Fig. 4). The “Natural Language” package enables an ACU to provide explanations in morphological and syntactical correct language. With the “Own Status” knowledge package an

ACU is aware of own behavior and the cause for it. This enables the cognitive agent to provide at any time information which are of interest to answer question like “What am I doing?” and “Why am I doing this?”. The knowledge package “Evaluation of the need for explanation” is only of relevance in the SEC-Mode. It permits an agent to reason about the necessity of an explanation and therefore initiates explanations in the SEC-Mode. In contrast, in the SED-Mode the pilot initiates an explanation by simply asking a question regarding the status of a cognitive agent. This request is processed by a speech recognition system and forwarded to the ACU.

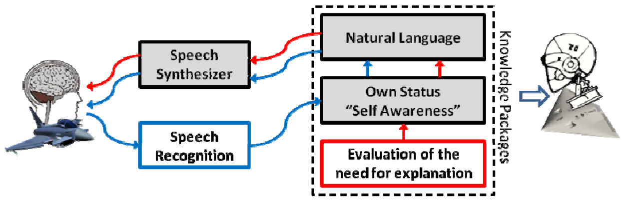


Fig. 4. Concept of a two mode self-explanation capability, SEC-Mode (red), SED-Mode (blue)

3.2 Team-Coordination-Module

The entire mission can be broken down into individual sub-tasks as mentioned before in section 3. These sub-tasks are distributed among all team members according to their capabilities. Mainly “prerequisite” and “simultaneity” dependencies exist between sub-tasks within the mission context. A typical prerequisite dependency is the flight of the manned fighter aircraft into hostile territory. The pilot should enter threatened areas only, if known threats are successfully suppressed. The attack of the mission target is an example for a simultaneity dependency. An unmanned team member needs to designate the target, while the manned fighter is deploying the ordnance. In order to enable suchlike coordination between the pilot and the unmanned sub-team we developed a team-coordination module (TCM) as part of the assistant system.

The TCM supports coordination in different roles of assistant systems [10]. Basically it works in the role of associative assistance and continuously provides necessary coordination information to the pilot. If the TCM identifies a problem between the coordination of sub-tasks, the TCM informs the pilot in the role of alerting assistance. If this warning is not handled in a specific time period, e.g. because the pilot may be overtaxed, the TCM takes over certain activities to cope with the situation. In this case the TCM works in the role of a temporarily substituting assistant.

The necessary coordination information including the pilot’s sub-task dependencies within the team will be displayed on multifunctional head down displays (MHDD) in the cockpit. The visualization on the display can be either a spatial (Fig. 5, left) or a temporal (Fig. 5, right) representation. The left screen-shot shows a section of the advanced map display. The aircraft symbols and the appropriate icons on the map represent the position and the actual carried out sub-task of each team member. For example, the green team mate is designating the target. Furthermore the red circular lines mark the threatened area of hostile SAM-sites. If the SAM-site is suppressed the line is dashed otherwise it is continuous. The threat status is usually communicated by the unmanned team members as prerequisite coordination information. The two trapezoidal elements in the map support the coordination between the two different time-critical simultaneity sub-tasks, laser designation and weapon release. By the right figure a section of the onboard timeline display is illustrated. The upper timeline indicates the mission sequence of the manned aircraft and the bottom timeline that of the UAV-team in a temporal manner. Through the upper timeline the pilot is continuously informed when a mission relevant event for him will occur and when a specific sub-task has to be performed. The timeline at the bottom is primarily dedicated to increase the pilot's situation awareness. Therefore, all sub-tasks including estimated start times and execution durations of the entire unmanned team are figured out.

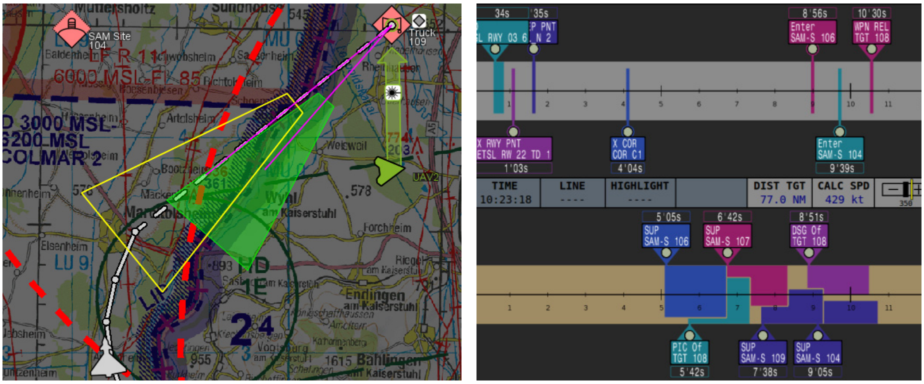


Fig. 5. Spatial and temporal visualization of coordination information

4 Evaluation

The system designed along the lines of this concept has been integrated in a fighter cockpit and mission simulator at UBM. This simulation environment facilitates human-in-the-loop experiments with a modern generic single-seat fighter cockpit representing the manned component of the strike package with up to three unmanned aircraft in different scenarios.

In first pre-experimental runs our concept was evaluated by German Air Force Pilots. In these runs pilots approved the usability of the simulation environment for realistic human-in-the-loop experiments. In addition military relevance of selected evaluation scenarios was confirmed. All conducted missions were accomplished successfully, indicated by the successful abatement of the target object, adherence of time constraints and no observed losses of UAVs. According to participating pilots' workload situation during mission execution was always on a manageable level, provided by subjective ratings.

To evaluate the implemented concept systematically, a human-in-the-loop experimental campaign with six experienced German Air Force pilots is scheduled for end of April 2013.

In this campaign pilots have to accomplish CAS missions on basis of a work objective assigned to the manned-unmanned team from a super-ordinated unit. Starting with an initial situation as shown in Fig. 6, subsequently unmanned platforms enter enemy territory and start to suppress relevant SAM-sites. Meanwhile the manned aircraft crosses the corridor and pictures of the intended target are taken by one UAV and provided to the fighter pilot. The pilot verifies the target and one UCAV prepares for target designation. As soon as the fighter aircraft is in position for weapon deployment the laser designation is started and the LGB is released. After weapon impact, new reconnaissance pictures are provided to the pilot for battle damage assessment. In case of mission success, the team returns to friendly area.

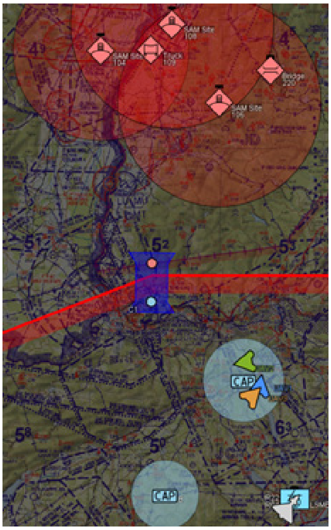


Fig. 6. Overview reference mission

Table 1. Variation of the independent variable

Setup	SEM	TCM	
		spatial	*temporal*
1	X	X	X
2		X	X
3	X	X	
4			X

Due to the limited number of available test subjects, a within-subject design was selected as experimental design. As independent variables variants of the assistant system's components, namely the Team-Coordination-Module (spatial and temporal information representation) and Self-Explanation-Module, will be investigated. Every participating test subject has to pass four experimental runs, with different sets of independent variables as shown in Table 1. The dependency between the quality of cooperation towards unmanned team mates and the selected set of cooperation supporting components will be the subject of investigation. The quality of cooperation is therefore the depended variable of this experiment. To evaluate this variable we characterize the quality of the cooperation by the parameters mission performance, workload of the pilot from team management tasks, trust toward team members and team situation awareness.

These parameters will be quantified in all experimental runs by different measurements. Team Situation Awareness will be measured by the well-established situation awareness tests SART [12] and SAGAT [6]. Mission performance is indicated by a selection of parameters like mission success or overall mission execution time. The task load of the pilot by team management tasks is determined by summing up the times during which he is busy with tasks like monitoring the team. This will be accomplished by analyzing attention distribution by use of gaze measurements. To analyze trust between the pilot and the unmanned assets the trust level during all runs will be measured by the subjective trust rating scale developed by Jian et al. [7].

In addition to this, trust of the pilot is rated by analyzing video streams of the experiments towards “double-check” behavior of the pilot, by this we understand that the pilot tends to verify information received from the unmanned team members.

Results of this experimental campaign will be provided in upcoming publications.

References

1. Baxter, J., et al.: Fly-by-agent: Controlling a pool of UAVs via a multi-agent system. Knowledge-Based Systems 21(3), 232–237 (2008)
2. Baxter, J., Horn, G.: Controlling teams of uninhabited air vehicles. In: Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 27–33. ACM (2005)
3. Billings, C.: Aviation Automation: The Search for a Human-Centered Approach. Lawrence Erlbaum Associates Publishers, Mahwah (1997)
4. Cummings, M., et al.: STANAG 4586 Human Supervisory Control Implications (2006)
5. Deutsch, M.: Cooperation and trust: Some theoretical notes. In: Nebraska Symposium on Motivation, pp. 230–275 (1962)
6. Endsley, M.R.: Situation awareness global assessment technique (SAGAT). In: Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, pp. 789–795. IEEE (1988)
7. Jian, J., et al.: Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics 4(1), 53–71 (2000)
8. Meitinger, C., Schulte, A.: Cognitive machine co-operation as basis for guidance of multiple UAVs. RTO-Meeting Proceedings HFM-135, Biarritz (2006)
9. Meitinger, C., Schulte, A.: Human-UAV Co-operation Based on Artificial Cognition. In: Harris, D. (ed.) EPCE 2009. LNCS, vol. 5639, pp. 91–100. Springer, Heidelberg (2009)
10. Onken, R., Schulte, A.: System-Ergonomic Design of Cognitive Automation. SCI, vol. 235. Springer, Heidelberg (2010)
11. Putzer, H., Onken, R.: COSA–A generic cognitive system architecture based on a cognitive model of human behavior. Cognition, Technology & Work 5, 140–151 (2003)
12. Taylor, R.M.: Situational Awareness Rating Technique(SART): The development of a tool for aircrew systems design. Situational Awareness in Aerospace Operations, 23–53 (1990)

Distributed Cognition in Flight Operations

Don Harris

Faculty of Engineering and Computing
Coventry University
Priory Street
Coventry CV1 5FB
United Kingdom
don.harris@coventry.ac.uk

Abstract. Human Factors is no longer simply concerned with the design of equipment and work stations. This old view is being superseded by a systems-based approach which examines all aspects of the working environment and makes little or no attempt to separate the human, machine and task environment. This socio-technical systems approach complements the latest thinking from cognitive science which regards the human use of technological artifacts as a joint cognitive system. People work in teams, who all have a slightly different perspective of the system; the tools that they use serve as 'cognitive amplifiers' to enhance human abilities. This brief overview begins by examining the operation of commercial aircraft as a joint cognitive system and examines the role of CRM in promoting distributed cognition on the flight deck.

Keywords: Distributed cognition, Joint Cognitive Systems, Crew Resource Management, Distributes Situation Awareness.

1 The Operation of Commercial Aircraft as a Joint Cognitive System

Any aircraft is a small socio-technical system operating within a larger socio-technical system. The air transport system as a whole is a very large, complex system of socio-technical systems [1]. Socio-technical systems contain people, equipment and organizational structures linked by functional processes (which are essential for transforming inputs into outputs) and social processes, which are informal but which may serve to either facilitate or hinder the functional processes [2]. Hollnagel [3] illustrated this issue in terms of the layers relating to the aviation Joint Cognitive System (JCS). Hollnagel suggested that the JCS relating to an airliner could be characterized in a similar manner to the skins of an onion (see Figure 1). It was merely a question concerning the desired unit of analysis the determined the bounds of the system under examination, not the ultimate bounds of the system *per se*.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 125–133, 2013.
© Springer-Verlag Berlin Heidelberg 2013

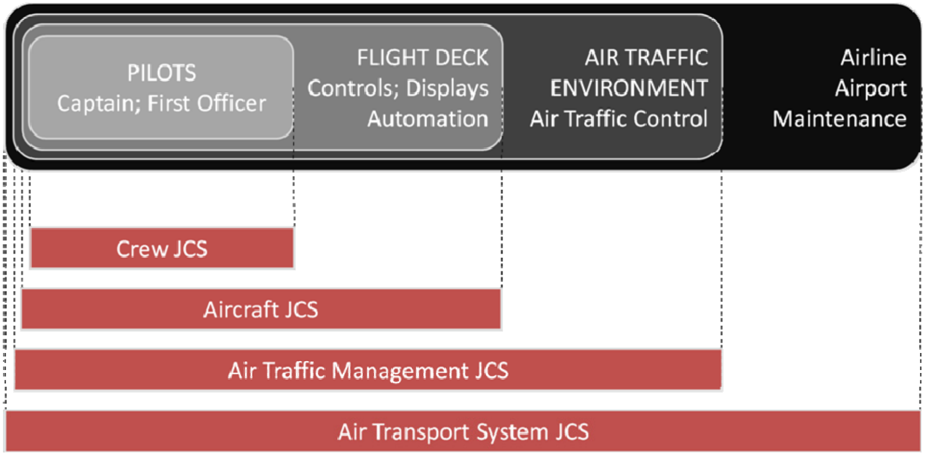


Fig. 1. The many bounds of a JCS for the air transport system (adapted from Hollnagel, E. (2007). Flight Decks and Free Flight: Where are the System Boundaries? *Applied Ergonomics, 38,* 409-416).

2 Distributed Cognition

Distributed cognition proposes that knowledge and cognition are not confined to the individual but are distributed across a number of interacting people and/or tools and objects. This approach takes a system-wide based view of people interacting with non-human (often technological) artifacts within an environment, where the emphasis in analysis is placed upon understanding how data, information and/or knowledge is represented and used. Cognition is a dynamic and emergent construct taking place in context. Key to this is the interchange of information between human and machine agents in a system and its representation. Hence it is necessary to describe how cognition is distributed and coordinated.

Salomon [4] proposed two general categories of distributive cognition: shared cognition and off-loading. The former category describes interactions between people engaged in a common activity; cognition across a group (the individual and shared representation of the situation) changes as their interactions progress. The latter category, off-loading, describes cognitive tool use.

2.1 Shared Cognition

Rogers [5] described four generic properties of distributed cognition in people working as a team, an instance of shared cognition:

- Cognitive systems comprising more than one person have properties over and above those individuals making up the system (e.g. an aircraft and its crew).

- The knowledge possessed by members of such a system is variable and redundant: teams working together on a task possess different kinds of knowledge and so engage in interactions that allow them to pool their cognitive resources.
- Knowledge is shared through via formal and implicit communication with prior knowledge of each other, enabling them to engage in heedful interrelating.
- Distribution of access to information and sharing access and knowledge promotes coordinated action.

Distributed cognition is predicated upon a degree of common understanding of a situation; about the aims and objectives of the task and the required method of achieving the goal. However, not every member of crew needs to know everything (indeed this would be very inefficient). For instance, when undertaking an ILS approach with one pilot flying and the other pilot monitoring both pilots should have a common understanding of the situation which overlaps completely. However, during the normal conduct of the flight the crew will not have such a close (shared) appreciation of their situation. There will be some common elements to their Situation Awareness (SA) for example, where they are and what their immediate and longer term intent is; many elements ‘overlap’. However, each crew member will also be concentrating on the individual responsibilities associated with their role as Pilot Flying (flying the aircraft; dealing with navigation and general aircraft operation) or Pilot Not Flying (monitoring the flying pilot; monitoring aircraft performance; handling the radios; being responsible for monitoring the weather and for running the checklists). Thus, each pilot will be solely aware of several different things. They require SA for those factors relevant to undertake their duties for a specific task in a particular mission phase. These knowledge components are role specific but inter-dependent. The major challenge to achieve wider crew SA is in the co-ordination of these crew resources. To this end Endsley and Jones [6] developed a model of team SA comprised of four components:

- *Requirements* – What information and goals need to be shared between crew members?
- *Devices* – What devices are available for sharing this information (communication devices; visual and/or auditory displays, etc).
- *Mechanisms* – What mechanisms do crew members possess which support their ability to interpret information in the same way (such as shared mental models to facilitate communication and coordination)?
- *Processes* – What formal processes are used for sharing information; verifying understanding; prioritizing tasks and establishing contingencies, etc?

The ‘*requirements*’ aspect is often a product of analyses which become instantiated via formal standard operating procedures (SOPs). ‘*Devices*’ refers to the design of the physical equipment to support shared/team SA (and hence distributed cognition, more specifically cognitive off-loading). ‘*Mechanisms*’ are a product of training, ensuring a degree of common understanding of processes and procedures in the crew; effective distributed cognition is predicated upon such an underlying basis. However, over the last three decades a great deal of effort has been directed at the '*processes'* component: specifically CRM (Crew Resource Management).

CRM was developed to promote pilots acting in a well co-ordinated manner. This was a direct result of several accidents where aircraft had crashed as a result of poor team working rather than technical failures, the root of which was often inadequate communication and/or cross-checking of crew actions (i.e. a failure to utilise all the human resources available on the flight deck in an appropriate manner). In several accidents these was also a failure to use the automation in an appropriate manner, thereby freeing up the crew's cognitive resources o undertake better management of the developing incident.

As early as 1998 the European Joint Airworthiness Authorities (JAA) inadvertently defined CRM in the terms of the processes in a JCS promoting distributed cognition: CRM was '*the effective utilization of all resources (e.g. crewmembers, aeroplane systems and supporting facilities) to achieve safe and efficient operation*'. UK CAA Civil Aviation Publication 737 [7] suggested that a CRM syllabus for flight crew should comprise:

- Human error and reliability, error chain, error prevention and detection.
- Company safety culture, SOPs, organizational factors.
- Stress, stress management, fatigue and vigilance.
- Information acquisition and processing, Situation Awareness and workload management.
- Decision making.
- Communication and co-ordination inside and outside the cockpit.
- Leadership and team behavior synergy.
- Automation, philosophy of the use of automation.

The above list it is not just about the human flight crew members in the aircraft; coordination and communication outside the aircraft is also considered and the use of the automation is also specifically included. The trend in flight deck design has been one of progressive 'de-crewing' coupled with increasing levels of computerisation and system integration [8]. Now just two pilots, with much increased levels of on-board automated assistance and surveillance from the ground, undertake the same job once accomplished by twice this number. However the introduction of automation did not just replace members of flight crew; it changed the nature of the piloting task. The emphasis is now upon being a flight deck manager rather than a 'flyer'. The aircraft and its systems are now more usually under supervisory control rather than manual control. The key skills required are crew and automation management rather than minute-to-minute navigation, communication and flight path control.

Kanki and Palmer [9] listed five methods by which communication facilitates CRM performance. They could equally have described its function as five means by which it promotes distributed cognition. Communication:

- Provides information.
- Establishes relationships.
- Establishes predictable behavior patterns.

- Maintains attention to task and monitoring.
- Is a management tool.

2.2 Cognitive Off-Loading

At the simplest level possible, a pencil and paper improves human memory, either in the long-term (e.g. as in a diary) or in the short term (e.g. when noting intermediate steps when doing long division). However, by doing long division using a pencil and paper, the main information processing limitations are now not the storage capacity and characteristics of Working Memory but the accuracy of recalling and executing the required arithmetical procedures (from Long Term Memory) for doing such a calculation, and the symbolic representation of the digits. The long-division process is now distributed between a human and a non-human component. The artifacts being used serve as a 'cognitive amplifier' [10] enhancing human abilities by distributing the tasks between the artifacts and the user. However, the use of such external aids simultaneously changes the nature of cognition. A different skill set is now required to develop the skills and knowledge allowing exploitation of the artifact(s) enhancing the user's cognitive system. In this instance a new skill set is required to fly the current fourth and fifth generation, highly automated airliners compared to earlier second and third generation commercial jet aircraft from only three decades years ago.

Modern technology can also transform data on behalf of the human, a process that would previously be done either in Working Memory directly or via a series of processes and calculations using rudimentary technology. To help characterize these transformational processes Ackoff [11] suggested the following categorization:

- *Data* – Basic building blocks/symbols.
- *Information* – Data that have been combined and processed concerning questions such as 'who', 'what', 'where' and 'when'.
- *Knowledge* – This applies information to questions concerning 'how'?

As an example, certification requirements mandate the display of fuel flow and quantity (*data*), however, these parameters are of limited utility: what is required is *information* concerning what the remaining amount of fuel represents in terms of range or endurance and *knowledge* about how the subsequent management of the flight. Modern flight deck automation can transform raw *data* from sensors to supply *information* (and even *knowledge*) via the displays to the pilots. The production of *information* for the pilots to use is now off-loaded to the machine. First and second generation jet airliners supplied only *data* requiring a mental manipulations on the part of the pilots to convert it to the required *information/knowledge* and thereby also increasing the mental workload and error potential.

Hutchins [12] illustrated the manner in which the cognitive representation of speed and the processes for calculating target speeds were distributed across human and machine agents on a flight deck. The agents used included pilots' LTM (Long Term Memory); speed bugs on altimeters; speed reference cards and flight deck/ATM (Air Traffic Management) procedures. Speed calculation and speed awareness was not simply a problem in Working Memory. It was regarded as a problem in distributed cognition across the flight deck which ultimately resulted in a more resilient system: air speed representation and calculation was best understood from a system-wide perspective.

3 Shared Cognition and Cognitive Off Loading in the wider Joint Cognitive System

The operation of complex systems by a team has cognitive properties over and above those accounted for by individual cognition. The example in table 1 describes a situation where the aircraft's weather radar detects a cloud formation ahead (heavy precipitation and electrical activity) and effectively warns that that it may pose a risk to flight. Noticing the information on the display the First Officer commences in-flight re-planning activities but even before the First Officer communicates his concerns to the Captain, the latter notices the First Officer's activities and becomes aware that a change of course will be required. In this case, output from the weather radar is input for the First Officer. The First Officer's unexpected activity is the Captain's input. However, it needs to be noted that (a) information is dispersed across human and non-human components in the system and (b) there is implicit communication rather than a detailed exchange of mental models.

Table 1. Distributed Situation Awareness on the flight deck as part of a response to potential bad weather ahead [13]

Agent	Perception	Comprehension	Projection
Weather Radar	Senses radar returns of storm clouds	Compiles picture of extent of cloud formation, distance and bearing	Displays information (along with projected tack) in appropriate color to alert pilots
First Officer	Sees storm cloud formation on weather radar/navigation display	Determines thunderstorm may present a risk to the aircraft	Needs to quickly determine new route to avoid the storm
Captain	Sees First Officer interrogate Navigation Display, Flight Management System and charts	Determines thunderheads present a risk to passengers and crew	Re-plans flight and initiates a diversion

Stanton, Baber, Walker, Salmon and Green [14] proposed a set of basic tenets that may form the basis for supporting distributed cognition in a JCS.

- Data, information and/or knowledge can be held by both human and non-human elements in a system.
- There are multiple views of any circumstance held by all the different agents, human or machine.
- Non-overlapping and overlapping knowledge depends on the human or machine agent's goals, which may be different but still compatible.
- One component in the system (human or machine) can compensate for degradation in another.
- Communication between agents in the system may take many forms including non-verbal behavior or ingrained customs and practices.

4 Promoting Distributed Cognition on the Flight Deck

Distributed cognition is not a product: it is also a process. Task goals (and hence the automation functions activated and the parameters within which they work) are set and managed by human operators. There is a reciprocal relationship between schemata (knowledge about the environment held by either person or machine) which directs exploration of that environment (by either person or machine) and which gathers information from the environment, which then in turn modifies the schemata held by the wider joint cognitive system (q.v. Neisser's perceptual cycle [15]). The wider human-machine system awareness determines what is attended to which subsequently dictates how data (or information) is perceived, interpreted and what further information is subsequently actively sought out. Hollnagel [3] described these components as an 'inner view' (knowledge in the head) comprised of issues such as workload, attention, SA, decision making, etc. and an 'outer view' comprising the job context, system boundaries, nature of the task, responsibility and control, etc.

Dekker [16] suggested that he question for successful automation should not be 'who has control' but 'how do we get along together' (p. 194). Automation needs to be transparent [17] if it is to be trusted and managed effectively. A 'good team player' makes their activities observable for their fellows and they are easy to direct. These are all issues in promoting distributed cognition, either from a shared cognition or a cognitive off-loading perspective. Machines have to be managed in a similar manner to that by which people are managed. Machines have certain levels of responsibility delegated to them: sometimes this is a 'boss/slave' relationship ('I say – you do'); other times it is a more collaborative relationship where responsibility is assigned to the computer within certain parameters. To exceed these parameters requires assent from the manager (choice of options or confirmation of proposed course of action). However, automation can also be 'strong and silent' apparently pursuing its own course of action with little oversight of, or communication to its manager (obviating distributed cognition).

To design for effective distributed cognition in a JCS is difficult. There is almost a fundamental contradiction in that distributed cognition adopts a more holistic analytical approach, however the design process requires reduction of larger components to smaller components and the formal specification of how they interact.

To start off with an overall representation of the system is required which includes (as a minimum, and in no particular order):

- Description of the system boundaries (which may not necessarily be fixed)
- System objectives
- Potential system states
- The relationships (required transformations) between data/information/knowledge
- Control requirements
- Display and communication requirements
- A philosophy for the role of the human(s) in the system.

The core information requirements that all components in the system (human and non-human) need to be aware of must be identified for each phase of operation (task). Peripheral items that may be allocated to an individual or machine artifact also need to be identified, as does that mechanism for promoting awareness of these issues when necessary (communication). Machines do not have to be 'transparent' in their operation concerning 'how' they are doing something but the human components in the system do need to be aware of 'what' they are doing. However, all the above is predicated on a shared mental model possessed by the humans in the system (the pilots). This is a product of training, and only the humans in the system can be trained. Furthermore, only humans can define the objectives of a system, either in terms of its design or its operation. A machine can only perform what a human can imagine.

References

1. Harris, D., Stanton, N.A.: Aviation as a System of Systems. Ergonomics 53(2010), 145–148 (2010)
2. McDonald, N.: Modelling the Human Role in Operational Systems. In: Martorell, S., Guedes Soares, C., Barnett, J. (eds.) Safety, Reliability and Risk Analysis: Theory, Methods and Applications. CRC Press, London (2008)
3. Hollnagel, E.: Flight Decks and Free Flight: Where are the System Boundaries? Appl. Ergon. 38, 409–416 (2007)
4. Salomon, G.: No Distribution without Individual's Cognition: A dynamic Interactional View. In: Salomon, G. (ed.) Distributed Cognitions: Psychological and Educational Considerations, pp. 111–139. Cambridge University Press, Cambridge (1993)
5. Rogers, Y.: A Brief Introduction to Distributed Cognition, `http://mcs.open.ac.uk/yr258/papers/dcog/dcog-brief-intro.pdf` (retrieved December 14, 2009)
6. Endsley, M.R., Jones, W.M.: A Model of Inter and Intra-Team Situation Awareness: Implications for Design, Training and Measurement. In: McNeese, M., Salas, E., Endsley, M.R. (eds.) New Trends in Cooperative Activities: Understanding System Dynamics in Complex Environments, pp. 46–67. Human Factors and Ergonomics Society, Santa Monica (2001)
7. Civil Aviation Authority. Crew Resource Management (CRM) Training: Guidance For Flight Crew, CRM Instructors (CRMIS) and CRM Instructor-Examiners (CRMIES) (CAP 737). Civil Aviation Authority, London (2006)

8. Applegate, J.D., Graeber, R.C.: Integrated Safety Systems Design and Human Factors Considerations for Jet Transport Aeroplanes. Hum. Fac. Aero. Saf. 1, 201–221 (2010)
9. Kanki, B.G., Palmer, M.T.: In: Wiener, E.L., Kanki, B.G., Helmreich, R.L. (eds.) Cockpit Resource Management, pp. 99–136. Academic Press, San Diego (1993)
10. Pea, R.D.: Beyond amplification: Using the computer to reorganise mental functioning. Edu. Psychol. 20, 167–182 (1985)
11. Ackoff, R.L.: From Data to Wisdom. J. Appl. Syst. Anal. 16, 3–9 (1989)
12. Hutchins, E.: How a Cockpit Remembers its Speeds. Cog. Sci. 19, 265–288 (1995)
13. Harris, D.: Human Performance on the Flight Deck. Ashgate, Aldershot (2011)
14. Stanton, N.A., Baber, C., Walker, G.H., Salmon, P., Green, D.: Toward a Theory of Agent-Based Systemic Situational Awareness. In: Vincenzi, D.A., Mouloua, M., Hancock, P.A. (eds.) Proceedings of the Second Human Performance, Situation Awareness and Automation Conference (HPSAAII), Daytona Beach, FL, March 22-25 (2004)
15. Neisser, U.: Cognition and Reality: Principles and Implications of Cognitive Psychology. W.H. Freeman, San Francisco (1976)
16. Dekker, S.W.A.: On the Other Side of a Promise: What Should we Automate Today? In: Harris, D. (ed.) Human Factors for Civil Flight Deck Design, pp. 183–198. Ashgate, Aldershot (2004)
17. Wood, S.J.: Flight Crew Reliance on Automation. (CAA Paper No 2004/10). Civil Aviation Authority, London (2004)

The Use of Eye Tracking in the Study of Airline Cabin Safety Communication

Yueh-Ling Hsu[1,*], Wen-Chin Li[2], and Ching-Hui Tang[3]

[1] Department of Air Transportation, Kainan University, Taiwan
[2] Psychology Department, National Defense University, Taiwan
[3] Department of Transportation Science, National Taiwan Ocean University, Taiwan
irishsu@mail.knu.edu.tw, fs049504@yahoo.com.tw,
chtang@mail.ntou.edu.tw

Abstract . The purpose of this study is intended to address the current state of comprehensibility of airline safety briefing cards by adopting the eye-tracking experimental method and comprehension test to solve the relationship between comprehensibility and fixations of airline safety briefing cards. The Land Evacuation Section of a safety card was selected to measure respondents' eye movements together with a survey to test the comprehension of pictorials/pictograms. 51 subjects participated with this study. The results indicate that the universal situation that safety information is not well transferred to passengers and potential passengers. The result of study also showe that the pictograms related to "how to operate emergency door" took the longest fixation time and fixation counts, yet with the highest comprehension score. Meanwhile, other pictorials also showed the positive correlation between their comprehensibility and fixation time and fixation counts. The implications from these results were discussed. It is hoped that the present work will generate interest in the designer and user for providing guidance in the development of cabin safety briefing card.

Keywords: Cabin Safety Communication, Safety Briefing Card, Briefing card comprehension, Eye Tracking.

1 Introduction

It is generally agreed by aviation authorities that an alert, informed, knowledgeable person has a much better chance of surviving any life or injury-threatening situation which could occur during passenger-carrying operations in civil aviation [1-2]. Therefore, to minimize aircraft occupants' risks, the regulations require airlines to provide safety communication, i.e. safety briefings to inform passengers of routine and emergency safety procedures on board transport airplanes in accordance with the applicable standards before take-off. The briefing should include restriction on smoking, the

[*] Corresponding author.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 134–143, 2013.
© Springer-Verlag Berlin Heidelberg 2013

location of the emergency exits, the use of safety belts and when to use them, and the location and use of any required flotation means. In addition, each passenger-carrying operator must provide a safety card in a convenient location for use by each passenger. This card should supplement the oral briefing and should contain diagrams of, and methods of operating, the emergency exits, and other instructions necessary for the use of emergency equipment [3].

Research of cabin safety communication has established the importance of the provision of passenger safety information, and the importance of passenger attention being paid to such communications, such as NTSB's study showed that passengers who did not read the safety card suffered 3 times the number of injuries as passengers who did read the card [4]. However, many of the survivors indicated that their evacuation was not influenced by the passenger safety information which had been presented [5]. NTSB pointed out that many air carrier safety briefing cards do not clearly communicate safety information to passengers, because the majority of the experiment subjects failed to understand the meaning of most of the images presented on the cards [6]. ATSB's report also reveal the majority of passengers reported not having read the safety card, results suggest that the safety card is generally ineffective as a means of delivering safety information [7].

Safety information on briefing cards is typically presented graphically, using symbols, pictorials, and pictograms, while some cards employ text as well. Few researchers have done research regarding pictorials comprehension test and communication design of passenger briefing card. Fennell and Muir [8] found that the depictions of multiple brace positions on the card were confusing and forty-six percent of participants had trouble locating and removing the life vests from the packages. Over fifty percent of participants were wrong or unable to provide the information about the oxygen mask usage. Silver and Perlotto [9] stated that pictorials such as "stow away tables, no smoking in aisles, exit in a sitting position, brace against seat in front of you, and move away from the aircraft" receive less comprehension rate. Caird, Wheat, McIntosh, and Dewar [10] found that comprehension for most of the pictorial information was uniformly low, and specific comprehension difficulties are related to emergency floor lighting and do not use electronics during take-off and landing. Corbett, McLean and Cosper [11] found comprehension of pictorials/pictograms was related to the familiarity that cabin safety professionals and high flight-time passengers have with safety briefings and briefing cards, and they concluded that safety briefing card pictorials/pictograms need to be designed and implemented with respect to novice passengers who do not have a prepotent understanding of the design, emergency equipment, and/or aircraft emergency procedures. Hsu[12] conducted a A330 safety card comprehension test by using Taiwanese respondents and reconfirms the universal situation that safety information is not well transferred to passengers and potential passengers. The improvement area of safety card comprehension should focus on the pictorials/pictograms related to Evacuation (both water ditching& land), Brace Position, Life vest used in the water/night, and Fasten seat belt sign. Hsu and Hsu [13] also tested and compared the comprehension of pictorials/pictograms on Boeing 747-400 safety cards of two national Taiwanese carriers. The findings show that both carriers' safety information are not well transferred to passengers and potential passengers, especially the pictorial/pictograms related to Land and Water Evacuation, Fasten-seat-belt sign compliance and Brace Position. Horizontal layout of pictogram is likely to have higher comprehension level compared to vertical layout for Taiwanese participants. It is also suggested to add textual clarification and serial numbers to serial pictograms so that safety information will be easier to understand.

In the last few years, the interest of cabin safety communication brought the emergence of studies on the comprehension level of safety briefing card by using comprehension surveys. Nevertheless, with the disadvantages and limits of survey technique, employing diverse methods of comprehension evaluation on the safety card is needed. Eye-tracking technique is a popular methodological approach in the area of HCI. Tracking people's eye movements can provide an insight into problem solving, reasoning, mental imagery, and search strategies which help researchers understand visual and display-based information processing and the factors that may impact upon the usability of system interfaces[14]. Because eye movements provide a window onto so many aspects of cognition, there are also rich opportunities for the application of eye-movement analysis as a usability research tool in HCI and related disciplines such as human factors and cognitive ergonomics s[14].

The present study is therefore intended to address the current state of airline safety briefing cards by adopting the briefing card comprehension test together with eye-tracking experimental method to gather behavioral data and explore the relationship between comprehensibility and fixations. The findings enable to generate some suggestions to improve usability in airline safety briefing card.

2 Method

2.1 Participants

Given no previous study ever conducted for safety card comprehension test by using eye-tracking technique, this research is served as a preliminary study. To this end, age variable was controlled. The sample participants are required to be aged between 18-30 years old, as safety card pictorials and pictograms need to be designed and implemented for those passengers who do not have a prepotent understanding of transport aircraft related knowledge. Meanwhile there is a growing trend of young travelers, who have better learning ability and not so many travel experiences. Sixty-eight participants volunteered for the study. All participants were undergraduates and postgraduates, and ranged in age from 18-30 years old. Due to the invalid data recorded by the eye-tracking system, there are fifty-one valid data for this research.

2.2 Stimulus Materials

According to Hsu [12] and Hsu and Hsu[13], one of the specific comprehension difficulties are associated with the pictograms of Emergency Evacuation, which should be the first improvement area. The pictograms related to Land Evacuation on the

A330 safety card of a Taiwanese carrier have therefore been selected as the stimulus material in this research. The popularity of A330 in worldwide airline industry will facilitate further comparison research too.

2.3 Procedures and Data Analysis

The eye-tracking test was conducted with ASL Mobile Eye wearable eye tracking system, which contains eye camera, scene camera, and PC with monitor. After each participant had worn the eyeglasses and made the best adjustment, the section of Land Evacuation on the safety card was shown to participants. The participants can take as long as they like to read the diagrams, just like passengers on board the airplane. The data gathered from the eye-tracking system was classified into five categories, i.e Area of Interests (AOI) in the stimulus material (see Figure 1), and then analyzed by a software named "Gaze Map".

Once the participants completed the reading, they were asked to conduct the comprehension test of these diagrams. The comprehension survey was adopted from the study of Hsu [12], which is an open-end questionnaire. Instructions were given to ask the participants to describe the meaning of the pictorials/pictograms. There was no time limit to complete the survey either. The comprehension test result was then scored according to the method used by Corbett, McLean and Cosper [11].

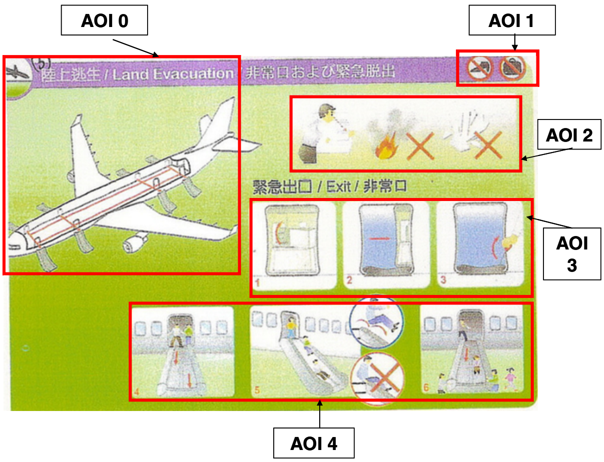


Fig. 1. Five Area of Interests (AOI) in the stimulus material

3 Results and Discussions

The respondents' demographic data are presented as follows (see Table 1):

Table 1. Respondents' Demographic Data

		Number	Percentage
Gender	Female	14	27.45
	Male	37	72.55
Age	21-30	36	70.59
	under 20	15	29.41
Flying Frequency	7-8 time	1	1.96
	5-6	2	3.92
	3-4	3	5.88
	1-2	20	39.22
	0	25	49.02
Flying Experience	Yes	42	82.35
	No	9	17.65
Purpose	VFR	1	1.96
	Study	2	3.92
	Travel	35	68.63
	Business	2	3.92
Attribute	Group	30	58.82
	Individual	10	19.60
Missing	System	11	21.57

3.1 Comprehension Scores

A matrix of the comprehension scores of each AOI is displayed in Table 2. Comprehension scores based on the item for each AOI ranged from 17.16 % to 66.67 %, with a mean score of 40.19%. No items met the more stringent American National Standards Institute (ANSI) suggested comprehension standard of 85%. One of AOI pictorial/pictogram (AOI3) almost met the International Organization of Standards (ISO) suggested comprehension criterion of 67%. Others were understood accurately below the 50% level (see Table 2).

In order to see whether the five AOIs differed across demographics, T test were conducted for each item across gender, age, flying experience and passenger attribute. One-way Analysis of Variance (ANOVA) was conducted across flying frequency, and purpose of the flight. When the effects are significant, the means must be then examined in order to determine the nature of the effects, by employing the post hoc multiple comparison- LSD. The results indicate that:

Table 2. Comprehension score of AOI 1-5

Item		Comprehension Scores (%)
AOI 3	How to open the door	66.67
AOI 4	How to use slide	63.73
AOI 2	Watch outside circumstances before open the door	27.94
AOI 0	Emergency exit and evacuation route	25.49
AOI 1	No baggage and high heel when evacuation	17.16
Average		40.19

(1) **Gender-** the significant differences between male and female were found on **AOI 0 -Emergency Exit and evacuation route**, (T=2.52*, male>female).

(2) **Flying experience**- the significant differences between yes and no were found on **AOI 3- How to open the door** (T=-2.21*, no>yes), AOI 4- How to use slide (T=-2.20*, no>yes).

(3) **Flying frequency**- The item with significant difference is **AOI 1- No baggage and high heel when evacuation** ($F=3.88$, $p<0.05$). LSD indicated that the variability may come from the difference between 3-4 times and 1-2 times (3-4times> 1-2 times).

3.2 Eye Movement Analysis

Table 3 shows the results of Eye Movement Index for AOI 1-AOI 5, including "Time in region", Number of Fixations", "Total Fixation Duration" , and "Time to First Fixation".

Table 3. Eye Movement Index and Data

Eye Movement Index	AOI 0	AOI 1	AOI 2	AOI 3	AOI 4
Time in Region(sec)	6.65	0.34	6.28	14.39	8.79
Total Fixation Duration(sec)	4.2	0.19	3.81	9.18	5.54
Number of Fixations(frequency)	26.94	1.31	23.94	57.13	35.21
Time to First Fixation(sec)	2.99	27.58	7.3	9.62	19.83

As shown in Figure 2, in terms of "Time in Region" and "Total Fixation Duration", **AOI 3 (How to open the door)** has the highest ranking, then AOI 4(How to use slide)> AOI 0 (Emergency Exit and evacuation route)>AOI 2 (Watch outside circumstances before open the door) >AOI 1 (No baggage and high heel when evacuation). For the index of "Number of Fixations", the results remain the same (see Figure 3).

AOI 3 (How to open the door) takes more time to read and renders the highest Number of Fixations. It shows that there are more messages/details in this area, or it is more difficult to read within the context, which cause the readers desire or need to spend more time in dealing with the message in this area. In particular, according to T test, the significant differences of flying experience between yes and no were found on Time in Region(T=-2.04*,no>yes), Total Fixation Duration(T=-2.26*, no>yes) and Number of Fixations(T=-2.20*, no>yes). In other word, those who have no flying experiences may need more observation time to deal with messages contained in AOI 3 (How to open the door).

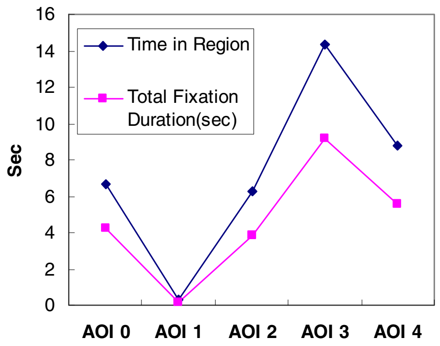


Fig. 2. Time in Region and Total Fixation Duration of AOI 1-AOI 5

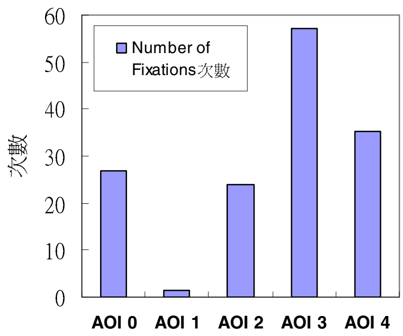
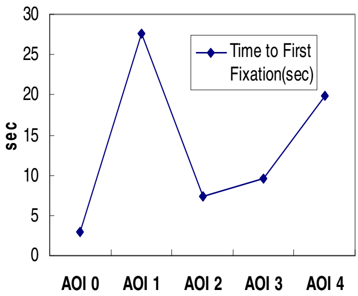


Fig. 3. Number of Fixations of AOI 1-AOI 5

In terms of the index of "Time to First Fixation", AOI 0 (Emergency Exit and evacuation route) rendered the shortest time, then the order is AOI 2 (Watch outside circumstances before open the door)< AOI 3(How to open the door) < AOI 4(How to use slide) <AOI 1 (No baggage and high heel when evacuation)(see Figure 4). As "Time to First Fixation" means the time it takes for a participant to first fixate on a specific area or interest. It can show people across the board which AOI was drawing respondents' attention in the context of the task they are asked to perform. Therefore, the result implies that AOI 0 (Emergency Exit and evacuation route) attracts the respondents the most, and the path of reading interest of AOIs appears like Figure 5.

The results reflect that most languages are read left-to-right, and top-to-down. Also, compared to other AOIs, the diagrams in AOI 1 (No baggage and high heel when evacuation) are much smaller, which obviously the least attractive items for the respondents.



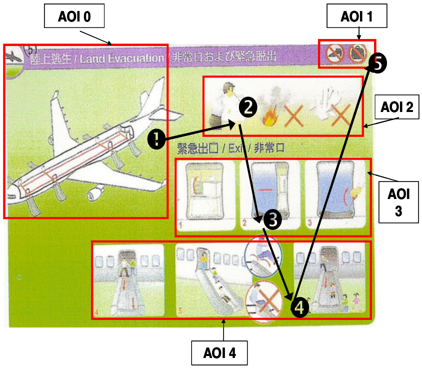


Fig. 4. Time to First Fixation of AOI 1-AOI

Fig. 5. The path of reading interest of AOI1-AOI 5

3.3 Eye Movement versus Comprehensibility

To explore the relationship between eye movement index and comprehension score of AOIs, correlation analysis was conducted. Table 4 shows the significant correlations. The comprehension score of AOI 1 (No baggage and high heel when evacuation) has significant low and positive correlation with Time in Region, Number of Fixation, as well as Fixation Duration, and low plus negative correlation with Time to First Fixation. The results show that for AOI 1(No baggage and high heel when evacuation), the shorter of observation length and longer Time to First Fixation are related to the least comprehensibility. For AOI 2 (Watch outside circumstances before open the door), its comprehension score has significant low and negative correlation with Time

Table 4. The correlation between eye movement index and comprehension score of AOIs

	Comprehensibility		
	AOI 1	**AOI 2**	**AOI 4**
AOI 1_time in region	0.30**		
AOI 1_number of fixation	0.28**		
AOI 1_fixation duration	0.28**		
AOI 1_time to first fixation	-0.06*		
AOI 2_time to first fixation		-0.28*	
AOI 4_time in region			0.38**
AOI 4_number of fixation			0.39**
AOI 4_fixation duration			0.39**

* Correlation is significant at the 0.05 level (2-tailed).

** Correlation is significant at the 0.01 level (2-tailed).

to First Fixation. The result shows that the less interest attention (longer Time to First Fixation) are associated with the lower comprehensibility. Meanwhile, the comprehension score of AOI 4(How to use slide) has significant low and positive correlation with Time in Region, Number of Fixation, and Fixation Duration. It reveals the longer of observation length are in association with better comprehensibility.

4 Conclusions and Suggestions

The purpose of this research was to investigate the efficacy and usability of airline safety cards through a comprehensibility test together with an eye-tracking study. Tracking people's eye movements provide an insight into mental imagery and display-based information processing of airline safety card, which enhanced the current comprehensibility test. Findings of this research led to following conclusions for the improvement of usability in the safety card:

1. Within the context of Emergency Evacuation, the diagrams relating to "How to open the door" take most time for respondents to comprehend but render in the highest comprehension scores. Also flying experiences exert an influence on the comprehensibility and cognition processing time in terms of "How to open the door". So it is crucial to make passengers spend time reading the safety card to increase the comprehensibility.
2. Diagrams relating to "No baggage and high heel when evacuation" should be as larger as better, because this is the part ought to be fulfilled by the passengers when emergency evacuation. It is better to move the diagrams to left hand side for first priority to read.
3. Diagrams relating to "Watch outside circumstances" render the most problematic answers, it is suggested to add some textual clarification to make safety information more meaningful.

Acknowledgement. This research is supported by the National Science Council of Taiwan, R.O.C., under Grant no. NSC 99-2221-E-424 -002

References

1. FAA, Passenger Safety Information Briefing and Briefing Cards, AC121-24C (2003)
2. Australian Civil Aviation Safety Authority (CASA), CAAP 253-2(0), Passenger safety information: Guidelines on content and standard of safety information to be provided to passengers by aircraft operators (2004)
3. Flight Safety Foundation (FSF), What more can be done? Cabin Crew Safety 26(3), 1–14 (1991)
4. National Transportation Safety Board (NTSB), Special Study-"Safety Aspects of Emergency Evacuations from Air Carr November 13, NTSB-AAS-74-3 (1974)
5. National Transportation Safety Board (NTSB), Safety Recommendation (5), Report no: A-83-45 (1983)

6. National Transportation Safety Board (NTSB), Safety Study: Emergency Evacuation of Commercial Airplanes. NTSB/SS-00/01 PB2000-917002 (2000)
7. Australian Transport Safety Board (ATSB), Public Attitudes, Perceptions and Behaviors towards Cabin Safety Communication, B2004/0238 (2006)
8. Fennell, P.J., Muir, H.C.: Passenger Attitudes Towards Airline Safety Information and Comprehension of Safety Briefings and Cards. CAA Paper 92015 (1992) (Civil Aviation Authority, London). quoted in National Transportation Safety Board (NTSB), Survivability of Accidents Involving Part 121 U.S. Air Carrier Operations, 1983 Through 2000, NTSB Safety Report NTSB/SR-01/01 (March 2001)
9. Silver, N.C., Perlotto, C.N.: Comprehension of aviation safety pictograms: Gender and prior safety card reading influences. In: Proceedings of the Human Factors and Ergonomics Society 41st Annual Meeting, pp. 806–810 (1997)
10. Caird, J.K., Wheat, B., McIntosh, K.R., Dewar, R.E.: The Comprehensibility of Airline Safety Card Pictorials. In: Proceedings of the Human Factor and Ergonomics Society, pp. 801–805 (October 1997)
11. Corbett, C., Mclean, G., Cosper, D.: Effective Presentation Media for Passenger Safety I: Comprehension of Briefing Card Pictorials and Pictograms. Federal Aviation Administration, DOT/FAA/AM-08/20 (2008)
12. Hsu, Y.L.: A Study on the Comprehensibility of Airline Safety Card Pictorials and Pictograms. Journal of Aeronautics, Astronautics and Aviation 42(1), 31–38 (2010)
13. Hsu, Y.L., Hsu, C.C.: The effectiveness of cabin safety communication - an example of Boeing 747-400 Safety Briefing Card. In: The 15th Annual Conference of the Air Transport Research Society (ATRS), Sydney, Australia, June 29-July 2 (2011)
14. Poole, A., Ball, L.J.: Eye Tracking in Human-Computer Interaction and Usability Research: Current Status and Future Prospects. In: Ghaoui, C. (ed.) Encyclopedia of Human Computer Interaction. Idea Group, USA (2005)

Human Factors Modeling Schemes for Pilot-Aircraft System: A Complex System Approach*

Dan Huang and Shan Fu

School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China
{huangdan,sfu}@sjtu.edu.cn

Abstract. The human factor is becoming a main topic in modeling and simulation, especially in airline safety as more aviation accidents are classified as pilot (human) errors. Traditional modeling schemes treat pilots and aircraft individually, assuming the other as given. However, to define a system-level architecture for the safety analysis, it is advantageous to expand the system boundary to include both pilots and aircraft as a coupled entity. In this paper, we propose a framework for pilot-aircraft system modeling scheme from a complex systems point of view. Key pilot factors are identified and quantified, and complex relationships and interaction among these factors are incorporated into usually well-modeled aircraft system. We also introduce a fast-time simulation model of man-aircraft-environment complex system with the human strategy model as its core to generate large sample sizes of flight data for this modeling purpose. The given results not only provide a proactive approach to the research of flying safety, it can also be applied to other complex dynamic systems.

Keywords: human factors, complex dynamic systems, coupled modeling, distributed information and control.

1 Introduction

With the increasing advancement and complexity of technological systems that operate in dynamically changing environments and require human supervision or a human operator, the relative share of human errors is increasing across all modern applications, especially in the air transportation industry. Although maintenance, manufacturing design flaws, and operational deficiencies other than those of the pilot are typically cited as cause factors, the major cause of all aviation accidents is pilot-error. Some studies suggest that approximately 70% of aviation accidents are classified as pilot-error[1], while others indicate that all accidents have some form of human error attached to their cause[2]. Table 1[3] shows the major airline crashes in the U.S. from 1989-1999. This may be unfair to pilots because accidents are often a result of a chain of events in which the pilot is the last link in the chain. But, often the pilot's

* This work was supported by the National Basic Research Program (973 Program) of China under Grant No. 2010CB734103 and the National Natural Science Foundation (No.61004026) of PR China.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 144–149, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Table 1. Major US airline crashes: 1989-1999

Date	Carrier	Location	Fatalities (On Board)	Cause category[a]
07/19/89	United	Sioux City, IA	111 (298)	Maintenance
01/25/90	Avianca	Cove Neck, NY	73 (158)	Pilot-error
02/01/91	Skywest	Los Angeles, CA	12 (12) + 22[b]	Miscellaneous, other
02/01/91	USAir	Los Angeles, CA	22 (89) + 12[b]	Miscellaneous, other
03/03/91	United	Colorado Springs, CO	25 (25)	Undetermined
03/22/92	USAir	La Guardia APT, NY	27 (51)	Pilot-error
07/02/94	USAir	Charlotte, NC	37 (57)	Pilot-error
09/08/94	USAir	Aliquippa, PA	132 (132)	Undetermined
10/31/94	American Eagle	Roselawn, IN	68 (68)	Pilot-error
05/11/96	ValuJet	Everglades, FL	110 (110)	Not given
07/17/96	TWA	Long Island, NY	230 (230)	Not given
01/09/97	Comair/Delta Connection	Monroe, MI	29 (29)	Not given

judgment in handing emergency situations is the final deciding factor as to whether the event will result in an accident. Reliable model of pilot-aircraft is therefore essential for aviation safety.

Conventional approaches focus on analyzing four major classifications of factors that have traditionally associated with pilot-error, that is, environmental factors, aircraft factors, airline-specific factors, and pilot-specific factors. Researchers tend to concentrate on only one category when designing their safety study. Our aim is to develop a model incorporating two of them, that is, the pilot-aircraft system. The incorporation of the human factor into the aircraft models is very difficult. The behavior of the pilot is influenced by a big number of environmental, physiological, and other kind of parameters that are difficult to measure. Furthermore, other human factors, like pilot-induced delays, varied control precisions, and the capability limitations and constraints of human behaviors, can all have a significant effect on the flying qualities of the aircraft. Conventional human behavior models, though investigated broadly, could not be rich enough in modeling pilot-aircraft systems as such systems are complex in a sense that they exhibit new properties, not easily deducible or found from properties of their individual parts. Due to interactions of different parts in complex systems between themselves and with dynamically changing environment, coupled effects and phenomena are becoming increasingly important in studying such systems.

In this paper, we study how human factors can be effectively coupled into the modeling and control of the pilot-aircraft system from a complex system point of view. From analytical point of view, such systems interact with their environments, and due to a degree of uncertainty in this dynamic interaction, many associated problems require the development of approaches with a substantial reduction in a priori information required for their applications. Such situations are typical for most complex systems. In this case, rich data must be gathered for accurate approximation, considering all exogenous perturbations and environmental data. Therefore, a fast-time simulation model of man-aircraft-environment complex system with the human strategy model as its core is constructed to generate large sample sizes of flight data for this modeling purpose. It is worth mentioning that such data, which is from distributed information resources, though remain as a challenge for humans, can be leveraged for collaboration and control strategies for further research.

The presented coupled modeling and control schemes in this paper aim to reduce design cycle time, support quantitative predictions of human-system effectiveness, and improve the design of crew operating procedures. It is also noteworthy that our new approach, that takes into account the more complex relationships among relevant factors, not only provide a proactive approach to the research of flying qualities, it can also be applied to other complex dynamic systems.

2 Complex Systems Approach

Various informal descriptions of complex systems have been put forward, and these may give some insight into their properties. One of the defining features of complex systems is that a complex system is one in which there are multiple interactions between many different components. Therefore, complex systems may have certain coupled features, such as:

- Cascading Failures[4]

Due to the strong coupling between components in complex systems, a failure in one or more components can lead to cascading failures which may have catastrophic consequences on the functioning of the system.

- Dynamic network of multiplicity[5][6]

As well as coupling rules, the dynamic network of a complex system is important. Small-world or scale-free networks which have many local interactions and a smaller number of inter-area connections are often employed. Natural complex systems often exhibit such topologies. In the human cortex for example, we see dense local connectivity and a few very long axon projections between regions inside the cortex and to other brain regions.

- Relationships are non-linear

In practical terms, this means a small perturbation may cause a large effect (see butterfly effect), a proportional effect, or even no effect at all. In linear systems, effect is always directly proportional to cause. See nonlinearity.

It is not hard to tell that the pilot-aircraft system in nature is a complex system. Coupling or interaction is the defining feature of this system. Unfortunately, existing theories of such man-involved complex systems may never be reducible to analytical equations, but are more likely to be sets of conceptually simple mechanisms interconnected that produce context-specific dynamics and outcomes. A human strategy model is developed via critical coupling of multiple characteristic patterns to test various flight conditions and to generate flight data with wide representation. This coupled human-aircraft modeling method proposed in this paper provides an algorithmic/numerical approach to establish a unifying framework for decoding the internal dynamics of man-involved complex system and lead forward unifying algorithmic theories of the relations among system complexity, adaptive behavior, and parameter uncertainty.

3 Modeling Schemes

The human factors have been deeply studied by researchers from different disciplines, however most of these models are traditionally very qualitative. Obviously, integrating qualitative models in simulation models could introduce significant criticality in term of confidence band definition, tolerance and model validation. However, theories of man-involved complex systems are difficult if not impossible to be reducible to analytical equations. We here propose a modeling framework which is sets of conceptually simple mechanisms interconnected that produce context-specific dynamics and outcomes. Quantitative results can be obtained from extensive simulation experiment. In fact, to guarantee the reliability of the model, it is necessary to have a very large amount of data that is detailed and/or difficult to measure. Our approach enables such data generation with wide representation.

In order to model such complex system, the first step is to identify key pilot factors. Researches on human factors in aviation have been conducted, and some major aspects of human factors with significant impacts to flight safety are identified[9]. Human factors influencing flight safety include but not limited to:

1) Pilot's control time and control precision;
2) Visual scene of the cockpit and out-of-the-window area;
3) Pilot's workload and workload increment;
4) Pilot's fatigue and fatigue related error;
5) Pilot's prediction on parameter variation;
6) Pilot's situation awareness;
7) Pilot's fast recognition on unacceptable conditions;
8) Pilot's detection and isolation of fault transient.

Once we identify such components, to get a reliable simulation models, certain issues must be tackled[10], including the determination of the behavioural model complexity level required for the specific case, quantification of the feasibility of acquiring knowledge of the application case and of quantitatively characterizing, verification, validation and accreditation (VV&A) process, and evaluation of the benefits provided by these models with respect to traditional simulators based on average performance levels.

Ergonomic and human factors evaluation of flight deck is a complicated study of the interaction between human and machine and the factors that affect the interaction. Aircraft is to be interacted via interfaces like controls and switches, and feedback information is to be obtained via interfaces. The designs and arrangements of these items may have varying degrees of effects to human factors, hence they may have unknown impacts on aircraft performance and flight safety. Dynamic evaluation of cockpit human factors is the inevitable demand.

To solve the problem, or to generate very large sample sizes of flight data, what indispensable is a fast-time simulation model of pilot-aircraft complex system with considerable flexibility to configure aircraft properties, human capability, and flight environment. This framework aims to represent human performance in complex environments within computer-based discrete event simulation (DES) models using data drawn from empirical studies in the literature and from many military laboratory based studies. Meanwhile, the framework should account for physical environmental factors with system performance as well as the interaction and mediating effects of other variables. The development of a modeling tool that is able to the elaborate of the functional relationships, which will provide a means of hypothesizing the relationships that need to be sought in order to begin to form a modeling tool. Once such a fast-time simulation model is in place, various combinations of aircraft with different characteristics and performance, pilots with different skills and habits, and flight scenarios with nominal and off-nominal cases, can then be tested at fast rate under controllable lead time.

Data analysis on aircraft performance, handling quality, flight safety will not be discussed in depth here though they are ultimate aims to develop human models. Aircraft performance can be defined as a measure of the ability of the aircraft to carry out a specific task. Four aspects of aircraft performance should be included in future effort: performance estimation, performance measurement, performance expression, and performance implication on flight safety. Handling quality are characteristics of an aircraft that govern the ease and precision with which a pilot is able to perform the tasks. Objective and quantitative rather than subjective and qualitative rating of the man-aircraft system is made possible, since human model enables human out of the man-aircraft closed loop. Objective and quantitative ratings can be deduced from data available from very large sample sizes of flight data generated by massive running the fast-time simulation model of complex system; because such flight data not only contain flight status of the aircraft but also contain relevant efforts made by the pilots. Safety margins for intended flight operations can be defined by the space needed to manipulate the aircraft is within valid space and the aircraft's fuel is sufficient. Analysis of flight safety can be summarized as whether the aircraft's performance is scheduled according to the aircraft's handling quality, the aircraft's performance margin and the pilot's capability, to make sure whenever a flight task is accomplished the aircraft is within safety margin.

4 Discussions and Conclusions

The proposed modeling scheme in this paper is essentially data-based approach from complex systems point of view. While we can quantify the human factors to some extent, rigorous analytical is not presented. This is difficult but not impossible. In some cases, the analysis of such systems requires dealing with highly oscillatory functions acquired as a result of measurements. This leads to mathematical difficulties. Integration approach [8] has been proposed based on a set of Hamilton-Jacobi-Bellman equations. The difficulty of this method lies in Hamiltonian estimation which is based on the information accumulated up to the point to reflect dynamic changes in the environment in which the system operates. The solvability of these HJB equations can be developed using mixed finite-element methods or Perron's method. Analytical results along with mathematical formulations are given to verify the effectiveness of the proposed modeling methodologies.

The modeling framework proposed in this paper may have an impact in other field of research, on of which is nuclear power plant. Developing advanced control room in nuclear power plant is the most efficient way to improve the safety and reliability of nuclear power plants. Besides the introduction of advanced technology and equipment, coupling human factors and the control room has been considered in the development of advanced control room significantly. It improves the safety and reliability in nuclear power plant, because the operation is simplified and the operators' load is lightened. The other promising area of application is health care field. Medical research shows that about 2/3 of the complications suffered by hospital patients are due to errors in patient care[7]. Like the aviation industry, the health care field is struggling to develop and implement new approaches to error prevention.

References

1. BASE: Boeing Airplane Safety Engineering, Statistical Summary of Commercial Jet Airplane Accidents - Worldwide Operations, 1959 - 1996. Boeing Commercial Airplane Group, Seattle (1997)
2. Braithwaite, G.R., Caves, R.E., Faulkner, J.P.E.: Australian aviation safety - observations from the 'lucky' country. Journal of Air Transport Management 4(1), 55–62 (1998)
3. McFadden, K.L., Towell, E.R.: Aviation human factors: a framework for the new millennium. Journal of Air Transport Management 5, 177–184 (1998)
4. Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, S.: Catastrophic cascade of failures in interdependent networks. Nature 464, 08932 (2010)
5. Newman, M.: Networks: An Introduction. Oxford University Press (2010)
6. Cohen, R., Havlin, S.: Complex Networks: Structure, Robustness and Function. Cambridge University Press (2010)
7. Brennan, T.A., Leape, L.L., Laird, N.M., Hebert, L., Localio, A.R., Lawthers, A.G., Newhouse, J.P., Weiler, P.C., Hiatt, H.H.: Incidence of adverse events and negligence in hos-pitalized patients, results of the Harvard medical practice study I. The New England Journal of Medicine 324(6), 370–376 (1991)
8. Melnik, R.V.N.: Coupling control and human factors in mathematical models of complex systems. Engineering Applications of Artificial Intelligence 22, 351–362 (2009)
9. Byrne, M.D., Pew, R.W.: A History and Primer of Human Performance Modeling. Reviews of Human Factors and Ergonomics 5(1), 225–263 (2009)
10. Bruzzone, A.G., Briano, E., Bocca, E., Massei, M.: Evaluation of the impact of different human factor models on industrial and business processes. Simulation Modelling Practice and Theory 15, 199–218 (2007)

The Influence of Guanxi Gradient on Crew Resource Management and Values in the Cockpit

Hung-Sying Jing and Berlin Chen

Institute of Civil Aviation, National Cheng Kung University, Tainan, Taiwan, R.O. China
hsjing@mail.ncku.edu.tw

Abstract. The goal of this research is to reveal the influence of a newly developed concept of guanxi gradient on crew resource management and the corresponding values in a Chinese cockpit. Guanxi gradient is a cultural variable describing the decay of attitude considering different degree of interpersonal intimacy. A questionnaire measuring the attitude change is designed in the study. The questionnaire includes three parts which are basic perception, situational response and open questions. The objective of these questions is to find out how the operations of CRM were affected by the different degree of intimacy. Also, the questionnaire was designed to expose the corresponding underlying Chinese value system. It is found that harmony is the top value in cases without safety concerns.

Keywords: crew resources management, guanxi gradient, value system.

1 Introduction

Crew Resource Management (CRM) is already very popular throughout the world's commercial aviation community. No matter how it is perceived in different cultures, the core of CRM is synergy, the interaction of discrete agents to produce a total effect greater than the sum of the individual effects. There are only two crewmembers in the cockpit, the Captain and the First Officer. The Captain is supposed to give orders and the First Officer expected to follow. Consequently, in discussing the interaction between the crewmembers, a power gradient is always assumed. There is an imbalance of status in the cockpit. The core concern of interaction on the flight deck turns naturally to the power distance issues. One focus of CRM worldwide is thus often to lower the power gradient in the cockpit.

It is well known that Western culture is centred around individualism while Chinese culture centres on situationalism (Hsu, 1970). In the West, people accept that all humans are created equal and should be seen as individuals. For the Chinese, everyone is a member of a group in different situations, not just an individual. The stability of groups is based on unequal relationships between members within it, as in a family, the prototype of all social organizations. Hence, there are differences between Westerners and Chinese concerning relationships between people. One of the goals of CRM is to enhance the cooperation between the crewmembers. However, it is deeply individualistic in its basic philosophy. Therefore, for Westerners and people

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 150–156, 2013.
© Springer-Verlag Berlin Heidelberg 2013

with the same individualistic point of view, it is natural to find that power distance is important, as was found in the work of Helmreich and Merritt (1998), and it is possibly very difficult for them to see something other than the power distance in the cockpit.

From a Chinese point of view, something might be missing. The core of Confucius teaching is ‘ren’ and it is at the heart of Chinese culture (Liang, 1963). ‘Ren’ also explains why Chinese culture is ethical. It is generally agreed scientifically that there exist two cognitive dimensions to describe and categorize Chinese human relations (Chuan and Yang, 1997). One of them is superior-subordinate, a concept much greater in scope than leader-follower. There are also sub-categories within this, and each is attached with a certain morality. In the dimension of superior-subordinate, Chinese authoritarianism is described by the concept of dragonality, a concept also much larger than power distance. Dragonality was proposed by Jing et al. (2002) that to a certain extent this is useful to comprehend the Chinese counterpart of authoritarianism. However, there is the other dimension used by the Chinese to categorize interpersonal relations. It is the intimacy-estrangement. People use this to categorize people without an obvious superior-subordinate relation, and there are certain moralities associated with it. Differentiated order (Fei, 1947) is the well-recognized concept established to delineate the Chinese perception about intimacy-estrangement.

Based on these previous studies, Jing (2004), using empirical evidence, proposed a four-level differentiated order to describe the basic structure of Chinese human relations along the dimension of intimacy-estrangement: kin/acquaintance/fellow/alien. With self at the centre, ‘kin’ refers to the people located closest to self and related by blood or marriage. ‘Acquaintance’ represents those people quite familiar to us, up to a status that we can have sentimental and beneficial exchange with them, like old friends, neighbours and close colleagues. People not very familiar or even not known to each other, but who still have certain things in common, for example working in the same company, graduating from the same university or joining the same party, belong to the category of ‘fellow’. As for ‘alien’, this represents the people we do not want to work with or even to understand – those who are psychologically remote. Although each level represents a specific category of intimacy or trust between people, it is not a fixed structure and the boundaries between these levels are movable and permeable. As a result, the four-level differentiated order does provide a preliminary frame of understanding of the complicated Chinese concept of guanxi.

2 Guanxi Gradient

Guanxi gradient is believed by the authors to be a cultural variable quite useful for understanding interpersonal relations worldwide. Based on the previous study about the four-level differentiated order, it can be seen that people differentiate their relations with others and a gradient type of attitude does exist according to the different degree of intimacy. It is conceived to be a basic part of human nature having a type of gradient in attitude when dealing with people located at different psychological distances. In addition, this gradient must be negative, i.e., there is a positive attitude to closer people, such as family members, and a less positive, even negative attitude, to less close people, such as strangers. The gradient here represents the rate of change of attitude with different degrees of intimacy (Jing, 2007).

Guanxi gradient is defined as the slope of the declining attitude towards different levels of intimacy concerning interpersonal relations. It is found that for Western people, the gradient is always flatter compared to their Eastern counterpart, especially the Chinese. This can be considered as a strong indication of individualism. When compared with Hofstede's individualism index (1980), it is found to be in fairly good agreement, with the results showing that North America and Australia possessing the highest scores amongst the other nations for individualism have the lowest gradient.

Guanxi gradient is a concept larger than individualism and collectivism together. Individualism, as generally accepted, represents a concept stressing human independence and the importance of individual self-reliance and liberty. Compared to guanxi gradient, individualism represents one end of the spectrum and in this case differentiation only takes place between 'self' and 'others'. The Chinese, on the other side of the spectrum, further differentiate the 'others' and use different attitudes to deal with different 'others'.

In Hofstede's work, and also as generally perceived, collectivism is a concept opposite to individualism. Therefore, for the countries with a low individualism score such as China, one would naturally assume that they must have a high degree of collectivism. However, this point of view is already denied by many Chinese scholars (Yu, 1984, Fei, 1947, Hsu, 1970, Yang, 1993). It can only be said that China is a non-individualistic society instead of collectivistic.

However, since individualism-collectivism is a very rough dichotomy, non-individualism is often easily confused with collectivism. For a country like China, interpersonal relations are surely non-individualistic but are also definitely not collectivistic. As for a society like China, people further differentiate 'others' into several categories (Jing, 2004), and some 'others' are treated enormously different from other 'others'. Hence, the concept of collectivism is certainly not enough and sometimes even misleading. The concept of guanxi gradient is hence much more suitable to describe different types of non-individualism, including that of the Chinese, and even collectivism. Furthermore, the concept of guanxi gradient has nothing to do with dichotomy. This gives guanxi gradient plenty of scope to delineate different types of interpersonal relations of the whole spectrum globally, including individualism. In this aspect, the guanxi gradient is thus believed to be more useful than the concept of individualism-collectivism.

Crew resource management is designed to promote interpersonal cooperation in the cockpit. However, as designed by Westerners, CRM conveys subconsciously the individualistic mentality, i.e., without consideration of the influence of the degree of intimacy between the crew. In Chinese language, it can be said there is no guanxi in CRM. Distortion is therefore almost certain to take place in a crew with a steep guanxi gradient. For closely intimate relation within the crew, communication and cooperation will be enhanced. On the other hand, the monitoring function is likely to deteriorate. Work sharing is found to be another function easily confused, because each crewmember is more than happy to help the other when they have a very intimate relationship. In a cockpit with estranged relationships, the bright side is that monitoring functions will be preserved almost as designed, and the procedures will be followed automatically because each crewmember does not want to cause any trouble. The down side of this, however, includes deterioration in communication and cooperation. Also, the degree of teamwork is highly likely to be lowered. Generally speaking, from this study it is found that in a cockpit with a steep guanxi gradient the designed activities will be more easily influenced by personal feelings or emotions. In this case, self-discipline is needed to avoid any negative influences. However, this is not always easy for a less mature person to face. Hence, there is a simple way to deal with this to avoid a steep guanxi gradient in the cockpit, i.e., not to put two pilots with too intimate or too estranged relations in a narrow and coercive space like a cockpit.

3 Questionnaire

In order to understand the latent meaning of guanxi gradient and its hidden Chinese value together with its influence on CRM, a questionnaire was designed. Also, the same questions used in the 2000 test(Lu et al. 2000) were casted in this version to measure the change of guanxi gradient with time. There are three parts in this questionnaire which are basic perception, situational response and open questions. There are 34 questions in the first part. The interpersonal relations are categorized into three different kinds of guanxi: acquaintance, fellow and alien. The questions in this part ask how the work in the cockpit is to be done when facing partner having different guanxi. There are 9 questions asking about the attitude facing acquaintance, while 6 and 13 questions about fellow and alien. There are 6 general questions inquiring the attitude about safety and harmony.

In these 34 questions, 7 were designed to check attitude about monitoring function, 12 about communication, 7 about team work while 8 about workload. In addition, these questions were also arranged to reflect the underlying values, which were harmony with 11 questions, morality 9 questions and professionalism 14 questions. Here, harmony means the basic believe is to maintain the crew as a whole to work properly, morality means that pilots accept the position determines their behavior. For professionalism, it represents the capability is the main concern to make judgment.

The second part is composed of 12 questions. Each question is given a situation occurred in the cockpit requiring the operation defined in the crew resource management. There are only two answers to be chosen between the consideration of safety and harmony. One of the choices asks pilots to answer according to the regulations to keep safety first, while the other one requires pilots to consider the feeling of the partner before taking actions which do not demand strictly adherence to the regulations. The objective is to find whether the subject is to be inclined to safety or harmony when facing partner with different degree of intimacy.

The reason to study the attitude about harmony is as follow. In flight operation, safety is corroded once in a while by something like authoritarianism or guanxi. The main value underlying the behavior of authoritarianism and guanxi is harmony. It is agreed generally that one of the most important basic values of Chinese culture is harmony (Lu, et al. 2001). Possessing ideographic mode of thinking, the Chinese have a holistic, or systematic, view about almost everything including operations in the cockpit. One key characteristic of the systematic view is the emphasis on dynamic interaction between the composing parts within it. To keep the system work positively, the Chinese believe that equilibrium should be maintained as the healthy interaction needed for the system. If equilibrium is destroyed, the whole system will enter into a harmful cycling loop and finally break down, in which case everybody in the system will get hurt. Hence, the Chinese always try instinctively to avoid damaging the equilibrium in most of the situations. Conceptually, avoid damaging the equilibrium is almost identical to maintaining harmony.

In the last part of the questionnaire, there are only three open questions for the subjects to answer. The goal of these questions is to find out how important the interpersonal relation is in the cockpit, and what kind of partner the pilots wanted to work with.

Two pretests were run. There were 29 subjects for the first test and 42 for the second test together with interviews. Collecting the opinions, there are totally 75 pilots were tested and interviewed finally. Among them, 56% were Captains, 15% Chief pilots, 16% Instructor pilots while 13% First officer. There were 48% pilots with military background. All the pilots have the same nationality: R. O. China. The reliability of all these questions were tested with Cronbach and split-half reliability and found to be acceptable. The results were again confirmed by the follow-up interviews.

4 Results and Discussion

A similar test was run in the year 2000. The guanxi gradient was again calculated in the present study in which the survey was executed in 2008. In 2000, the guanxi gradient was found to be -0.1835 while -0.1648 in 2008. The gradient was flattening obviously, which means the attitude of Taiwaness pilots treating partners with different degree of intimacy was gradually similized. The rate of flattening was calculated to be 0.00234 per year. In the year 2000, it was estimated to be 0.00278, the agreement is quite acceptable although the actual change of guanxi gradient was found to be lower than the linear extrapolation. Influenced by the western individualism, the interpersonal relation in Taiwan is becoming closer and closer to that of the west. However, the process takes quite a long time. According to the results, the evolution speed of guanxi gradient is slower then was expected, which represents the change of cultural traits in general is more difficult then we thought.

As for the four main operations of CRM: monitor, communication, teamwork and workload, the guanxi gradient for each item was found to be different. The highest one is -0.3078 possessed by workload. It means that for a Chinese pilot, he is more willing to share the workload, i.e., helping each other, with a more familiar partner, while more inclined to stick to the rule when facing an alien partner. The monitor and communication have rather similar but lower guanxi gradient as workload. The lowest guanxi gradient was from teamwork with -0.1183. For a Chinese pilot, the degree of intimacy does not have strong influence when doing a job required by regulations in the cockpit. No matter which one, however, the above results indicate that the activities in CRM were performed differently with different partners in a cockpit with steep guanxi gradient.

The results also reveal the underlying believes of the pilots. No matter which one, the consideration of harmony, morality and professionalism all declined as the degree of intimacy decayed. It means that the local pilots gradually ignore the existence of the partners as the psychological distance increased. However, the decay rate was the highest for the value of morality. It is thought to be reasonable. For Chinese, aliens represent the people from somewhere totally different from us. They were so different that they have different moral. Even with different moral, we still have to maintain harmony with them and respect their profession. From the results, it can be seen that the average score of harmony is the highest, which represents harmony is the top value in the minds of Chinese pilots. Harmony should be preserved even when they work with partners they do not really care.

From the results of the second part, it is found that when most of the pilots face conflicting but relatively normal and not so dangerous situations, in average about 68% of them chose harmony over safety. That means unless the situation is somewhat dangerous, the local pilots are always prone to harmony in order to maintain smooth operation in the cockpit since safety is not really an issue.

If the given situation is about monitor, 39% of the pilots chose safety. In this case, the question asks about how the pilots behave when the partner accidently violates the regulations but without obviously deteriorating safety. Only 39% of the pilots chose to record faithfully and report to the superior. Most of the pilots chose harmony unless there is seemingly safety concern. When the situation is about communication, most of the local pilots still are incline to harmony. If the partner is very familiar, about 88% of the pilots will like to maintain a relaxed atmosphere in the cockpit. However, if the pilots fly with someone quite foreign, 68% of the pilots will choose sticking to the regulations to avoid any possible misunderstanding. As for teamwork, when flying with a good friend, 97% of the subjects will ask for help directly. If the partner is someone not quite familiar, 88% of the pilots will choose to respect the actions of the other if agreed by each other to avoid conflict. As for flying with someone never know, when something unexpected induced by the unknown behavior of the partner, 77% of the pilots will help unconditionally without asking anything.

From the third part, the first question asks the subjects about the factors influencing the performance of CRM. Among all the factors mentioned, 57% were related to harmony, 54% related to moral and 43% related to professionalism. It is again confirmed that the top value underlying the behavior in the cockpit is harmony. The third question is about the most wanted partner the local pilots like to work with. Among them, 71% of the pilots chose partners possessing quality and ability to maintain harmony, 51% of the pilots chose partners with good professionalism. Only 29% of the pilots chose to work with partners with good moral. It is natural for a Chinese to take harmony seriously even in a small group of two as in the cockpit. Harmony is also found to be the most wanted atmosphere for a Chinese pilot in the cockpit.

References

1. Chuan, Y.-C., Yang, K.-S.: Role-Playing and Cognition Structure. Indigenous Psychological Study, 282–338 (June 1997) (in Chinese)
2. Fei, X.-T.: The Native China. San-Lian Books Company, Hong Kong (1947) (in Chinese)
3. Helmreich, R.L., Merritt, A.C.: Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences. Ashgate Publishing Limited, Aldershot (1998)
4. Hofstede, G.: Culture's Consequences: International differences in work-related values. Sage, Beverly Hills (1980)
5. Hsu, F.L.K.: Americans and Chinese: Two Ways of Life. Natural History Press, New York (1970)
6. Jing, H.-S., Lu, P.-J., Yong, K., Wang, H.-C.: The dragon in the cockpit: the faces of Chinese authoritarianism. Human Factors and Aerospace Safety 2, 257–275 (2002)
7. Jing, H.-S.: Differentiated order in the cockpit. Human Factors and Aerospace Safety 4, 131–143 (2004)
8. Jing, H.-S., Chang, L.-S.: Guanxi gradient and flight safety. Human Factors and Aerospace Safety 6, 17–34 (2007)
9. Liang, S.M.: The Element of Chinese Culture. Chi-Chen Books Company, Taipei (1963) (in Chinese)
10. Lu, P.-J., Wang, S.-E., Co, H.-J., Jing, H.-S., Yuan, S.-F., Dai, M.: Crew Resources Management and Corporate Culture. Report of Aviation Safety Council, Taipei (2000) (in Chinese)
11. Lu, J.-P., Hong, L., Liang, R.-Y.: Chinese Personality Study. Yuan Liou Press, Taipei (2001) (in Chinese)
12. Yang, C.F.: Are Chinese really collectivism? In: Yang, K.-S. (ed.) Chinese Value System: Social Science Point of View. Kwei-Kuwan Books, Taipei (1993) (in Chinese)
13. Yu, Y.-S.: Modern Meaning of Chinese Culture from Value System. Times Culture Publications, Taipei (1984) (in Chinese)

Pilot Operating Characteristics Analysis of Long Landing Based on Flight QAR Data

Lei Wang[1,2,3], Changxu Wu[1], and Ruishan Sun[3]

[1] Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing, 100101
[2] Graduate University of Chinese Academy of Sciences, 19 A, Yuquan Road, Beijing, 100049
[3] Civil Aviation University of China, 2898 Jinbei Road, Tianjin, 300300
wanglei0564@hotmail.com

Abstract. Long landing events make up the largest percentage of all exceedance incidents and multiply the risk of runway excursions in landing phase. For the aim of exploring operating factors causing long landing, this study examined the pilot operating characteristics of long landing events by the methods of variance analysis, regression modeling and flare operation analysis based on flight QAR data. Finally it concluded that flare is the most critical operation in landing, which determining the touchdown distance by two key factors of flare time and flare height. Both of the control column and throttle operation plays an important role in the flare process. Pilots' faster pulling up columns and softer throttle closing is probably helpful for a successful flare. In addition, pilots need to control the aircraft to an appropriate groundspeed and descent rate before descending to the flare initial point. The conclusions are expected to be applied into practice to prevent the happening of long landing events.

Keywords: Pilot Operating Characteristics, long landing, QAR data, safety.

1 Introduction

Final approach and landing is the most important flight phase because a pilot needs to deal with more operations, decision-making, and workloads than other phases [1-4]. Accident statistics have also indicated that approach and landing was the most dangerous phase of flight, which accounted for only 4% of exposure time but resulted in over one-third of all commercial jet accidents. In particular, the landing phase alone accounted for 22% of total fatal accidents occurring between 2001 to 2010, despite the fact that the landing phase accounts for just 1% of average flight time [5].

A long landing event, which is one case of undershooting, is defined as an aircraft's contact with the runway over the normal touchdown area. A NLR (National Aerospace Laboratory of Netherlands) study has revealed that if the landing was long, the landing overrun accident risk was 55 times greater than when it was not long [6]. Referring to Iceberg Theory and Heinrich Accident Triangle [7], a runway excursion

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 157–166, 2013.
© Springer-Verlag Berlin Heidelberg 2013

accident is the smallest visible part of ice above the surface of water, while long landing events are he large invisible part of ice beneath the surface of water which is always omitted. Statistics also showed that long landing events regularly accounted for the largest part of exceedance events [8]. Therefore, long landing events are expected to attract more concern from aviation carriers and researchers.

The long landing event is generally monitored by using Quick Access Recorder (QAR) data in most commercial air carriers, but these data are also confidential for them. Meanwhile, there are few aviation administrators whom enforced their airlines to install QAR equipment on every transport jet. Therefore, QAR data were difficult and rarely utilized into research. Civil Aviation Administration of China (CAAC) has implemented the program of Flight Operations Quality Assurance (FOQA) since 1997, with all commercial airplanes of Chinese airlines obliged to install QAR or similar equipment. The practice has proved that QAR data were helpful for improving flight safety management and quality control.

In this study, we use QAR data to analyze long landing events and try to find the differences of pilot operating characteristics between normal landing and long landing. Meanwhile, the critical operation variables leading to long landing are expected to be analyzed.

2 Methods

2.1 QAR data

The QAR data in this study were collected from three commercial aircrafts (Boeing 737-800) of a local airlines. The data covered all normal and exceedance flights of these three aircrafts from the 1st of April to the 30th of May in 2012. First, 293 flight samples were selected and relevant QAR data files were downloaded from QAR ground station of airlines. The original data is a CSV (Comma Separated Value) file with thousands of rows and columns. Therefore, VBA (Visual Basic for Applications) programing functions in Microsoft Excel was applied and 19 columns of original QAR data of every file were refined as following. Finally we also compiled the VBA program to calculate 19 parameter variables and touchdown distance of each flight sample.

These parameter variables covered all flight and operational parameters in the critical visual and manual landing stages from the height of autopilot-disconnected to touchdown. Generally the threshold of identifying normal and long landing was set as 2600 feet for this aircraft type by the airlines. Based on this threshold, 293 cases of QAR data were divided into two groups with 119 cases of normal landing (Group 1) and the other one was 174 cases of long landing (Group 2). QAR data of 119 normal landing events and 174 long landing events were regarded as two groups of independent samples. The mean and standard deviation of touchdown distance of these two groups respectively was 2248.88±247.27 (Group 1) and 3082.62±357.64 (Group 2).

Table 1. Selection of parameters

Classification of parameters	Parameter name of QAR data	Explanations
Height	RADIO HEIGHT LRRA	Radio height
Speed	GROUND SPEED	Groundspeed
	VERT SPD CMD	Descent rate
	AIR SPD	Airspeed
Operation	SELTD TRA FILTERED	Throttle resolver angle
	CONTRL COLUMN POSN	Control column position
	CONTRL COLUMN FORCE	Control column force
	CONTRL WHEEL FORCE	Control wheel force
	CONTRL WHEEL POSN	Control wheel position
	FLAP HANDLE POSN	Flap handle position
	SPD BRAKE HANDLE POSN	Speed brake handle position
	RUDD PEDAL POSN	Rudd pedal position
Attitude	CAP DISP PITCH ATT	Pitch angle
	CAP DISP ROLL ATT	Roll angle
Configuration	FLAP	Flap
	AILERON POSN	Aileron
	ELEV POSN	Elevator
	RUDD POSN	Rudder
Acceleration	VERT ACCEL	Vertical acceleration

2.2 Variable Selection and Flare Operation

Currently most of commercial aircrafts have an advanced autopilot system and even an automatic landing system, so that the main task of the human pilot during most of flight time is monitoring instrument panels and checking the aircraft states for any abnormality. However, during take-off and final landing (below 60 *m*), the automatic system is rarely used and the aircraft is still operated by pilot. Especially, the final landing control is known as the most difficult maneuvers for airline pilots in normal operation. Pilots are required to control groundspeed and descent rate in a quick and few seconds depending on visual and other situational information. Among that a characteristic and critical maneuver is called flare, which involves lifting of the nose to both land the aircraft on the main gear first and decrease the sink rate and vertical load at the landing. Flare operation would make large influences on final landing performance including touchdown distance and also is one of the most skilled operation in flight [9-12]. Therefore, the pilot operation below 200 feet, especially the flare operation was selected as the main subject for analysis.

2.3 Statistical Methods

Before data analyzing, the final landing track of two groups of flight samples was depicted based on their height average at every time scale before touchdown. The height change and difference between two groups would be found from the figure of track.

These parameters regarding with flare operation were the main variables to examine. Firstly, they were classified into four categories of height and time, operation parameter, configuration and attitudes and flight performance parameters. There were 20 variables included in the four categories in total. It needs to explain that the variable of *Flare Time* means the total time from flare initial point to touch down point. In addition, the flare operation initial point in this study is higher than definition in most of flight manuals where it is always defined as 30 feet. This is because any slight pulling up of control column could be recorded by Quick Access Recorder, and it leads that the time and height of flare is earlier than theoretical value. Secondly, the normal distribution test was carried on. Then for the aim of difference analysis, one way ANOVA was used to examine variables which were subjected to normal distribution and non-parameter K-W test for other ones. Thirdly, in order to further analyze the correlations between touchdown distance and the 20 variables related with flare operation, a multiple linear regression model was developed. Considering the probable collinearity between independent variables, the stepwise regression method was used for eliminating and the stepping criteria were based on probability of F ($F \leq 0.05$ for entering and $F \geq 0.10$ for removal). Meanwhile, the effectiveness of the model was analyzed.

For further observing dynamic change of flight variables in final landing phase and their differences between two groups, the altitude of 200 to 0 feet was divided into four flight levels (200-150-100-50-0 feet) and selected 20 variables were measured in every level. The multivariate analysis process of general linear model was introduced to compare the differences in the two groups. Especially, the variable *Control Column* and *Throttle Resolver Angle* was analyzed in detail and presented in this paper.

3 Results and Discussions

3.1 Analysis of Final Landing Track

Scaling with the time before aircraft touchdown and the height mean of each group, the final track of two kinds of landing is as showing as Figure 1. The first second on horizontal axis is touchdown point.

As seen in Figure 1, following points are easily to be found.

(1) The average height of normal landing is higher than long landing at each same second before touchdown.
(2) The height change of two groups was basically linear before flare operation initial point, and their slope was also basically the same.
(3) There is no significant difference between the flare initial height of two groups, which are both around 50 feet. The height change shows difference after flare operation.
(4) The most remarkable difference is the time of flare starting and also the flare operation time, the normal landing is around 8 seconds and long landing is around 10 seconds.

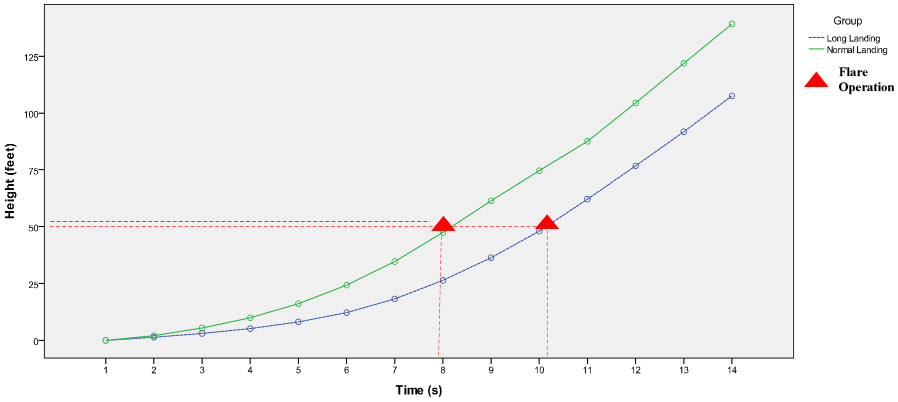


Fig. 1. Final track of normal and long landing

3.2 Data Analysis of Flare Operation Point

3.2.1 Results of Difference Analysis

The results of difference analysis on variables at flare point are as showing as Table 2.

Table 2. Difference analysis on variables of flare point

Parameter Categories	Variable Names	Group	N	Mean±SD	*p* (K-S)	*p* (ANOVA/K-W)
Height and Time	Flare Height	1	119	52.076±21.975	0.044	0.351
		2	174	50.787±25.005	0.000	
	Flare Time	1	119	7.891±2.102	0.002	0.000
		2	174	10.684±2.589	0.003	
Operation Parameters	Throttle Resolver Angle	1	119	49.570±1.926	0.996	0.052
		2	174	49.062±2.355	0.132	
	Control Column	1	119	1.004±0.773	0.144	0.818
		2	174	1.023±0.667	0.108	
	Column Force	1	119	2.072±0.942	0.284	0.419
		2	174	2.164±0.978	0.270	
	Control Wheel	1	119	0.461±8.938	0.894	0.330
		2	174	-0.628±9.674	0.207	
	Wheel Force	1	119	-0.017±0.424	0.423	0.139
		2	174	0.055±0.466	0.012	
	Flap Handle Position	1	119	31.597±3.678	0.000	0.008
		2	174	30.632±2.441	0.000	
	Speed Brake Position	1	119	2.949±0.822	0.000	0.286
		2	174	2.843±0.912	0.000	
	Rudder Pedal	1	119	0.563±0.250	0.000	0.564
		2	174	0.579±0.142	0.041	

Table 2. (*Continued*)

Configurations and Attitudes	Elevator	1	119	2.492±0.938	0.110	0.547
		2	174	2.431±0.794	0.060	
	Aileron	1	119	1.504±1.864	0.713	0.307
		2	174	1.267±2.006	0.527	
	Flap	1	119	31.597±3.678	0.000	0.008
		2	174	30.632±2.441	0.000	
	Rudder	1	119	-0.160±0.605	0.858	0.189
		2	174	-0.248±0.528	0.996	
	Pitch Angle	1	119	1.464±0.653	0.683	0.596
		2	174	1.421±0.704	0.229	
	Roll Angle	1	119	-0.345±1.221	0.628	0.074
		2	174	-0.091±1.173	0.097	
Flight Performance	Air Speed	1	119	148.462±4.871	0.429	0.000
		2	174	150.575±4.402	0.407	
	Groundspeed	1	119	146.277±7.453	0.834	0.000
		2	174	152.080±7.375	0.842	
	Descent Rate	1	119	-813.849±148.094	0.007	0.131
		2	174	-825.448±142.712	0.065	
	Vertical Acceleration	1	119	1.047±0.037	0.273	0.005
		2	174	1.061±0.040	0.167	

As seeing from Table 2, seven variables show the difference at the significant level of 0.05, which are *Throttle Resolver Angle, Flap Handle Position, Flap, Air Speed, Groundspeed, Vertical Acceleration* and *Flare Time*. However, there are only *Air Speed, Groundspeed, Flare Time* represented the significant difference at the level of 0.001. This point means that the major difference of two groups is reflected on longitudinal speed including airspeed and groundspeed. In fact, the three variables of *Throttle Resolver Angle, Flap Handle Position, Flap* would make direct effect on longitudinal speed. Meanwhile, we can find that most of operation variables such as *Control Column* and *Control Wheel* do not represent a significant difference at flare point. It is probably because that most of operation are a consequent movement, their differences exist in a time period or a stage, not at a point. Therefore the difference analysis based on flight height change was also carried on in coming steps.

3.2.2 Multiple Linear Regression Model

The results of stepwise linear regression shows that five significant predictors are included in the final regression model, which are *Flare Time, Flare Height, Groundspeed, Descent Rate, Vertical Acceleration*.

The R square of the final model achieves 0.974, which indicates that the relatively good fitness of this linear model ($F(5,287) = 1074.868$, $p < 0.001$). The linear regression model is expressed as the following equation:

$$TD = -1823.106 + 235.295x_1 - 18.940x_2 + 19.111x_3 + 0.204x_4 + 566.668x_5 \quad (1)$$

Table 3. Coefficients of model

No.	Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
	(Constant)	-1823.106	239.156		-7.623	0.000		
x_1	Flare Time	235.295	3.606	1.255	65.257	0.000	0.478	2.094
x_2	Flare Height	-18.940	0.442	-0.869	-42.863	0.000	0.430	2.328
x_3	Groundspeed	19.111	0.909	0.292	21.019	0.000	0.914	1.094
x_5	Descent Rate	0.204	0.053	0.057	3.863	0.000	0.813	1.231
x_5	Vertical Acceleration	566.668	204.316	0.043	2.773	0.006	0.735	1.361

The standardized regression model, which could present this correlation directly, is introduced and written as following equation:

$$Z_{TD} = 1.255Z_{x_1} - 0.869Z_{x_2} + 0.292Z_{x_3} + 0.057Z_{x_4} + 0.043Z_{x_5} \tag{2}$$

In Table 3, all of the coefficients are highly statistically significant ($p < 0.01$). The variable x_1 (*Flare Time*) carries the biggest one (1.255) and had the greatest impact on touchdown distance. This point is consistent with the results of difference analyses. It should point out that the variable x_2 (*Flare Height*) also carried a great contribution on touchdown distance, despite of there is no significant difference between normal and long landing groups.

The Durbin-Watson test shows that there are no autocorrelations existing among predictors (Durbin-Watson = 1.884). All VIF coefficients of these five predictors are less than three which meant that collinearity level of independent variables is acceptable. A P-P plot demonstrates that the regression standardized residual is basically subjected to a normal distribution. It means that the normality assumption of regression is not violated.

3.3 Flight Operation Analysis of 200-0 Feet

According to the landing description in flight manual, both groundspeed and descent rate are the two most critical flight parameters which human pilots should monitor and the most important operation actions are finished by control column and throttle. The difference of 20 variables from 200 feet to touchdown were analyzed by using repeated measure and one way ANOVA. Due to length limitation, here only the results of groundspeed, descent rate, control column and throttle are presented.

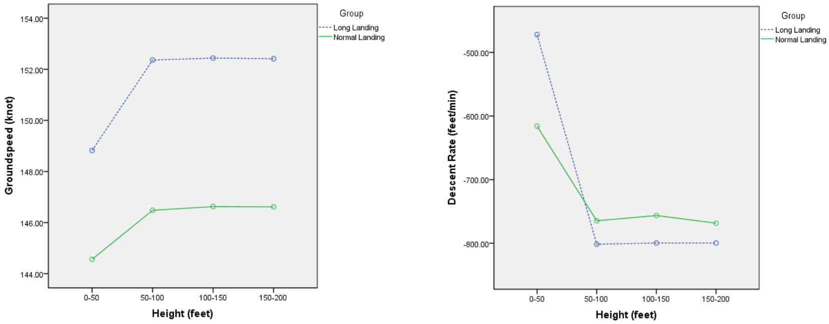


Fig. 2. Difference analysis of groundspeed and descent rate

As shown in Figure 2, the significant difference of variable *Groundspeed* is existing in the whole stage of 200-0 feet ($F(1,291) = 37.265$, $p < 0.001$), the groundspeed of long landing group is higher than normal group. The results of repeated measure ANOVA showed that the group effect of variable *Descent rate* is not significant ($F(1,291) = 1.802$, $p = 0.18$). However, the results of one way ANOVA on each stage showed that the difference is significant. The descent rate of long landing is bigger than normal group before 50 feet which is also the flare initial point, but it changes a lot after past 50 feet and makes a more significant difference between groups ($F(1,291) = 234.373$, $p < 0.001$).

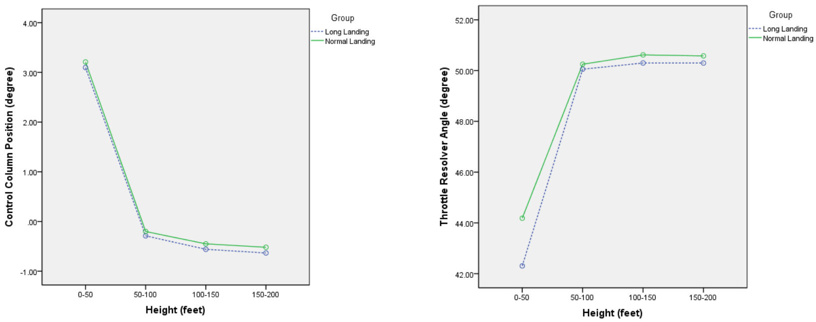


Fig. 3. Difference analysis of control column and throttle resolver angle

In Figure 3, the control column and throttle change greatly after passing 50 feet (flare operation initial point). There is no difference between control column of two groups ($F(1, 291) = 2.771$, $p = 0.097$) . The column change degrees and trend of two flights both are keeping the same. However we need to note that the time of operating column is definitely different, which means that the speed of pulling column is significant different. The normal landing group is faster than long landing group (7.9 seconds for normal landing and 10.7 seconds for long landing). There is also no difference found for throttle operation before 50 feet, the main difference is reflected after flare starting when pilot begins to decrease thrust. The result of one way ANOVA is $F(1, 291) = 46.351$, $p < 0.001$. The throttle change of normal group is smaller than long landing group, which means that the throttle of normal operation is closed more steady and softly.

4 Conclusions

Long landing is one kind of unsafe incident which could increase the risk of runway excursions. It occurs frequently and sometimes lead to runway excursion accidents. Though many studies regarding runway excursion have been conducted, most of them have been based on accident investigations, models, or experiments rather than real flight data. Because real flight data was hard to be obtained from air operators. Basing on flight QAR data, this study provides a new way to analyze long landing and its operating characteristics. The main works are concluded as following.

- Flare is the most critical operation in final landing, which would make great influences on touchdown distance through the factors of flare operation time and height.
- Flare time is the most significant different variable between normal landing and long landing, where the variable mean is 7.9 seconds for normal landing and 10.6 for long landing.
- The control column and throttle operation below 50 feet represents a difference between groups and plays a great role on flare together. Pilots' faster pulling up columns and softer throttle reduction maybe helpful for a better flare and landing.
- Pilot is suggested to monitor and control aircraft to an appropriate longitudinal and vertical speed when entering into a manual operation phase in landing. Groundspeed and descent rate are always the two crucial parameters guiding to a good landing performance.

In future work, a more qualitative model is expected to be developed for explaining how control column and throttle operation contributes to touchdown distance and other landing performance.

Acknowledgments. We appreciate the support of this work from National Basic Research Program of China (No.2010CB734105) and National Natural Science Foundation of China (No. 60979009).

References

1. Hawkins, F.H.: Human Factors in Flight, 2nd edn. Ashgate, Brookfield (1993)
2. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn. Prentice Hall Press, Upper Saddle River (2000)
3. International Civil Aviation Organization. Accident Prevention Programme (Manual). ICAO, Montreal (2005)

4. Shappell, S., Detwiler, C., Holcomb, K., Hackworth, C., Boquet, A., Wiegmann, D.: Human error and commercial aviation accidents: an analysis using the human factors analysis and classification system. Human Factors 49(2), 227–242 (2007)
5. Boeing: Statistical summary of commercial jet airplane accidents, worldwide operations, 1959-2010. Boeing Commercial Airplanes, Seattle (2011)
6. Gerard, V.E.: When a runway is not long enough to land on. Transport Canada, Aviation Safety Letter (1), 27–28 (2006)
7. Heinrich, H.W.: Industrial Accident Prevention: A Scientific Approach, 4th edn. McGraw-Hill, New York (1959)
8. Sun, R., Han, W.: Analysis on parameters characteristics of flight exceedance events based on distinction test. Journal of Safety Science and Technology 7(2), 22–27 (2011)
9. Molesworth, B., Wiggins, M.W., O'Hare, D.: Improving pilots' risk assessment skills in low-flying operations: The role of feedback and experience. Accident Analysis and Prevention 38(5), 954–960 (2006)
10. Mulder, M., Pleijsant, J.M., Vaart, H., Wieringen, P.: The effects of pictorial detail on the timing of the landing flare: results of a visual simulation experiment. The International Journal of Aviation Psychology 10(3), 291–315 (2000)
11. Benbassat, D., Abramson, C.I.: Landing flare accident reports and pilot perception analysis. International Journal of Aviation Psychology 12(2), 137–152 (2002a)
12. Flight Safety Foundation. Stabilized approach and flare are keys to avoiding hard landing. Flight Safety Digest 23(8), 1–16 (2004)

The Investigation of Visual Attention and Workload by Experts and Novices in the Cockpit

Wen-Chin Li*, Fa-Chung Chiu, Ying-shin Kuo, and Ka-Jay Wu

Psychology Department, National Defense University,
Beitou District, Taipei City 112, Taiwan
w.li.2002@cranfield.ac.uk

Abstract. Under high pressure of flight mission and dynamic aircraft maneuvers in the tactic missions, pilot faces additional difficulties and increased mental workload. Workload could increase the error of flight operation, decrease efficiency of pilot's decision-making. Experts had significantly shorter dwells, more total fixations, more aim point and airspeed fixations and fewer altimeter fixations than novices, experts were also found to have better defined eye-scanning patterns. This research applies the eye-tracking technology for analyzing visual attention, emWave-2 for measuring physiological coherence, and NSAS-TLX for investigating subjective cognitive efforts. The participants of this research consisted of 41 fighter pilots. The present study is applying new technology to understand the pilots' workload and visual attention in the cockpit for conducting a simulated air-to-air tactic operation. There is a raising need for further research regarding mental workload and stress management program for real-time flight operations.

Keywords: Cognitive Processes, Eye Movement, Visual Attention, Workload.

1 Introduction

The natural limitation of human being's cognitive processes and the vast number of tasks are reasons for increasing critical workload levels for military pilots. Under high pressure of tactic missions and dynamic aircraft maneuvers, the pilot faces additional difficulties and increased workload during hostile environmental (Ahlstrom, 2003). Workload can negatively affect operator performance and increase the probability of operational hazards. Peter, Jennifer and Joey (2001) found that experts had significantly shorter dwells, more total fixations, more aim point and airspeed fixations and fewer altimeter fixations than novices. Experts were also found to have better defined eye-scanning patterns. The visual scanning difference between novice and experts are correlated with better performance by experts, as experts should have shorter dwell and more fixations than novices on all the instruments (Bellenkes, Wickens, & Kramer, 1997).

* Corresponding author.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 167–176, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Fox, Merwin, Marsh, McConkie, & Kramer (1996) found that experts are more likely to use peripheral vision to process a broader range of visual cues than novices do. The pattern of acquisition of cue-based information provides an opportunity to assess the application of distinct cognitive skills. Eye movements are useful to reveal the diagnostic information that enables the development of appropriate strategies which efficiently target a particular feature of the performance for completing a task (Rayner, 1998). However, deliberate missions result in higher workload, such as controlling aircraft for tactic maneuver is a stressful operation which needs high situation awareness to make risk assessment, might increase pilot's mental workload (Schipani, 2003). Pilots have to make in-flight decisions not only about the management of the systems in the cockpit, but also about the operating state of that airspace in dynamic situations, the workload may be increased dramatically during abnormal situations and system failures (Weiner, 1989).

The NASA Task Load Index (TLX) is a popular technique for measuring subjective mental workload. It relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales (Hart & Staveland, 1988). The scales range from 0 to 100 and were divided into 20 increments. By analyzing the six weighted scores to understand which dimension has higher degree of workload. Aside from using self-reported subjective workload ratings such as NASA-TLX for evaluating operator's workload, pupillary response has also been proposed as an index of the amount of cognitive processing (Beatty, 1982). Eye movement measurement offers deep insights into human- machine interaction and the mental processes of pilots. Measurements based on different aspects of ocular behavior, such as the number of fixations, dwell time, and the dilation of pupil, have been used to reveal the status of mental workload. There are existing diversity findings of eye movement in the literature. Beatty (1982) reviewed a large amount of experimental data and concluded that increasing in pupil size correlate with increasing in workload. However, some experiments have not detected a relationship between task difficulty and pupil dilation (Lin, Zhang, & Watson, 2003). There was evidence that increasing in workload could increases dwell time and the frequency of long fixations (Van Orden, Limbert, Makeig, & Jung, 2001). Athènes, Averty, Puechmorel, Delahaye and Collet (2002) found that workload could increase the error of operation, and the decreased fixation duration appeared to predict upcoming errors in the auditory task (Tsai, Viirre, Strychacz, Chase, & Jung, 2007).

Workload has impact to cognitive processing and creating psychological and physical stress. The deleterious effects of stress on human beings' performance are pervasive, as stress may result in physiological changes such as emotional reactions such as fear, anxiety, frustration, and cognitive effects such as narrowed attention, longer reaction time to peripheral cues and decreased vigilance, and degraded problem solving (McCraty, Atkinson, Tomasino, & Bradley, 2009). Training programs, often called stress management programs (SMPs), have been developed, primarily by clinical psychologists, for this expressed purpose. Among the most common procedures are progressive muscle relaxation, meditation, biofeedback, and cognitive-behavioral skills training. Human beings cognitive processes and performance have been the subjects of much concentration in research and practice including situation awareness, decision-making, mental workload, and operational performance. However, there are lots of argument for finding strong empirical evidence and lacking scientific status for military pilots. Therefore, this research applied the eye-tracking technology, em-Wave-2 and NASA-TLX to investigate objective and subjective cognitive effort involved in attending to information presented in stressful situation for military pilots. By examining pilots' eye movement's patterns and performance compared with pilots' physical coherence levels, and subjective stress levels, it is hope to discover the appropriate interventions in flight operations for improving the training effectiveness and aviation safety.

2 Method

2.1 Participants

There were 41fighter pilots participated in this research. The flying hours is between 335 and 3200 hours, the rank is between first lieutenant and colonel, the age is between 25 and 45 years old.

2.2 Apparatus

Eye-Tracker. Eye movement data were collected by an head-mounted ASL (Applied Science Laboratory) Mobile Eye with 76 gram in weight. When combined with an optional head-tracking device and eye/head integration software, the eye tracker can also measure pilot's eye line of gaze with respect to stationary surfaces in the environment. It is designed to be durable under a variety of active applications and its light weight is suitable to detect the eye movement when pilot operates the aircraft under active and dynamic flight tasks.

Flight Simulators. This study used fighter simulator, a dynamic high fidelity trainer that replicates actual aircraft performance, navigation and weapon systems. This simulator provides a realistic representation of the flight management system. The instructors can supervise the participated pilot's performance and the instrument data from three screens. In instrument fight task, the integrate control panel (ICP) 、data entry display 、horizontal situation display 、head-up display (HUD) , Left and Right Multi-display provide critical information for tactic mission. Pilots cross check those instrument to maintain the speed 、altitude 、heading and position. Therefore, this study set those six gauges as the area of interests (AOIs) to analyze the eye movement data. The scenario of flight simulator is Air-to-Air Combat.

NASA-TLX. The NASA Task Load Index (TLX) is a popular technique for measuring subjective mental workload. It relies on a multidimensional construct to derive an overall workload score based on a weighted average of ratings on six subscales:

(1) Mental demand: How much mental demand and perceptual activities you would use; (2) Physical demand: How much is the degree of physical demand; (3) Temporal demand: How much is the degree of time pressure; (4) Performance: How do you feel about the flying time and the performance in flight? (5) Effort: How much difficult do you think? (6) Frustration: How much frustration and disappointment do you feel (Hart & Staveland,1988).

EmWave-2. Physiological coherence is measured by HRV (heart rate variability) analysis, which reflect heart-brain interactions and autonomous nervous systems (ANS) dynamics and is reflected in a heart rhythm pattern. The Coherence Ratio bar chart is divided into three levels: Low, Medium and High coherence. The positive emotions such as appreciation and compassion, as opposed to negative emotions such as anxiety, anger, and fear, are reflected in a heart rhythm pattern that is more coherent. The coherent state has been correlated with a general sense of well-being, and improvements in cognitive, social, and physical performance. Heart rhythm coherence is indeed associated with significant improvements in cognitive performance. EmWave-2 is used in many areas including military operational personnel and Olympic athlete, it has a significant improvement on stress management (McCraty, Atkinson, Tomasino, & Bradley, 2009).

2.3 Procedure

Participants were asked to perform Air-to-Air Combat using the flight simulator, the procedure included: (1) an orientation to the experiment (10 minutes); (2) eye-tracker calibration in the cockpit of flight simulator (5-10 minutes); (3) operate Air-to-Air Combat task on flight simulator for collecting eye movement data by eye-tracker, and coherence level by emWave-2 (3-5 minutes); (4) rate NASA TLX (10-15minutes). Each session was conducted by an eye-tracker operator and a flight instructor. The instructor evaluates pilot's performance base on the figures of flight simulator, and the simulator control panels record the time for terminating target and weapon consumption. Eye movement patterns, video, verbal protocol data were collected for further analysis.

3 Results and Discussion

3.1 Sample Characteristics

Participants consisted of 41 fighter pilots from R.O.C. air force. The qualifications for flying experience were not combat ready, combat ready, 2-aircraft team leader, 4-aircraft team leader, daytime back seat instructor, night back seat instructor, and training instructor; flying hours between 335 and 3200 hours; the age of participants between 25 and 45 years old (Table 1).

Table 1. Demographical Variables of Participants *(N=41)*

Variables	Group	Frequencies (%)
Age	25－30	14(34.1%)
	31－35	11(26.8%)
	36－40	9(22%)
	41－45	7(17.1%)
Rank	First lieutenant	1(2.4%)
	Captain	17(41.5%)
	Major	9(22%)
	Lieutenant Colonel	13(31.7%)
	Above colonel	1(2.4%)
Qualification	Not combat ready	2(4.9%)
	Combat ready	12(29.3%)
	Two aircraft team leader	4(9.8%)
	Four aircraft team leader	9(22%)
	Daytime back seat instructor	3(7.3%)
	Night back seat instructor	0(0%)
	Training instructors	11(26.8%)
Flying hours	Under 1000	17(41.5%)
	Above 1000	24(58.5%)

3.2 The Different Workload Level by NASA-TLX and Coherence Scores

There were two groups of pilots by operational qualifications, junior pilots including not combat ready and combat ready pilots; and senior pilots including team leaders, back-seat instructors, and training instructor. The workload scores of NASA-TLX and em-Wave-2 (coherence) during Air-to-Air Combat by t-test shown as table 2. There were two dimensions of NASA-TLX, Performance ($p < .05$) and Frustration ($p < .05$), with significant difference between junior and senior pilots. Also, senior pilots showed significant higher coherence levels than junior pilots by emWave-2 ($p < .001$).

Results show that senior pilots have higher performance (M=8.80 vs M=5.45) and coherence levels (M=27.71 vs M=15.22), and lower frustration than junior pilots (M=7.41 vs M=10.45). The coherent state has been correlated with a general sense of well-being, and improvements in cognitive, social, and physical performance. Senior pilots have more experience of tactic training, competent of system integrations for flight operations, and mature maneuver skills than junior pilots. The findings of high performance and low frustration by NASA-TLX, and high physical coherence of heart rhythm by emWave-2 for senior pilots in this study, have coincided with both laboratory and natural settings (Tiller, McCraty, & Atkinson, 1996). There is abundant evidence that emotions alter the activity of the body's physiological systems and that beyond their pleasant subjective feeling, heartfelt positive emotions and attitudes provide a number of benefits that enhance physiological, psychological, and social functioning (McCraty, Atkinson, Tiller, Rein, & Watkins, 1995). Zakowski, Hall and

Baum (1992) concluded that stress management is a valuable weapon not only against temporary interference with performance attributable to stress, but also against certain types of disease and physical illnesses and against the progression of diseases already established. Furthermore, stress management training could bring individuals' ability to diminish stress and increase their adaptation based on a stressful situation (Linden, 2004). The different approaches between objective measurement of emWave-2 and subjective rating scale of NASA-TLX show senior pilots have better performance and physical coherence for coping with stress. There is a raising need for developing stress management interventions for military pilots for improving aviation safety.

Table 2. The Different Workload level by NASA-TLX and emWave-2

Dimensions	***Means (SD)***		***t***	*D*	***p***
	Seniors	Juniors			
Mental Demand	10.13 (7.48)	11.23 (9.05)	.55	.13	.59
Physical Demand	15.64 (9.95)	16.32 (8.44)	.33	.07	.74
Temporal Demand	12.70 (8.18)	10.20 (8.13)	1.20	.31	.23
Performance	8.80 (7.21)	5.45 (6.03)	2.25	.50	.03
Effort	11.45 (6.86)	13.55 (6.94)	1.37	.30	.18
Frustration	7.41 (6.34)	10.45 (6.22)	2.18	.48	.03
Coherence	27.71 (5.53)	15.22 (7.80)	8.47	1.85	.001

3.3 The Percentage of Gaze in Region to Total Gaze for Different Area of Interests (AOIs)

The differences between senior and junior pilots for percent gaze in region to total gaze at different area of interests (AOIs) by t-test (table 3). The results show that senior pilots have significant more gaze on Left Multi-display than junior pilots during Air-to Air combat ($p < .01$). The result is difference compared with the research conducted by Peter, Jennifer and Joey (2001) which proposal that experts had significantly shorter dwells, more total fixations, more aim point and airspeed fixations and fewer altimeter fixations than novices. Experts should have shorter dwell and more fixations than novices on all the instruments. Pilots have better performance as experience pilots aware how to pay attention to certain instrument (Left Multi-display) at the critical timing for tactic maneuver.

Table 3. The percentage of gaze in region to total gaze for AOIs

Dimensions	*Means (SD)*		*t*	*D*	*p*
	Seniors	**Juniors**			
Head-up Display	53.28 (19.23)	50.78 (21.48)	.56	.12	.58
Integration Control Panel	1.35 (1.31)	1.76 (2.51)	.94	.20	.35
Data Entry Display	.18 (.38)	.12 (2.6)	.87	.03	.39
Right Multi-display	.67 (1.23)	.63 (1.50)	.13	.03	.90
Left Multi-display	3.37 (5.22)	.91 (1.85)	2.72	.63	.01
Horizontal Situa-tion Display	.08 (.22)	.11 (.45)	.52	.08	.61

3.4 Average Pupil Diameter on AOIs

There were significant differences between senior pilots and junior pilots on the Average Pupil Diameter at Left Multi-display by t-test ($p < .05$). The data show that senior pilots' Average Pupil Diameter significant bigger than junior pilots at Left Multi-display (M=73.69 vs M=55.60) (Table 5). The results demonstrated that senior pilots' pupil diameter dilated significantly than junior pilots at Left Multi-display, it is for the collecting target's information from tactic radar. It is the focus of pilots' attention and

Table 4. Average pupil diameter on different AOIs

Dimensions	*Means (SD)*		*t*	*D*	*p*
	Seniors	Juniors			
Head-up Display	86.67 (11.42)	89.50 (10.92)	1.14	.25	.26
Integration Control Panel	81.28 (24.93)	83.79 (23.00)	.47	.10	.64
Data Entry Display	33.44 (42.14)	27.19 (40.56)	.68	.15	.50
Right Multi-display	46.91 (45.50)	43.69 (45.95)	.32	.07	.75
Left Multi-display	73.69 (36.42)	55.60 (44.59)	2.02	.53	.04
Horizontal Situation Display	12.84 (3.24)	11.51 (29.66)	.20	.06	.85

getting situational awareness during Air-to-Air mission. Hilburn, Bakker, Pekela, & Parasuraman (1997) found an increase in pupil diameter when subjects performed well on the auditory task. The results also found senior pilots have better performance than junior pilots.

4 Conclusion

This research applied subjective measurement by NASA-TLX, and objective measurement by eye-tracker and emWave-2 for approaching military pilots physical coherence, mental workload and eye movement pattern for developing future training interventions. The weakness of subjective measurement is the self-report scale might be affected by social desirability and introspective limits (Perugini & Banse, 2007). Therefore, the objective measurement of workload and attention distribution by emWave-2 and eye-tracker are to improve the defect of self-report approach. The emWave-2 found that the coherence scores have significant difference between expert pilots and novice pilots. The effect size by emWave-2 is 1.85 which is much bigger than .50 by self-report of NASA-TLX. Eye-tracker is a powerful research apparatus not only for objective measurement but also for investigating the cognitive processes for task performance (Rayner, 1998). The experienced pilots have paid more attention on Left Multi-display, and the pupil diameter larger than novice pilots. Rehder and Hoffman (2005) also found that operators would spend more time on the target which attracting attention for completing task. Expert pilots know how to pay attention at critical time for getting important information for tactic operations. Eye-tracking device has already been shown to be a useful measurement for attention allocation.

Workload is an increasingly salient factor in advanced complex environment, the potential for stress-induced error spans both high-technique cockpit and ground operation for aviation domain, and effects of stress on physiological reactions, cognitions, emotions, and social behavior are manifold. This research found that pupil dilation is a valid and reliable indicator of mental workload and increase in pupil diameter when pilots performed well on the cognitive process for getting situational awareness. More experience pilots have higher physical coherence, bigger pupil dilation and more gaze on Left Multi-display for Air-to-Air mission, and have better performance and lower frustration than less experienced pilots. Although the mechanism of human perception is in the same operation, the purpose of this study is to explore the correlation among pilots' performance, workload and eye movements. By examining pilots' eye movement's patterns and physical coherence levels compared with pilots' subjective stress levels, it is hope to discover the role of cognitive effort in flight operations for improving the training effectiveness and aviation safety.

References

Ahlstrom, U.: Current trends in the display of aviation weather. Journal of Air Traffic Control 45(3), 14–21 (2003)

Athènes, S., Averty, P., Puechmorel, S., Delahaye, D., Collet, C.: ATC complexity and controller workload: Trying to bridge the gap (2002)

Averty, P., Collet, C., Dittmar, A., Athènes, S., Vernet-Maury, E.: Mental workload in air traffic control: an index constructed from field tests. Aviation, Space, and Environmental Medicine 75, 333–341 (2004)

Beatty, J.: Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychological Bulletin 91, 276–292 (1982)

Bellenkes, A.H., Wickens, C.D., Kramer, A.F.: Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. Aviation Space and Environmental Medicine 68, 569–579 (1997)

Endsley, M.R.: Level of automation: Integrating humans and automated systems (1997)

Fox, J., Merwin, D., Marsh, R., McConkie, G., Kramer, A.: Information extraction during instrument flight: An evaluation of the validity of the eye-mind hypothesis. In: Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting (1996)

Hart, S.G., Staveland, L.E.: Development of a multi-dimensional workload rating scale: Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload (1988)

Hilburn, B.G., Bakker, M.W.P., Pekela, W.D., Parasuraman, R.: The effect of free flight on air traffic controller mental workload, monitoring and system performance (1997)

Lin, Y., Zhang, W.J., Watson, L.G.: Using eye movement parameters for evaluating human–machine interface frameworks under normal control operation and fault detection situations. International Journal of Human Computer Studies 59, 837–873 (2003)

Linden, W.: Stress management: From basic science to better practice. Sage, Thousand Oaks (2004)

McCraty, R., Atkinson, M., Tiller, W.A., Rein, G., Watkins, A.D.: The effects of emotions on short-term power spectrum analysis of heart rate variability. The American Journal of Cardiology 76(14), 1089–1093 (1995)

McCraty, R., Atkinson, M., Tomasino, B.A., Bradley, D.: The coherent heart heart–brain interactions, psychophysiological coherence, and the emergence of system-wide order. Intergral Review 5(2), 10–115 (2009)

Peter, K., Jennifer, S., Joey, H.: Comparison of expert and novice scan behaviors during VFR flight. In: The 11th International Symposium on Aviation Psychology. The Ohio University, Columbus (2001)

Perugini, M., Banse, R.: Personality, implicit self-concept and automaticity. European Journal of Personality 21, 257–261 (2007)

Rayner, K.: Eye movements in reading and information processing: 20 years of research. Psychological Bulletin 124(3), 372 (1998)

Rehder, B., Hoffman, A.B.: Eyetracking and selective attention in category learning. Cognitive Psychology 51, 1–41 (2005)

Schipani, S.P.: An evaluation of operator workload, during partially-autonomous vehicle operations: DTIC Document (2003)

Tiller, W.A., McCraty, R., Atkinson, M.: Cardiac coherence: A new, noninvasive measure of autonomic nervous system order. Alternative Therapies in Health and Medicine 2, 52–65 (1996)

Tsai, Y.F., Viirre, E., Strychacz, C., Chase, B., Jung, T.P.: Task performance and eye activity: predicting behavior relating to cognitive workload. Aviation, Space, and Environmental Medicine 78(1), 176–185 (2007)

Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.P.: Eye activity correlates of workload during a visuospatial memory task. Human Factors: The Journal of the Human Factors and Ergonomics Society 43(1), 111–121 (2001)

Wickens, C.D.: Situation awareness and workload in aviation. Psychological Science 11(4), 128–133 (2002)

Wiener, E.L.: Human Factors of Advanced Technology ("Glass Cockpit") Transport Aircraft. (NASA Contractor Report 177528). NASA Ames Research Center, Moffett Field, CA (1989)

Zakowski, S., Hall, M.H., Baum, A.: Stress, stress management, and the immune system. Applied and Preventive Psychology 1, 1–13 (1992)

New Technologies for FRMS

Min Luo[1], Mei Rong[1], Jing Li[1], Wen Dong Hu[2], and ChangHua Sun[3]

[1] No.24A XiBahe Beili, ChaoYang District. China Academy of Civil Aviation Science and Technology, Beijing, 100028, China
[2] The Fourth Military Medical University, Xi'an, China
[3] Aviation Safety Office, Civil Aviation Administration of China, Beijing, China
{luomin,Rongmei,lij}@mail.castc.org.cn, huwendong@163.com, ch_sun@caac.gov.cn

Abstract. Today, Fatigue is one of the hottest issues discussed in civil aviation of the world. However, because the numbers of the contributing factors and the diversity of symptoms, it makes the fatigue monitoring and the fatigue management as a problem. Based on the FRMS framework of Canada, this article will focus on the fatigue monitoring technologies of China, they are the methods on the assessment of work schedule, fatigue symptoms and the actual sleep time. These fatigue risk control measures and tools are designed for the pilots at present, and then it will be gradually developed for ATC and maintenance personnel.

Keywords: Fatigue Risk Management System, Monitoring, Circadian Rhythms, Fatigue Symptoms, Actually Sleep Time.

1 Introduction

Today, fatigue is one of the hottest issues discussed in civil aviation of the world. However, the numbers of the contributing factors and the diversity of symptoms make the fatigue monitoring and the fatigue management as a big problem. As we know, the traditional fatigue risk management based on Flight Time Limitations (FTL) cannot meet the demands of airlines' safety management. At the same time, the new fatigue risk management system (FRMS) has been developed based on SMS. The senior managers of airlines have learned from FRMS that fatigue management should not only consider the length of work hours, but also need to focus on the synchronization of work schedule and circadian rhythms, the actual effective duration of sleep and the severity of mental fatigue under certain work mode.

Just as FRMS advocates, FRMS is an organizational approach, it is also an on-going, adaptive, data-driven and continuous improvement programme based on science and empirical findings for managing fatigue. It aims to find and manage the relevant factors that may lead to fatigue, and enable greater operational flexibility of the organizational management.

China has been promoting the establishment and implementation of SMS for a few years, yet most airlines' safety managers feel that the effect of SMS is not significant. Usually, they often can find some references for the hazard identification and the risk

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 177–183, 2013.
© Springer-Verlag Berlin Heidelberg 2013

evaluation, but cannot find a particularly effective method for the risk management. We like to use the word "no grasp" in Chinese to describe this situation. The reasons are that the risk management measures are too vague to operate or the implementation of the measures affects some stakeholders' interests or causes trouble among different departments. In a word, the biggest challenge of SMS which the China's airlines face is whether there is any available technology or method to manage the different risks. For example, the airlines' managers might know their own pilots are facing fatigue, but the method to reduce the time of flight simply is impossible because it would be detrimental to the profitability of the company. In fact, there are many ways of fatigue risk management and it is the essence of a flexible FRMS to use diverse risk prevention at different levels. Also, as mentioned above, FRMS includes fatigue monitoring technology, and it is a continued program by supporting data analysis.

Based on the FRMS framework of Canada, we will focus on the fatigue monitoring technologies of China, they are the methods on the assessment of work schedule, fatigue symptoms and the actual sleep time. These fatigue risk control measures and tools are designed for the pilots at present, and then it will be gradually developed for ATC and maintenance personnel.

2 Technologies for Fatigue Monitoring

2.1 Pre-flight Fatigue Prediction System

During 2010 to 2011, we carried out a research on the fatigue monitoring model. As figure, we considered the effects on fatigue of circadian rhythms, sleep situation and workload, and designed a fatigue effect factors questionnaire. We collected and analyzed the 60 pilots' test results, and after the validity examination (Cronbach's alpha =0.91), we determined the effectiveness of the model.

In 2012, we developed a network-based fatigue monitoring software named "Pre-Flight Fatigue Prediction System" based on the fatigue predict model. It involve the

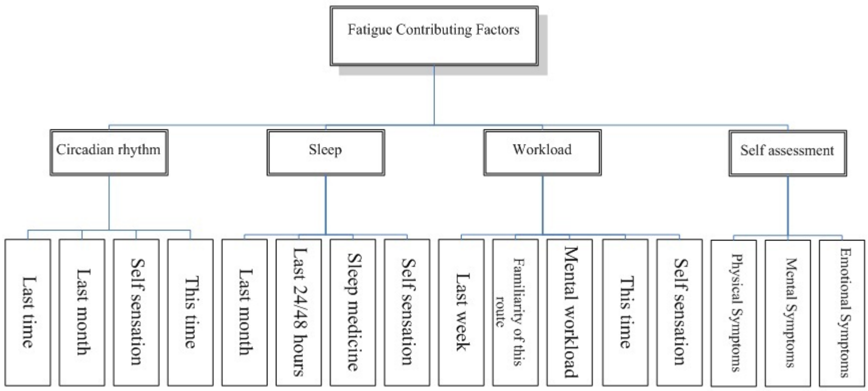


Fig. 1. The Framework of the fatigue monitoring model

main contributing fatigue factors which are circadian rhythm (such as a real flight time and rest time), sleep situation (such as the last month's sleep quality and duration) ,workload(such as the familiarity of this route and the aircraft), and plus the self-assessment factor.

The whole assessment will take about 10-15 minutes to finish. It needs pilot to do it on the given website before starting the flight of the whole day. The assessment results can be obtained immediately, and it can display the individual fatigue level (as figure 2) of each pilot directly and the average fatigue level of the company's pilots for their supervisor. So it can help the manager to decide to change the flight plan in the crew's preparatory stage according to the level of fatigue severity.

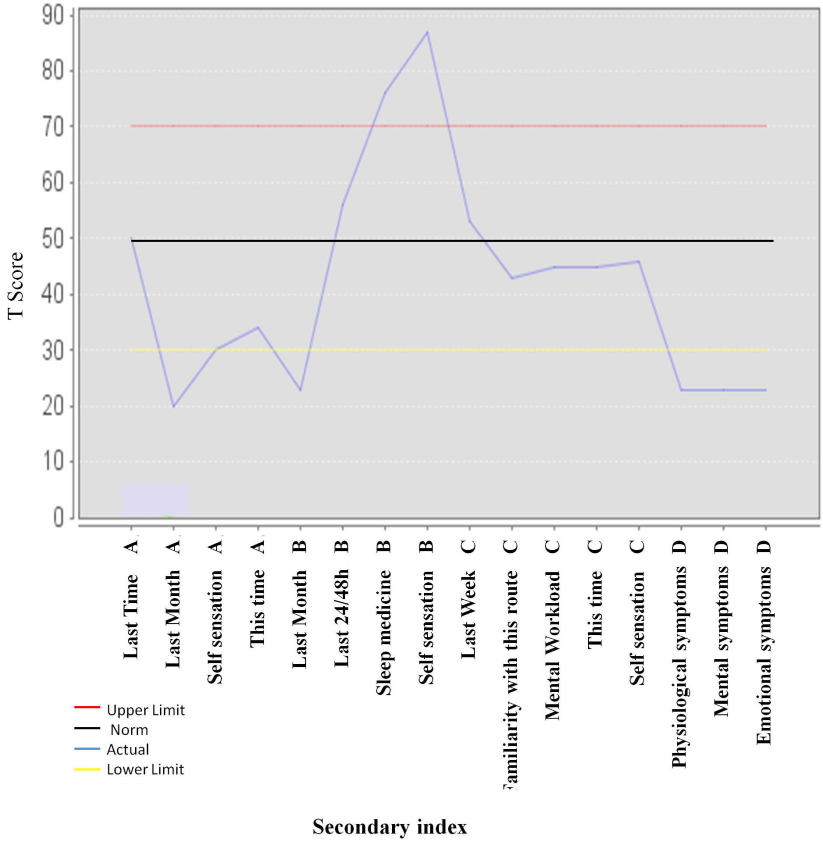


Fig. 2. The display of individual fatigue level

2.2 Orthostatic Detection Platform

During 2008 to 2009, we carried out a study of the relationship between fatigue and postural stability. We assessed the performance of 12 subjective twice at the begging and the ending time of 24 hours sustained wakefulness. After analysis, we indicate that sleep deprivation can arouse a feeling of fatigue and can affect postural stability; hence an objective posture graphic test score may be useful as an indicator of mental fatigue.

Based on this study, we developed an orthostatic detection platform (as figure 3), which consists of stress detection stand, computer and fingerprint input system. The stress detection stand has four separate sets of metal plates. When the individual stand

Fig. 3. Orthostatic Detection Platform

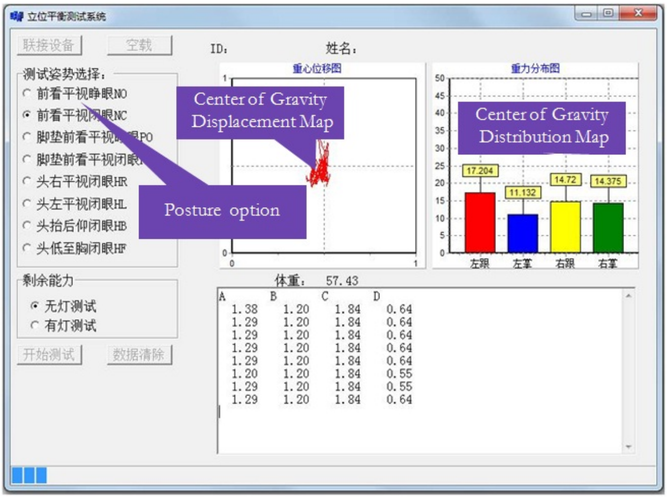


Fig. 4. The display of the assessment

on the pressure platform, making each heel and toes in accordance with the shape of the metal footprints on the corresponding metal plate, the vertical pressure changes on the platform are transmitted into a wave signal to the computer, and then processed by the computer to analyze the individual balance and posture.

According to the test type, the shortest assessment will take about 2.8 minutes and the longest assessment will take about 7minutes. The system can measure the pilot's basic stability, weight distribution and its harmonious degree, then to detect the Individual's vestibular problems, drunkenness, fatigue level(as figure 4), strength loss, and other physiological phenomena in a short time. This platform can be embedded in the pilot's check-in system to use before the flight and also be used to check after each mission.

2.3 Real Time Physiological Monitoring Belted Device

During 2009 to 2011, we developed a real time physiological monitoring belted device named "physiological monitoring watch". It can record and save the physiological and psychological signals such as skin resistance, three degrees of hand freedom, environmental pressure and other data as shown in figure 5.

Fig. 5. Physiological Monitoring Watch

The physiological monitoring watch need to record the change of state of the individual in the period of time, such as fatigue state changes, the duration of actual sleep, hand activities during actual flying and so on. Thus, the pilots need to equip with the watch during a continuous process. It asked the pilots to wear it 10 hours before the flight and take off it until 4 hours to the end of the flight.

We can analyze the pilots' equipped situation, the basic state of the body, the rest and sleep conditions, workload, operating trajectory, flight altitude, flight acceleration and speed indicators based on the basic data obtained by the physiological monitoring watch as shown in figure 6. These results can help the supervisor to assess the real sleep time and physiological state of pilot.

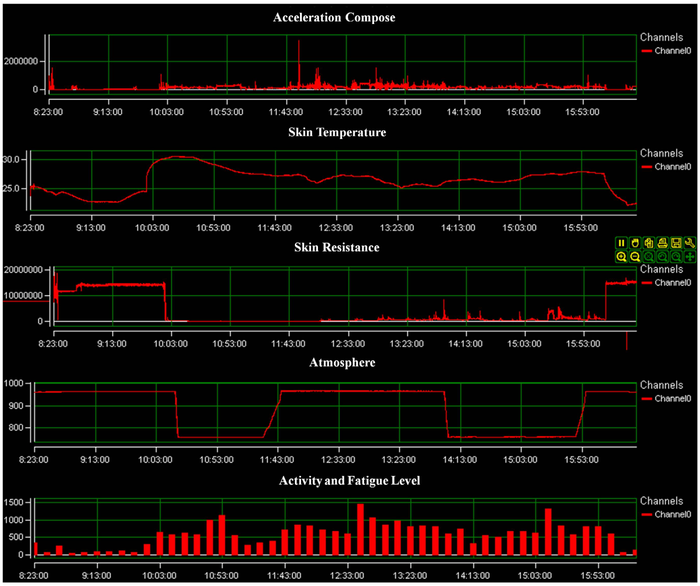


Fig. 6. The result display of Physiological Monitoring Watch

2.4 Comparison

In summary, these three assessment devices have their own advantages and disadvantages, and the obtained data are different. We suggest the supervisor use a

Table 1. Comparison of Three Assessment Devices

	Pre-Flight Fatigue Prediction System	Orthostatic Detection Platform	Real Time Physiological Monitoring Belted Device
Obtained Data	Fatigue level before flight	Current functional level	The changes of fatigue state, the duration of actual sleep, hand activities
Measuring Time	10-15 m	● Shortest: 171 s ● Longest:441s ● Fingerprint Identification:10s ● Balance Test 1:32s ● Balance Test 2:32s ● Balance Test 3:32s ● Heart Rate Measurement:30s ● Heart Rate Variability Measurement:5m ● Body Temperature Measurement:3s ● Reaction Time:32s	Continuous Monitoring, staring from 10hours before the work to 4 hours after the work
Advantages	● Lower cost ● Only once measured	The objective results	● Portable device ● The objective results ● Reflect the dynamic changes and trends of fatigue .
Disadvantages	● Longer time-consuming ● The results are easier by subjective.	● The equipment need to be installed in a fixed place. ● Pilots need to be measured twice before and after flight.	It needs to take a long time to wear the equipment.

combined assessment methods to assess the individual fatigue level accurately if the condition is permitted. The performance of three different assessment devices is listed as table 1.

3 Prospects

In order to assess the fatigue risk for the airlines company, we have formed a detailed fatigue risk control program based on the three monitoring technologies which mentioned above. According to the program, The Civil Aviation Administration of China has planned to carry out a one-year experiment in a selected airline in 2013. This program would verify the effectiveness of the fatigue assessment tools and it will help the supervisor to determine the fatigue degree of the pilots before and after the flight, and take effective fatigue risk control measures for aviation safety.

References

1. ICAO: Fatigue Risk Management Systems Manual for Regulators. DOC9966 (2011)
2. Transports Canada: Advisory Circular on the Development and Implementation of Fatigue Risk Management Systems in the Canadian Aviation Industry. SUR-001 (2008)
3. Van Dongen, H.P.A., Dinges, D.F.: Circadian Rhythms in Fatigue, Alertness and Performance, Principles and Practice of Sleep Medicine, 3rd edn., pp. 391–399
4. Ma, J., Yao, Y.-J.: Effects of Sleep Deprivation on Human Postural Control, Subjective Fatigue Assessment and Psychomotor Performance. The Journal of International Medical Research 37, 151–161 (2009)
5. Lal, S.K., Craig, A.: A critical review of the psychophysiology of driver fatigue. Biol. Psychol. 55, 173–194 (2001)
6. Grandjean, E.: Fatigue in industry. Br. J. Ind. Med. 36, 175–186 (1979)
7. Morad, Y., Azaria, B., Avni, I., et al.: Posturographyas an indicator of fatigue due to sleep deprivation. Aviat. Space Environ. Med. 78, 859–863 (2007)
8. Nakano, T., Araki, K., Michimori, A., et al.: Nineteen-hour variation of postural sway, alertness and rectal temperature during sleep deprivation. Psychiatry Clin. Neurosci. 55, 277–278 (2001)
9. Thomas, M., Sing, H., Belenky, G., et al.: Neuralbasis of alertness and cognitive performance impairments during sleepiness: II. Effects of 24and 72 h of sleep deprivation on wakinghuman regional brain activity. Thalamus Relat Syst. 2, 199–229 (2003)
10. Goldie, P.A., Bach, T.M., Evans, O.M.: Force platform measures for evaluating postural control: reliability and validity. Arch. Phys. Med. Rehabil. 70, 510–517 (1989)

The Glare Evaluation Method Using Digital Camera for Civil Airplane Flight Deck

Zhi Ma[1], Wei Zhang[1], Ye Zhou[1], Jinhai Yu[2], and Baofeng Li[2]

[1] School of Aeronautics, Northwestern Polytechnical University, Xi'an, China
mazhi1982@126.com, weizhangxian@nwpu.edu.cn,
vickyliz1013@gmail.com

[2] Shanghai Aircraft Design and Research Institute, COMAC, Shanghai, China
{yujinhai,libaofeng}@comac.cc

Abstract. Glare is a key factor influencing the visual performance in light conditions of civil airplane flight deck, but it is difficult to directly evaluate the complex glare sources in flight deck, such as non-uniform glare, irregular shape glare and indirect glare using current glare equations. In this paper, a method based on digital camera was proposed to evaluate glare is proposed to evaluate the glare from flight deck. Digital camera's imaging luminance measurement is based on High Dynamic Range (HDR) image processing. The computational procedures to calculate source luminance, background luminance, position index and solid angle of source, to detect the glare sources were developed in Matlab. And then, the desired glare index can be computed. Finally, Daylight Glare Probability (DGP) equation was utilized as an example to evaluate the glare for flight deck in daytime. The results indicate that the proposed method can compute glare index automatically and quickly.

Keywords: glare evaluation, digital camera, flight deck, fish-eye lens, glare index.

1 Introduction

Van Nakagawara[1,2] investigates the relationship between visual impairment from natural sunlight and aviation accidents. The glare was found to be a contributing factor. On the other hand, visual comfort has been a factor what need be taken into account in the flight deck designing. So, Glare is a strong factor to influence human-computer interaction of visual information in light conditions of the flight deck, and it is important to evaluate glare in flight deck. The glare sources of flight deck are very complex, including non-uniform glare sources, irregular shape glare sources and indirect glare sources, etc. And, as it is very difficult to compute luminance of reflective glare sources, these indices cannot evaluate indirect glare, such as reflective glare sources.

In our research work, we can use the glare evaluation method of flight deck based on digital camera to evaluate the glare of flight deck. The process of evaluation is shown in Fig.1.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 184–192, 2013.
© Springer-Verlag Berlin Heidelberg 2013

According to the different visual influence, glare can be generally divided into two types, discomfort glare and disability glare. In general illumination engineering, relative to the disability, the discomfort is more universal phenomenon. The method to control the discomfort glare can solve the disability glare problem. So, the glare we discussed is the discomfort glare in this paper. Several discomfort glare indices have been developed to assess glare from artificial light sources. These include: British Glare Index (BGI)[3,4], CIE Glare Index (CGI)[5], and Unified Glare Rating (UGR)[6]. And, there are several glare indices developed to evaluate the glare caused by daylight. These include: Daylight Glare Index (DGI)[7], New Daylight Glare Index (DGI_N)[8], and Daylight Glare Probability (DGP)[9]. In general, all these equations draw upon the four physical parameters, the luminance of the glare sources, the background luminance, the position index, and the solid angle. Especially, the vertical eye illuminance is also considered as a primary factor influencing glare index in the DGP[3].

In this paper, and DGP proposed by Wienold is utilized as an example to evaluate the glare for daylight in glare quantifiable evaluation of flight deck.

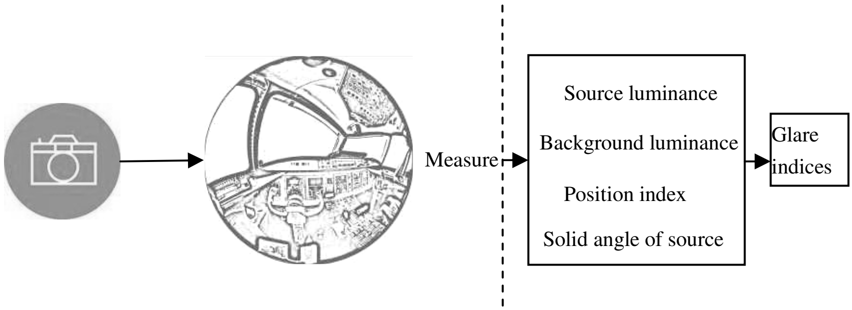


Fig. 1. The process of glare evaluation for flight deck

2 Equipment and Method

2.1 Method

Firstly, after analyzed the luminance of fields in scenes using a luminance meter at the time images are captured by digital camera whose lens is circular fish-eye lens, an empiric formula is developed describing the relationship between the luminance and the color record in the image. And then, the digital camera can measure the luminance of flight deck in 180° visual field. Secondly, glare sources are detected using computational procedures. The last stage computes source luminance, background luminance and position index and solid angle of source, and computes whatever glare index is desired.

Digital Camera's imaging luminance measurement is based on High Dynamic Range (HDR) image processing. Inanici[9] and Tse-ming CHUNG[10] have validated confidence in applying HDR photography as a luminance data acquisition system. The time scale exposure series of Low Dynamic Range (LDR) images to compose a high dynamic range image are used in the method. Then the tristimulus value is calculated from the RGB output by using the color space transformation. Finally the estimated value of luminance is computed by using the calibration coefficient obtained through physical luminance measurement from luminance meter.

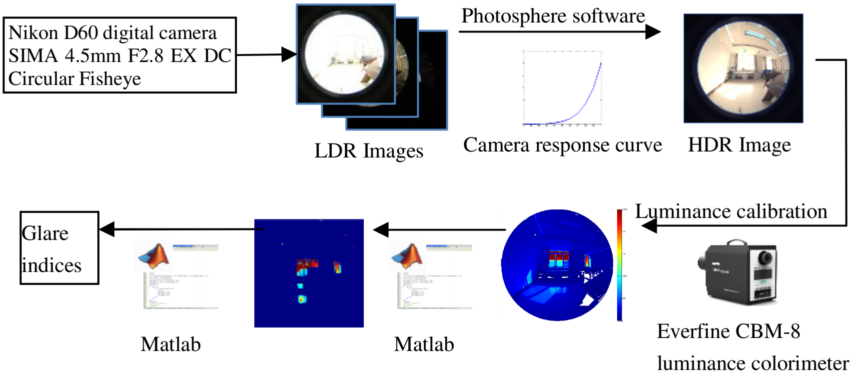


Fig. 2. The evaluation method based on digital camera

2.2 Equipments

The multiple exposure photographs were taken with a Nikon D60 digital camera mounted on a tripod and fitted with a fisheye lens (SIMA 4.5mm F2.8 EX DC Circular Fisheye). The fisheye lens has a focal length of 4.5 mm (equivalent focal distance 4.5×1.5mm) and an angle of view of 180°. Because of advantage to calculate solid angle easily, the fisheye lens we choose used equisolid projection mode. The projection mode of fish-eye lens is shown in Fig.3.

Reference physical measurements were taken with a calibrated luminance meter EVERFINE CBM-8 luminance colorimeter with 2°, 1°, 0.2°, and 0.1° field of view. The luminance of each color of the X-Rite ColorChecker chart measured with a calibrated luminance meter was compared with the luminance value of the corresponding pixels extracted from the HDR image.

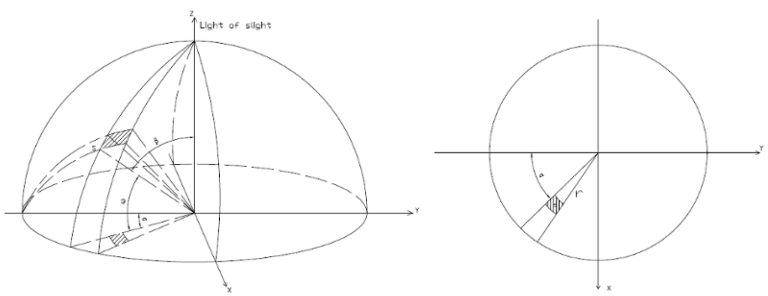


Fig. 3. The projection mode of fish-eye lens

2.3 Softwares

The multiple exposure photographs were processed using the free software Photosphere developed by Greg Ward who invented Radiance RGBE format[11]. All photographs were taken with the camera settings shown in Table 1. It is especially important to fix the white balance to Daylight for achieving consistent colour space transitions. Changing either the aperture size (f-stop) or the shutter speed (exposure time) can vary the exposure values. Shutter speed is a more reliable measure than aperture size. Therefore, exposure variations were achieved with a fixed aperture size (f/8.0), and varying only the shutter speed in manual exposure mode (2 s to 1/200 s).

Table 1. Nikon D60 camera settings

Featrue	Setting	Featrue	Setting
White balance	Daylight	Image	3872×2592
Lens	Fisheye	Aperture size	f/8.0
Exposure Mode	Manual	ISO	100

Photosphere can calculate camera response automatically[9]. Once the camera response curve is determined, Photosphere can fuse any photograph sequence into a HDR image. HDR images can be stored in image formats such as Radiance RGBE, where the pixel values can extend over the luminance span of the human visual system (from 10^{-6} to 10^{8} cd/m^{2})[9].

The computational procedures to calculate the key parameters of glare evaluation were developed in Matlab.

3 Luminance Calculation

For analyzing the HDR images from Photosphere software, computational procedures were implemented (referred to as GetLuminance()). They allow the user to extract and process per-pixel lighting data from the HDR images saved in Radiance RGBE format.

CIE XYZ values for each pixel were quantified from floating point RGB values based on the standard colour space (sRGB) reference primaries, CIE Standard Illuminant D65, and standard CIE Colorimetric Observer with 2° field of view[9]. The transformation process is seeing in Ref .9, and transformation result is as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2127 & 0.7151 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

So, we can calibrate images luminance with a physical luminance measurement of a selected region in the scene. This feature can be applied as a constant (‘*k*’) to the pixel values in an image [9]. Luminance (*L*) is calculated as:

$$L = k \times (0.2127 \times R + 0.7151 \times G + 0.0722 \times B)(cd/m^2) \quad (2)$$

4 Calculating the Key Parameters of Glare Evaluation

4.1 Detection of Glare Sources

Three principal methods were used for automatic detection of glare sources[3]:

1. Calculate the average luminance of the entire picture and count every section as a glare sources that is *x*-times higher than the average luminance;
2. Take a fixed value and count every section as glare sources that is higher than the fixed value;
3. Calculate the average luminance of a given zone (task area) and count every section as glare sources that are x-times higher than the average luminance of this zone.

The first method was implemented in the RADIANCE *findglare* tool. For very bright scenes only few parts or nothing could be detected, although the facade was obviously glare sources. Reducing the x-factor can increase the sensitivity to detect glare sources in a scene, but might lead to "over-detecting" potential glare sources in darker scenes.

The second method, which applied a fixed luminance value as threshold does not take into account eye adaptation. This method was therefore not considered to be a reliable method for lighting scenes with substantial luminance variations.

The last method was used in Wienold's new evaluation method for daylight environment[3]. Each pixel with a luminance value four times higher than the average task-zone luminance was treated as a glare source. This detection sensitivity factor can be changed.

In this paper, we use the first method to detect glare sources as same as RADIANCE. When the threshold is different, the detected result of glare is different. Reducing the threshold can increase the range to glare sources. The detected result is shown in Fig.5.

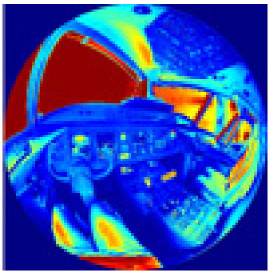


Fig. 4. The mapping of luminance in flight deck

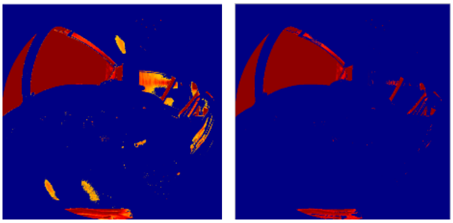


Fig. 5. Glare detected result of the flight deck

4.2 The Key Parameters of Glare Evaluation

In general, the glare evaluation draws upon the four physical parameters:

- The luminance of the glare source: the brighter the source, the higher the index;
- The background luminance: the general field of luminance controlling the adaptation levels of the observer's eye. The brighter the room, the lower the index;
- The position index: the angular displacement of source from the observer's line of sight. The further from the centre of vision, the lower the index;
- The solid angle: the larger the area, the higher the index.

Especially, the vertical eye illuminance is also considered as a primary factor influencing glare index in the DGP[3].

Source Luminance (L_s). The source luminance is the average luminance of all pixel of source. Through computational procedures in Matlab, luminance levels of HDR images from Photosphere can be obtained, see Fig.4. The luminance value of each pixel is obtained by the function "GetSourceLuminance ()". The average luminance of glare sources can be calculated through traversing all pixels of glare sources.

Background Luminance (L_b) . The background luminance is luminance of the field of view, not including source luminance, and called adaption luminance in some study. It can be calculated by

$$L_b = \text{total luminance} - L_s \tag{3}$$

It is obtained by the function "GetBackgroundLuminance ()".

Position Index (*p*). The position index of a source, *P*, is an inverse measure of the relative sensitivity to a glare source at different position throughout the field of view. There are two methods to calculate position index.

The first method is Guth position index, where is given by [6]

$$P = exp[(35.2 - 0.31889\alpha - 1.22e^{-2\alpha/9})10^{-3}\beta + (21 + 0.26667\alpha - 0.002936\alpha^2)10^{-5}\beta^2] \tag{4}$$

Where

α=angle from vertical of the plane containing the source and the line of sight, in degrees,

β=angle between the line of sight and the line from the observer to the source.

Through the new study, Iwata[5] found that sensitivity to glare caused by a source located below the line of vision is greater than sensitivity to glare caused by a source located above the line of vision. Then, a new method to calculate the position index was developed; it could be expressed by this equation [3,4]:

$$\begin{gathered} P = 1 + 0.8 \times R / D \qquad (R < 0.6D) \\ P = 1 + 1.2 \times R / D \qquad (R \geq 0.6D) \\ R^2 = H^2 + Y^2 \end{gathered} \tag{5}$$

Where

R=distance between source and fixation point (m),

D=distance from eye to vertical plane on which a source is located (m),

H=vertical distance between source and fixation point (m),

Y=horizontal distance between source and fixation point (m).

In this paper, Guth position index is used above the line of vision, and the new position index of Iwata is used below the line of vision. The position index of each pixel is obtained by the function "GetPositionIndex()". The mapping of position index is shown in Fig.6.

Based on the equisolid projection mode of fish-eye lens, the meaning of α、β is shown in Fig2.

Where, $\beta=\pi/2-\omega=\pi/2-2\arcsin(r/2f)$, r is the distance between the centre of a circle and the a source, f is the focus of lens.

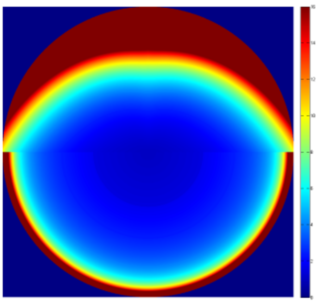


Fig. 6. The mapping of position index[5]

Solid Angle of Source (*Ω*). Solid angle (*Ω*) is a measure of that portion of space about a point bounded by a conic surface whose vertex is at the point. It is defined as the ratio of intercepted surface area of a sphere centered on that point to the square of the sphere's radius. It is expressed in steradians[6].

For the fisheye used equisolid projection mode, solid angle is given by the following equation:

$$\Omega = s/f^2 \tag{6}$$

Where, s is the area of source pixel, f is the focus of the lens.

The solid angle of each pixel is obtained by the function "GetSolidAngle()".

Vertical Eye Illuminance (E_v). The vertical eye illuminance can be calculated by the equation:

$$E_v = \sum_{i,j}^{n} L(i,j) \times \Omega(i,j) \tag{7}$$

Where, L is the luminance of pixel, Ω is the solid angle of pixel.

It is obtained by the function "GetVerticalEyeIlluminance ()".

5 The Glare Evaluation

So, we can evaluate the glare of existent flight deck. Now, DGP is utilized as an example to evaluate the glare for daylight in glare quantifiable evaluation of flight deck shown in Fig.4. The observe position is the design eye position of pilot, and observe direction is the horizontal direction.

The DGP equation is

$$DGP = 5.87\times10^{-5} E_V + 9.18\times10^{-2}\log_{10}(1+\sum_i \frac{L_{S,i}^2 \omega_{S,i}}{E_V^{1.87} P_i^2}) + 0.16 \tag{8}$$

Where DGP is Daylight Glare Probability; E_v is the vertical eye illuminance (lux); L_s is the luminance of source (cd/m^2); ω_s is the solid angle of source; p is the position index of the source.

Aiming to the flight deck scene shown in Fig.4, the calibration factor is 371.471. The detection threshold luminance of the glare sources which is 1273 cd/m^2 is four times higher than the average luminance which is 318.2587 cd/m^2. The vertical eye illuminance E_v is 885.3243 lx. The solid angle of each pixel for glare sources $\omega_{s,i}$ is 8.25×10^{-7} Sr. The position index P_i is shown in Fig.6.

The calculation result of DGP is 0.2483, which express that only 24.83% people feel fidget and discomfort. So, in this condition, the most people are comfort.

6 Conclusions

Several glare indices based on current method cannot be utilized to evaluate non-uniform glare, irregular shape glare and indirect glare correctly. In this paper, using digital camera to evaluate glare offers a simple method. Digital camera's imaging luminance measurement is based on High Dynamic Range (HDR) image processing. The computational procedures to calculate source luminance, background luminance, position index and solid angle of source, to detect the glare sources were developed in

Matlab. After calculating source luminance, background luminance, position index, solid angle of source, the desired glare index can be computed. Finally, Daylight Glare Probability (DGP) equation is utilized as an example to evaluate the glare for flight deck in daytime. The results indicate that the proposed methods are convenient for evaluating the visual performance and comfort in human-computer of flight deck.

Acknowledgments . This research was supported by National Basic Research Program of China (973 Program) (Number: 2010CB734106).

References

1. Van Nakagawara, B., Wood, K.J., Montgomery, R.W.: Natural sunlight and its association to civil aviation accidents. Optometry 75(8), 517–522 (2004)
2. Van Nakagawara, B., Montgomery, R.W., Wood, K.J.: Aircraft accidents and incidents associated with visual effects from bright light exposures during low-light flight operations. Optometry 78, 415–420 (2007)
3. Wienold, J., Christoffersen, J.: Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. Energy and Buildings 38, 743–757 (2006)
4. Iwata, T., Tokura, M.: Position Index for a glare source located below the line of vision. Lighting Research and Technology 29, 172–178 (1997)
5. Lowson, J.C.: Practical application of the Einhorn (CIE) Glare Index formula. Lighting Research and Technology 13(4) (1981)
6. IESNA.: The IESNA Lighting Handbook, 9th edn. IESNA(2000)
7. Chauvel, P., Collins, J.B., Dogniaux, R., Longmore, J.: Glare from windows: current views of the problem. Lighting Research and Technology 14, 31–46 (1982)
8. Nazzal, A.A.: A new evaluation method for daylight discomfort glare. International Journal of Industrial Ergonomics 35, 295–306 (2005)
9. Inanici, M.N.: Evaluation of high dynamic range photography as a luminance data acquisition system. Lighting Research and Technology 3(2), 123–136 (2006)
10. Chung, T.M., Roger, T.H.: Variation of Calibration Factor over Time for High Dynamic Range Photography in a Single Daylit Interior Scene. Journal of Light and Visual Environment 34(2) (2010)
11. Ward, G.: The LogLuv Encoding for full gamut, high dynamic range images. ACM Journal of Graphics Tools 3(1), 15–31 (1998)

Pilot Preferences on Displayed Aircraft Control Variables

Anna Trujillo and Irene Gregory

NASA Langley Research Center, Hampton, VA, USA
{anna.c.trujillo,irene.m.gregory}@nasa.gov

Abstract. The experiments described here explored how pilots want available maneuver authority information transmitted and how this information affects pilots before and after an aircraft failure. The aircraft dynamic variables relative to flight performance were narrowed to energy management variables. A survey was conducted to determine what these variables should be. Survey results indicated that bank angle, vertical velocity, and airspeed were the preferred variables. Based on this, two displays were designed to inform the pilot of available maneuver envelope expressed as bank angle, vertical velocity, and airspeed. These displays were used in an experiment involving control surface failures. Results indicate the displayed limitations in bank angle, vertical velocity, and airspeed were helpful to the pilots during aircraft surface failures. However, the additional information did lead to a slight increase in workload, a small decrease in perceived aircraft flying qualities, and no effect on aircraft situation awareness.

Keywords: Bank Angle, Pitch Angle, Vertical Velocity, Aircraft Speed, Workload, Cooper-Harper Controllability Rating.

1 Introduction

Adaptive control in flight applications has a long and rich history dating back to the 1950s. Currently, adaptive control is considered for highly uncertain, and potentially unpredictable, flight dynamics characteristic of adverse conditions, such as upsets, stall, post-stall high angle-of-attack or damage, induced on transport or high-performance aircraft. Some recent flight experiences of pilot-in-the-loop with an adaptive controller have exhibited unpredicted interactions [1, 2]. In retrospect, this is not surprising once it is realized that there are now two adaptive systems interacting, the adaptive flight control system and the pilot. The pilot controls the attitude of the vehicle and the method of control may change due to varying system parameters. The experiments, described in the paper, explored how pilots want information about available control authority transmitted and how this information affects pilots before, during, and after an aircraft failure.

The aircraft dynamic variables that pilots want to know relative to flight performance were initially narrowed to energy management variables. Further down select of variables to translational velocity, rotations around longitudinal and lateral axes,

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 193–202, 2013.
© Springer-Verlag Berlin Heidelberg 2013

was informed by 5 loss-of-control envelopes proposed in [3]. Because these variables included 0th, 1st, and 2nd order derivatives, a survey was conducted in order to pare the relevant variables down to single parameters in the longitudinal and lateral axes rotations, and for translational velocity related information.

From the survey results, two stand-alone displays were designed to indicate the available maneuverability envelope dependent on the health of the aircraft. During normal aircraft operations, the displays showed full control-authority maneuverability envelope. After a control surface failure, the displays indicated how much maneuver authority was available to the pilot based on the newly calculated safe flight envelope.

The two experiments described in this paper looked at how pilots want information about available maneuver authority transmitted and how this information affects pilots before and after an aircraft control surface failure. Recommendations to improve the displays, based on pilot comments, will be discussed in the Conclusions.

2 Parameter Survey Experiment Procedure

In order to narrow the aircraft attitude parameters related to the flight envelope displayed to the pilot, an initial survey was designed and conducted to acquire in-house pilot preferences. In particular, respondents were asked their opinions about the usefulness of, difficulty in understanding, acceptability of, and the amount of time spent referencing the displayed parameter.

Each respondent indicated his ratings on the displayed parameter. The parameters were bank angle (a scalar), turn rate (a velocity), and turn load limit (an acceleration) for lateral control. For longitudinal control, the parameters were pitch angle, vertical velocity, and vertical acceleration. Finally, for translational velocity, speed and acceleration preferences were the considered parameters. Respondents also provided comments about each parameter.

The displayed parameters indicated how close the aircraft was to the limit where the controller could give near normal control responses. Furthermore, the respondents were told that the aircraft was at top of descent but that the flight envelope had changed due to a control surface failure. An example of the displays respondents were asked to rate is shown in Figure 1.

Three respondents filled out the survey and two test pilots were asked their opinions on the same questions regarding what flight envelope information they thought would be the most beneficial. The survey respondents were an average of 51 years old and had an average of 1517 flight hours and 19 years of flight experience. The two test pilots were an average of 53 years old and had an average of 5850 flight hours and 30 years of flight experience.

3 Parameter Survey Experiment Results

In general, respondents indicated that they wanted bank angle and vehicle speed information (Figure 2). For lateral control, respondents commented that bank angle was

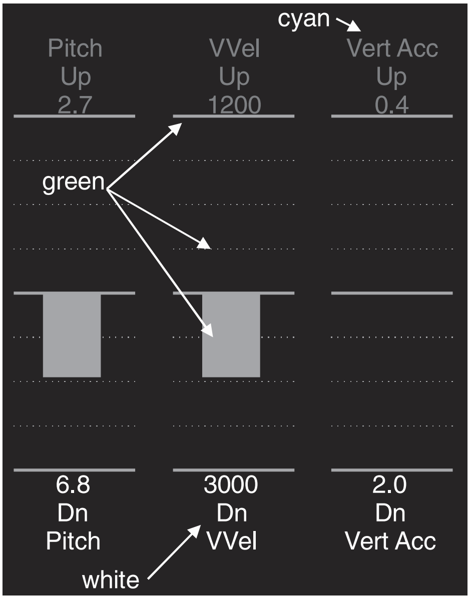


Fig. 1. Longitudinal Parameters and Their Associated Displays for the Survey

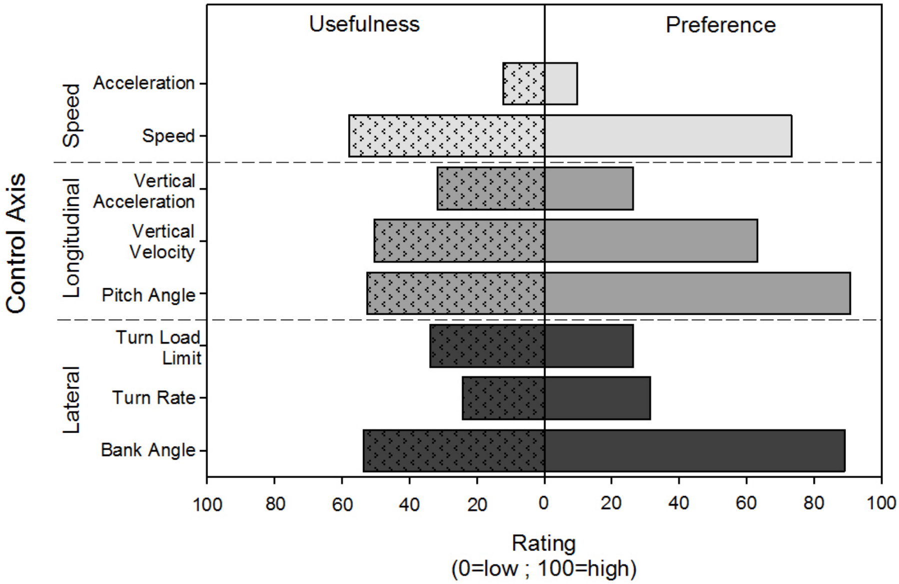


Fig. 2. Survey Control Information Usefulness and Preference

the "most helpful" although "knowing the turn load limit might be useful in windy conditions." This sentiment is reflected in the Usefulness and Preference Lateral ratings in Figure 2. The difference of usefulness between bank angle and turn load ratings in Figure 2. The difference of usefulness between bank angle and turn load limit is smaller than the difference of preference between these two parameters. Hence, turn load limit might be useful during high loading conditions. However, the load factor is a function of bank angle so bank angle was chosen as the preferred lateral control parameter to display for information related to the flight envelope.

For translational velocity, all respondents indicated that Speed was the "most useful and critical to controlling the aircraft."

As for longitudinal control, respondents were fairly evenly divided, especially for usefulness, between pitch angle and vertical velocity as indicated in the Longitudinal ratings in Figure 2. Several respondents said that pitch angle is "important for go-arounds." Others indicated that vertical velocity is "extremely useful, especially when establishing ascent or descent profiles." Thus, vertical velocity was chosen to be displayed rather than pitch angle because the longitudinal-related comments from the survey respondents slightly favored vertical velocity and conversations with the two test pilots indicated that vertical velocity was more useful for controlling aircraft in a steady-state climb or a descent.

4 Maneuver Authority Display Experiment Procedure

The maneuver authority display (MAD) experiment looked at whether an adaptive controller helps pilots during control surface failures and whether an additional display indicating how close the vehicle is to reaching the limit of safe maneuver authority was helpful during and after control surface failures. The variables shown on the new display, which informed the subject of the available maneuverability envelope, were the ones identified by the survey and flight-test pilots: bank angle, vertical velocity, and aircraft speed. These new displays were then used in a human-in-the-loop experiment to look at their effects on pilot performance in the presence of aircraft control surface failures, specifically in the cruise phase of flight while initiating a climb, a descent, or a heading change maneuver. These maneuvers were indicated on the primary flight display (PFD) via the flight director.

The physical setup of the simulator incorporated an out-the-window view in the upper center 30-inch diagonal screen and four 20-inch touchscreens below the out-the-widow screen. The middle-left touchscreen depicted the PFD and the middle-right touchscreen depicted the engine indication display (EID). The far-left touchscreen contained the available maneuverability envelope display, when present, and the far-right touchscreen displayed the after run questions. Subjects flew the aircraft with a right-handed joystick.

The two displays tested were the dial display (Figure 3) and the circle display (Figure 4). In both displays, the information shown was the same but the format was different. In each display, a green wedge filled in from zero the percentage of available safe flight envelope used in the maneuver. For example, for vertical velocity (VVel) in Figure 3, the aircraft is descending at about 80% of 3000 feet per minute. When the available control authority changed due to a control surface failure, the displayed number went from white to cyan in color and the value changed to the new

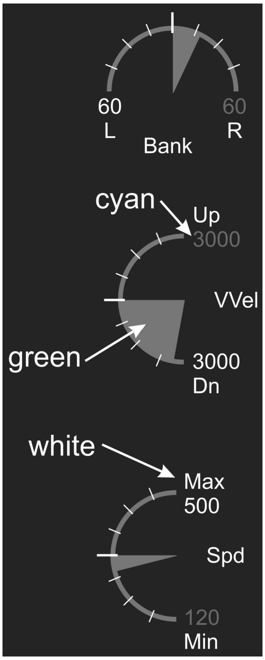


Fig. 3. Dial Display

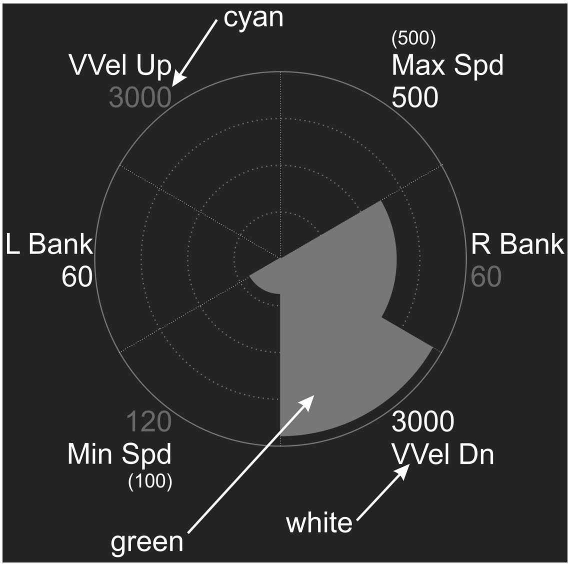


Fig. 4. Circle Display

limit indicating available maneuver authority. For example, for minimum speed (Min Spd) in Figure 4, the aircraft's safe minimum speed is now 120 knots as indicated by the cyan number. For this experiment, the available safe maneuver envelope was predefined for each scenario rather than calculated in real time during the subjects' simulation data runs.

Each subject performed several runs without the new displays (none), then several runs with a display (randomly either the circle or dial display), and finally several runs with the remaining display. During each data run, flight technical data were recorded in addition to a NASA-TLX workload rating [4, 5], Cooper-Harper (CH) handling qualities rating [6-8], and situation awareness questions. After all the data runs, subjects filled out a final questionnaire asking them about their preferences on the information in the new displays showing how close the vehicle is to reaching the limit of safe maneuver authority and the new displays themselves.

The seventeen subjects in the MAD experiment were an average of 48±10 years old with the youngest 29 years old and the oldest 61 years old. All of them were airline transport rated pilots with an average of 26±11 years of flight experience (minimum flight experience = 7 years and maximum flight experience = 45 years) and an average of 10,706±7164 hours of flight experience (minimum flight hours = 2,100 and maximum flight hours = 23,400).

5 Maneuver Authority Display Experiment Results

Overall, the MAD experiment results for new experimental display preferences mimicked the parameter survey results. Unless noted otherwise, display in the following sections refers to the new experimental display indicating how close the vehicle is to reaching the limit of safe maneuver authority and mentioned parameters are associated with the new experimental display.

5.1 Practicality

The vast majority of the subjects thought that the experimental display was practical. In fact, 13 subjects said the display was practical compared to 4 subjects who reported that the display was impractical. Five of the 17 subjects commented that the display provided "relevant additional information" although one subject did mention that the information was "not relevant during normal conditions" and another said that it was "not enough information during a failure."

5.2 Preferred Content

Subjects in the MAD experiment indicated that they preferred to have pitch angle, vertical velocity and vehicle speed flight envelope information available to them on the experimental display (Figure 5). Once again for longitudinal control, subjects slightly preferred vertical velocity over pitch angle. Subjects also said that the additional roll flight envelope information (10 subjects) was the most desirable over pitch (5 subjects) and yaw (1 subject) flight envelope information.

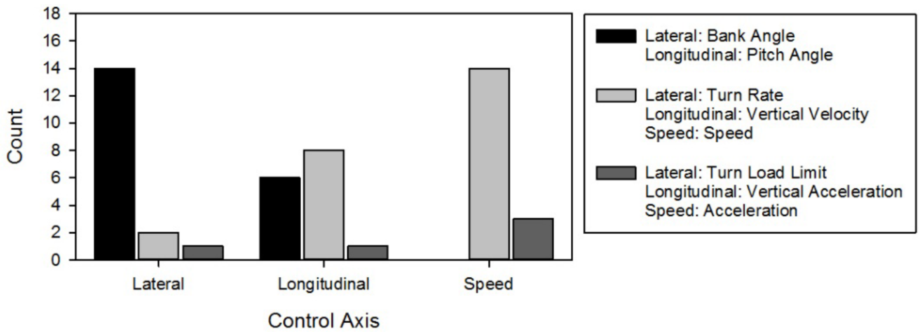


Fig. 5. Maneuver Information Preferences for the MAD Experiment

For lateral control, subjects commented that bank angle was "useful" (5 subjects) on the experimental display and a "well-known and understood measure" (5 subjects). A couple of subjects stated the bank angle was "useful for preventing upsets" but two other subjects said that turn rate information could be "useful for upset prevention."

Subject comments for longitudinal control on the display were split between pitch angle and turn rate. Three subjects mentioned that pitch angle was "useful" while one subject mentioned that vertical velocity was "useful." For how "well known and understood" a measure was, two subjects said this was true for pitch angle and three subjects indicated that this was true for vertical velocity. Also, regarding "upset prevention," two subjects said that pitch angle was good for this while three subjects said vertical velocity was good for upset prevention. Other comments regarding vertical velocity indicated that it was "required knowledge for safe flight" and it typically is used during "normal operations and for approaches."

As for translational velocity, subjects overwhelmingly preferred speed over acceleration on the experimental display. Five subjects mentioned that it is a "well-known and understood measure" and four subjects said it was "useful." Also, two subjects stated that speed was "useful for upset prevention." As for acceleration, two subjects said they did not need this displayed because "you can feel changes in acceleration."

5.3 Situation Awareness

In general, the added displays did not adversely affect a subject's situation awareness of the status of the aircraft with respect to airspeed, altitude, heading, and the aircraft system status. Hence, subjects were able to maintain their general situation awareness about the aircraft even with the additional displays present.

5.4 Workload

Mental demand did increase significantly ($F_{(2, 764)} = 3.5$; $p \leq 0.03$). This may have been an artifact of display design rather that information content because only the circle display was significantly different from the none case. The overall workload of subjects increased slightly, but not statistically significantly, with the additional display (Figure 6). Hence, maneuver envelope information of bank angle, vertical velocity, and vehicle speed did not appreciably increase workload. This may indicate that either the information was the right information on the display or that the display was not attended to by the subjects.

5.5 Cooper-Harper Controllability Rating

Subjects also reported a slight increase in CH ratings with the maneuver envelope information although this increase was not statistically significant (Figure 7). As with workload, this information did not significantly hinder the subjects' ability to control the aircraft. It does appear that the displays did worsen handling qualities slightly as seen by the shift of the number of CH ratings at 1 and 2 for no display to a CH rating of 3 for both the circle and dial displays. This shift may be due to additional information that must be attended to and understood while maintaining control of the aircraft.

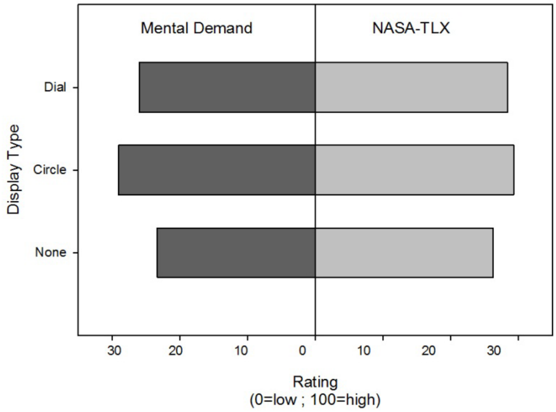


Fig. 6. Mental Demand and Overall Workload Ratings for the CAD Experiment Displays

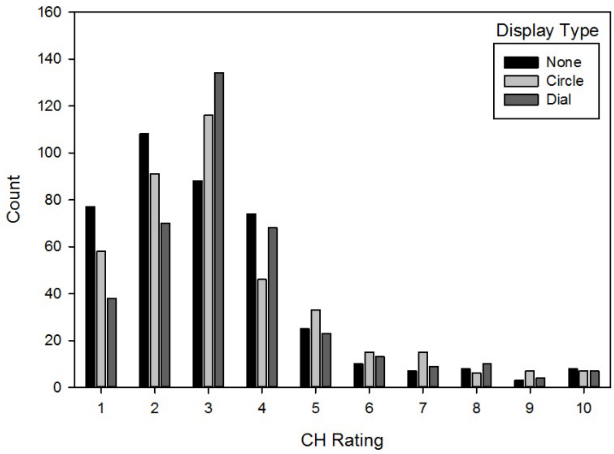


Fig. 7. Count of Cooper-Harper Controllability Ratings by Display Type

5.6 Pitch and Roll Error

Pitch and roll error were calculated by taking the difference between the actual aircraft pitch and roll angle and the commanded pitch and roll angle. This was further broken down by the error before and after a control surface failure.

The pitch and roll errors were not significant except for the pitch error before a failure by display type ($F_{2,\ 1166} = 3.2$; $p \le 0.04$). As can be seen in Figure 8, the pitch error typically decreased after the failure for all display types with the biggest decrease for the circle display although the circle display's error was the highest overall for the displays. For the roll error, the error increased after the failure although the increase was the least for the circle display. Therefore, the vertical velocity information for pitch control helps after the control surface failure but the bank angle information after the control surface failure for roll control does not.

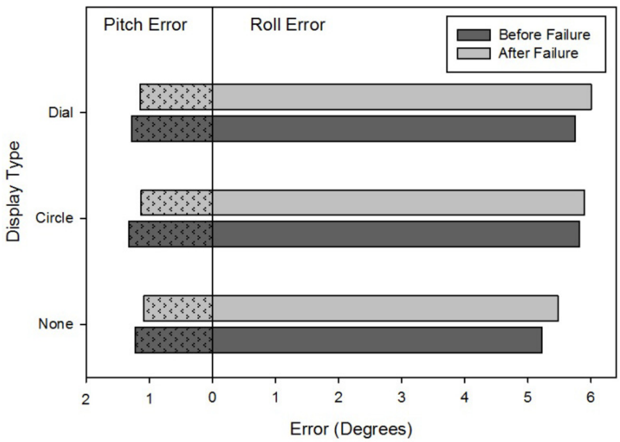


Fig. 8. Pitch and Roll Error by Display Type Before and After the Control Surface Failure

The result indicating decreasing pitch error after a failure may be fortuitous because three of the five loss-of-control envelopes include the longitudinal axis [3] so attending to maneuvering safely within the flight envelope in the longitudinal direction may aid the most in preventing upsets. Finally, the circle display had the largest decrease in pitch error and the smallest increase in roll error indicating that the integrated display showing large wedges was easier to process than the dial display.

6 Conclusions

To maintain control of an aircraft before, during, and after control failures, pilots want to know the limits of the aircraft's new maneuver envelope. The two experiments described above began to look at what maneuver envelope information pilots need in order to safely control the aircraft after a control surface failure.

The survey results indicated the pilots would want to know available safe bank angle, vehicle speed, and either pitch angle or vertical velocity during and after a control surface failure. For lateral and translation velocity control, respondents commented that bank angle for the former and speed for the latter were the most helpful. As for longitudinal control, respondents were fairly evenly divided between pitch angle and vertical velocity. Several respondents said that pitch angle is important for go-arounds. However, vertical velocity was used because others indicated that vertical velocity is extremely useful, especially when establishing ascent or descent profile and the survey respondents slightly favored vertical velocity.

From the survey results, the MAD experiment looked at two new displays showing how close the aircraft was to reaching the limit of safe control of bank angle, vertical velocity, and aircraft speed. The results from this experiment indicated that pilots did in fact want this information. Specifically, pilots said that they did want to know bank angle, vertical velocity, and aircraft speed maneuver envelope information. Furthermore, this additional information did not appreciably increase workload, adversely affect situation awareness, or affect vehicle controllability negatively. The CH handling qualities ratings may have increased slightly, however, due to either wrong information provided or because pilots had another display to look at during a failure. Lastly, the new additional display did not negatively affect the ability of the pilots to control the aircraft before or after a control surface failure.

Therefore, bank angle, vertical velocity, and aircraft speed are acceptable maneuver envelope information to display to pilots especially after a control surface failure which may alter the safe maneuver envelope. Exactly how to show this information is being further investigated. Several subjects commented that they wanted the information integrated into the primary flight display; in particular, into the horizon display for bank angle information, the vertical speed indicator for vertical velocity, and the speed tape for the speed envelope information. Other subjects indicated that they preferred the separate display because it was easier to see changes when they occurred although some subjects said they only wanted the display present when a failure occurred. Lastly, with regards to the display of the information, a few subjects did suggest changing vertical velocity to pitch angle during takeoffs and go-arounds. Whether this information switching would increase the chance of “mode” confusion should be fully investigated before incorporating this into a future display indicating available maneuver envelope of the aircraft.

References

1. Bosworth, J.T., Williams-Hayes, P.S.: Flight Test Results from the NF-15B Intelligent Flight Control System (IFCS) Project with Adaptation to a Simulated Stabilator Failure. In: AIAA Infotech@Aerospace 2007 Conference and Exhibit, Rohnert Park, CA, vol. AIAA-2007-2818 (2007)
2. Page, A.B., Meloney, E.D., Monaco, J.F.: Flight Testing of a Retrofit Reconfigurable Control Law Architecture Using an F/A-18C. In: AIAA (ed.) AIAA Guidance, Navigation, and Control Conference and Exhibit, vol. AIAA 2006-5062, p. 20. AIAA, Keystone (2006)
3. Wilborn, J.E.: An Analysis of Commercial Transport Aircraft Loss-of-Control Accidents and Intervention Stategies (2001)
4. Trujillo, A.C.: How Electronic Questionnaire Formats Affect Scaled Responses. In: 2009 (15th) International Symposium on Aviation Psychology, Dayton, OH (2009)
5. Hart, S.G., Staveland, L.E.: Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P.S., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. Elsevier Science Publishers B. V., Amsterdam (1988)
6. Cooper, G.E., Harper, R.P.: The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities. AGARD (1969)
7. Harper, R.P., Cooper, G.E.: Handling Qualities and Pilot Evaluation (Wright Brothers Lecture in Aeronautics). Journal of Guidance, Control, and Dynamics 9, 515–529 (1986)
8. Trujillo, A.: Paper to Electronic Questionnaires: Effects on Structured Questionnaire Forms. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part I, HCII 2009. LNCS, vol. 5610, pp. 362–371. Springer, Heidelberg (2009)

A Layered Multi-dimensional Description of Pilot's Workload Based on Objective Measures

Zhen Wang and Shan Fu

School of Aeronautics and Astronautics, Shanghai Jiao Tong University,
Shanghai, 200240, P.R. China
b2wz@sjtu.edu.cn

Abstract. Human factors have an important impact on aviation safety. The evaluation of pilot's workload is one of the most noteworthy human factors issues. After a brief overview of workload measurement techniques, a layered multi-dimensional description of workload is proposed, and the method is based on multiple objective measures. Heart rate, respiration, eye movements, control inputs and flight data are recorded in a simulated flight. The sensitivity and diagnosticity of several psychophysiological measurements are analyzed. Finally, a multi-dimensional pattern is constructed using the proposed method. The pattern can give a detailed description of pilot's workload throughout the flight.

Keywords: pilot's workload, layered multi-dimensional description, objective measures, simulated flight, workload pattern.

1 Introduction

Safety has always been the most considerable problem of civil aviation. After many years of effort, the reliability of the aircraft itself has been greatly developed. However, statistics shows that more than 60% of modern aircraft accidents are caused by human factors. In this situation, it's important to carry out human factors research in the cockpit to acquire information about human capability and limitations. Using this information in design, training and certification can increase the safety, comfort and efficiency of the aircraft [1].

One of the most important aspects of human factors research is to evaluate operator's workload. Universal definition of workload doesn't exist, but there is general agreement that operator's workload is a multi-dimensional construct. Megaw (2005) pointed out that workload is the interaction of task factors, operator response, operator performance and additional stressors [2].

The state of art workload measurement techniques are generally organized into three broad categories: task performance measures, subjective rating scales and psychophysiological measures [2-6].

Task performance measures evaluate operator's workload by considering the completion of the task. Task performance measures can be classified into two types: primary task performance measures and secondary task performance measures. Primary task performance measures overlook the operator's initiatives and effort, thus it's of poor validity. Secondary task performance measures require great care during the

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 203–211, 2013.
© Springer-Verlag Berlin Heidelberg 2013

design of the secondary task in order to ensure that the secondary task compete for the same resource as the primary task. And secondary task performance measures are intrusive to primary task performance. Furthermore, task performance measures are only sensitive to the global changes of workload, their diagnosticity are limited.

Subjective rating scales use the operator's experience to evaluate workload. Some of the subjective rating scales such as NASA-TLX and SWAT have been widely used in different fields. Multi-dimensional subjective rating scales have good diagnosticity and are easy to apply. But subjective rating scales can't evaluate workload in real time. They are always correlated with task performance. Sometimes participants may forget the high workload they experienced. And the weighting procedures of these techniques are always cumbersome.

Psychophysiological measures use the psychophysiological reaction of the operator to reflect the workload imposed on them. Frequently used psychophysiological measurements include ECG, EOG, EEG, EMG, etc. These measures can evaluate workload in real time. Many measuring equipment are now portable and wireless, thus they would bring little intrusiveness and can be accepted by the participants. But the relation between psychophysiological measurements and workload are not clear. Some measurements have poor diagnosticity and may be easily affected by other factors e.g. ambience changes. And the equipment is always expensive.

In this paper, a workload description method is proposed based on objective measures. These measurements are categorized into different aspect. And workload is represented in a multi-dimensional way. It aims at providing a real time, valid and diagnostic tool to evaluate pilot's workload.

2 Proposed Workload Description

In order to cope with the problems of the existing workload measures, the following methods are proposed to describe pilot's workload.

2.1 Hierarchical Relationship

Since the relationship between psychophysiological measurements and workload is indirect. It can be assumed that there are some intermediate parameters between psychophysiological measurements and workload. These intermediate parameters could be explained as different dimensions of workload.

2.2 Multi-dimensional Representation

In the proposed method, it is assumed that workload can be decomposed into four primary dimensions. They are:

- "Cognitive activity", which represents pilot's activities in perception, information processing, decision making, etc.
- "Control activity", which represents pilot's input activities in the cockpit.
- "Effort and fatigue", which represents pilot's initiative effort in difficult situation.
- "Flight performance", which represents the result of the pilot's control.

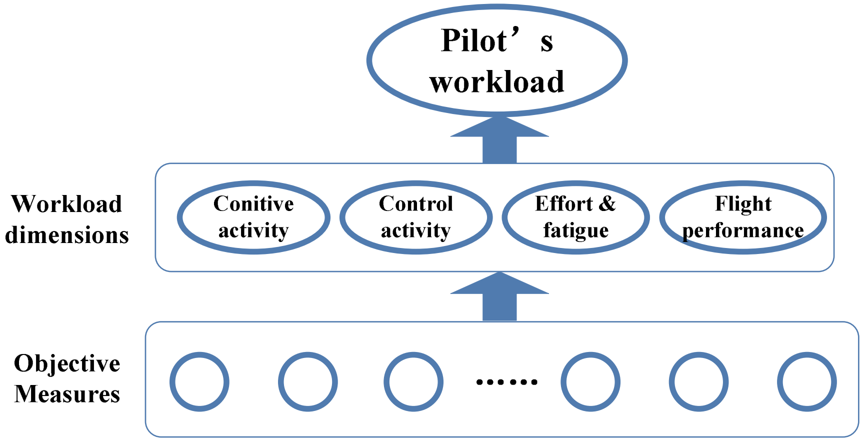


Fig. 1. Layered multi-dimensional description of pilot's workload

2.3 Integrated Description

Each primary dimension of workload can be integrated described by several objective measurements. Integrate different measurements which are sensitive to the same dimension can make them complement each other and can also reduce the interference of environment to single measurement.

By the above methods (i.e. hierarchical relationship, multi-dimensional representation and integrated description), the workload can be described by the structure as illustrated in Fig. 1.

3 Experiment

To empirically investigate the proposed workload description method, a simulated flight experiment has been conducted. During the experiment, several parameters have been objectively measured.

3.1 Participants and Apparatus

Eight students volunteered to take part in the flight simulation. They are aged from 23 to 29, with the average of 26. All of them are from the school of aeronautics and astronautics. They have basic knowledge about aviation and have been trained in simulated flight.

The experiment is carried out on a simulator with high fidelity. The simulator consists of two parts, the outside view and the cockpit. The outside view is simulated and projected on a cylindrical screen which has a diameter of about 8 meters. In the cockpit, the arrangements are referred to Boeing 777-200ER. There are control instruments and display instruments in the cockpit. The control instruments include the yoke, throttle, rudder pedal, flaps, landing gear, CDU and MCP. The display instruments include PFD, ND, EICAS, etc.

3.2 Procedure

In the experiment, each participant is asked to fly a complete flight task 5 times. The flight task consists of take-off phase, cruise phase, approach and landing phase. The aircraft would take off from KSJC 30R. After passing 5 way points, it would be landed at KSFO 28R. The flight environment is simulated as in summer, at noon, sunny and no wind.

During the experiment, several parameters are recorded. Altitude, airspeed, position and instrument inputs are recorded by flight recorder. Blink, saccade, fixation and pupil diameter are recorded by SmartEye eye tracker. Heart rate, body temperature and respiration are recorded by BioHarness physiology monitoring system.

There is a training procedure before each experiment. Participant is asked to relax and fly a free flight. In the meantime, psychophysiological parameters, control activities are measured to acquire the baseline.

4 Results

Several typical tracings of the psychophysiological measurements are depicted in Fig. 2.

5 Discussion

5.1 Psychophysiological Reaction

The experimental results in Fig. 2 show that several psychophysiological measurements vary with the different phases of the flight.

Blink latency is longer during the take-off phase and the approach and landing phase. Realizing that during the take-off phase, the participants have to check the status of the systems carefully; during the approach and landing phase, the participants are asked to achieve a visual approach and landing, they have to frequently acquire the information about the flight, in order to adjust the flight path. Thus, it can be informed that when pilots pay more effort to concentrate on visual information, their blink latency may become longer [8,10].

Fixation duration can reflect the time for a participant to acquire information. During the beginning of the take-off phase the fixation duration is longer, this may because that to check the systems status is very important, the participants pay more attention. During the last part of the approach and landing phase, pilots have to continuously monitor several critical parameters, e.g. altitude, airspeed, heading, attitude, etc. On the contrary, during the cruise phase, autopilot is engaged. Pilots are relatively more relaxed. They don't have to read a gauge accurately, and they glance at the whole cockpit and outside view. Thus, longer fixation duration can inform more cognitive activity.

Heart rate and respiration rate are higher when pilots have more control activities. This may be explained as that more activities require more energy. Researches indicate

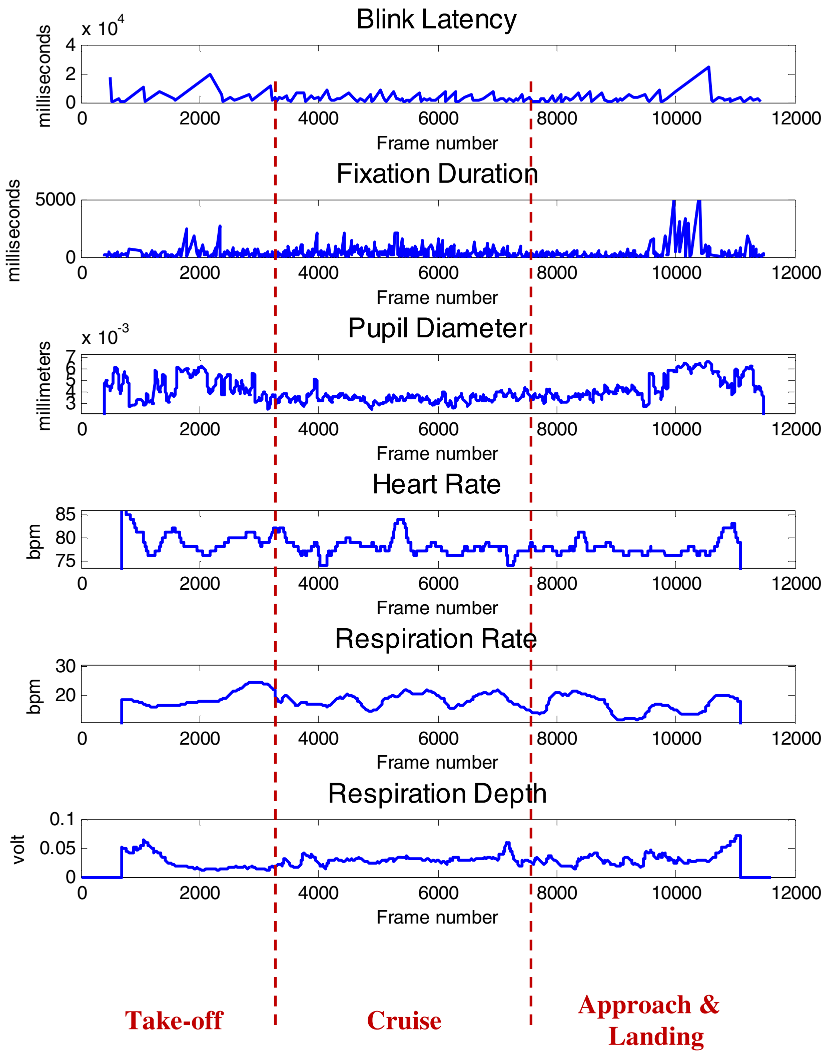


Fig. 2. Changes of psychophysiological measures throughout the flight

that heart rate, respiration and pupil diameter, etc. are associated with activities of the sympathetic nervous system [7-10]. When the task becomes harder or the operator feels more stressed, sympathetic nervous system accommodates heart and other organs to work harder.

Respiration depth seems to be sensitive to some particular events. When the participant discovers the performance falls, he would implement some adjustments. If these adjustments have good effects, the stress would be released and there often follows a deep breath.

5.2 Integrated Description of Workload Dimensions

The psychophysiological measurements and other objective measurements can be classified into the four primary dimensions according to their diagnosticity. In each dimension, the following equations can be used to integrate the measurements.

In "cognitive activity" dimension, measurements can be integrated by the following equation:

$$cognition\ activity = C_1 \cdot \frac{t_{cgn}}{T_{cgn}} + C_2 \cdot \frac{t_{fix}}{T_{fix}} \quad (1)$$

Where

- t_{cgn} is the cognitive time for a participant to perceive, process the information and make the decision before making a control. t_{fix} is the duration of each fixation. T_{cgn} and T_{fix} are the baseline of cognitive time and fixation duration respectively, which are calculated from the training data.
- C_1 and C_2 are the weights to express the different contributions of cognitive time and fixation duration respectively. They are empirically set by the reliability of the measurements. In this paper, C_1=0.6, C_2=0.4.

In "control activity" dimension, measurements can be integrated by the following equation:

$$control\ activity = K_1 \cdot \frac{t_{ctrl}}{T_{ctrl}} + K_2 \cdot \frac{m_{ctrl}}{M_{ctrl}} + K_3 \cdot \frac{f_{ctrl}}{F_{ctrl}} \quad (2)$$

Where

- t_{ctrl} is the time for a participant to achieve a input by a control instrument. m_{ctrl} is the magnitude of each input, represented by percentage of the instrument's extent. f_{ctrl} is the frequency of the control activities. T_{ctrl}, M_{ctrl} and F_{ctrl} are the baseline of control time, control magnitude and control frequency respectively, which are calculated from the training data.
- K_1, K_2 and K_3 are the weights to express the different contributions of control time, control magnitude and control frequency respectively. They are empirically set to balance the different scale of the components. In this paper, K_1=0.4, K_2=0.4, K_3=0.2.

In "effort and fatigue" dimension, measurements can be integrated by the following equation:

$$effort\ \&\ fatigue = E_1 \cdot \frac{t_{bl}}{T_{bl}} + E_2 \cdot \frac{hr}{HR} + E_3 \cdot \frac{pdiam}{PDiam} + E_4 \cdot \frac{res}{RES} \quad (3)$$

Where

- t_{bl} is the blink latency. *hr* is the heart rate. *pdiam* is the pupil diameter. *res* is the respiration rate. T_{bl}, *HR*, *PDiam*, *RES* are the baseline of blink latency, heart rate, pupil diameter and respiration rate respectively. They are calculated from the training data.

- E_1, E_2, E_3, E_4 are the weights to express the different contributions of blink latency, heart rate, pupil diameter and respiration rate respectively. They are empirically set by the reliability of the data. In this paper, E_1=0.3, E_2=0.4, E_3=0.2, E_4=0.2.

In "flight performance" dimension, measurements can be integrated by the following equation:

$$performance = P_1 \cdot e^{-(|\Delta s| \times 10)} + P_2 \cdot e^{-(|a| \times 0.5)} \quad (4)$$

Where

- Δs is the deviation of the flight path. a is the acceleration of the aircraft.
- P_1 and P_2 are the weights to express the different contributions of flight path deviation and acceleration respectively. They are empirically set to balance the scale of the components. In this paper, P_1=0.6, P_2=0.4.

5.3 Workload Pattern

After working out the values in the primary dimensions, another 6 secondary dimensions can be constructed. These secondary dimensions which represent the correlations between any of the two primary dimensions include "cognition-performance correlation", "control-performance correlation", "cognition-effort correlation", "control-effort correlation", "effort-performance correlation" and "cognition-control correlation". Note that, for the sake of simplicity and in order to display the multi-dimensional pattern into a flat plane, the "effort-performance correlation" and "cognition-control correlation" dimensions are omitted.

Each dimension has been scaled into the extent of 0 to 20. Then an 8-dimension workload pattern is formed. In almost every moment of the flight, the workload pattern can be provided. This can give a very detailed evaluation of pilot's workload throughout the whole flight.

5.4 Pattern Diagnosis

The pattern's diagnosis is discussed by several typical cases.

The pattern in Fig. 3 (a) is a frequently appeared pattern during flight. It illustrates that the primary task of the pilot is monitoring the system. There are very few control activities, and flight performance is mainly influenced by the monitoring activities.

Pattern in Fig. 3 (b) illustrates that the pilot doesn't have many cognitive activities or control activities. After paying a little effort, a good performance can be achieved. It can inform that the pilot is currently performing in a well-designed flight deck system.

Pattern in Fig. 3 (c) shows that the pilot is now paying a lot of efforts and feels tired. This is mainly caused by heavy cognitive activities. In the meantime, few controls are made. It can inform that the pilot is trying hard to estimate the flight situation.

Pattern in Fig. 3 (d) shows a bad situation. The pilot is trying hard to estimate the conditions and there are many control activities, but the flight performance is still worse.

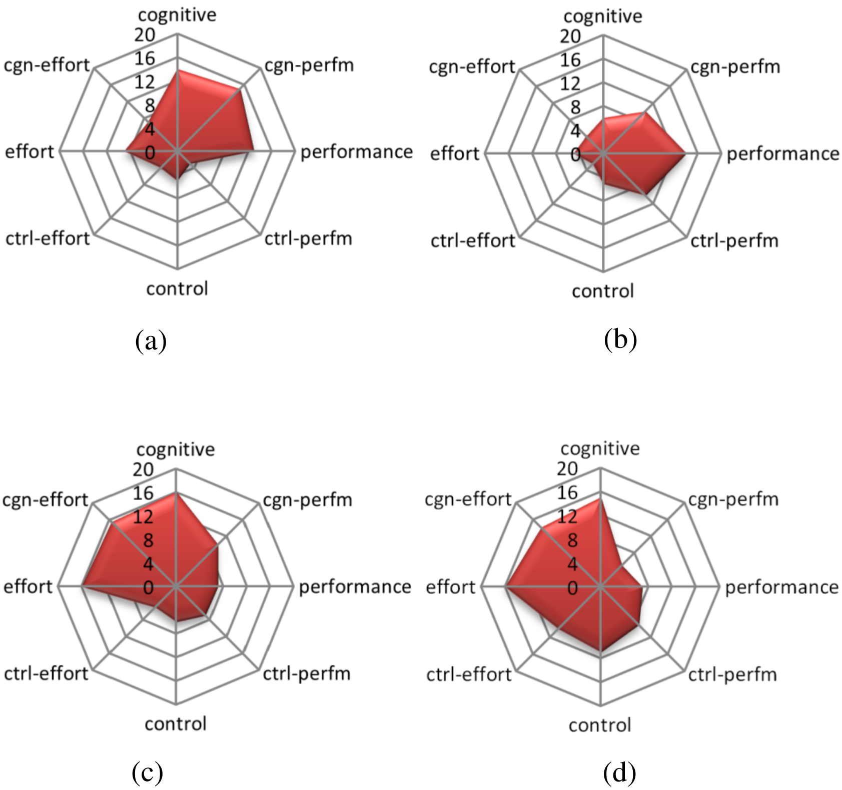


Fig. 3. Workload pattern cases

6 Conclusions and Future Work

In this study, several objective parameters are measured in a simulated flight. The experimental results show that blink latency, fixation time, pupil diameter, heart rate, respiration rate are sensitive to different aspects of workload. A layered multi-dimensional description of workload is proposed and the objective measurements are classified and integrated into four aspect of workload i.e. cognitive activity, control activity, effort and fatigue, flight performance. Furthermore, an 8-dimension workload pattern is formed. It can give a detailed and diagnostic evaluation of workload throughout whole flight. Yet, in the proposed method, several parameters are determined empirically. And the participants are not real pilots. The reliability of this method will be systematically examined in the future.

Acknowledgement. This research work was supported by National Basic Research Program of China-(973 Program No. 2010CB734103) and research program of Shanghai Jiao Tong University for innovation of post graduates (985 Program) with No. Z-413-006 under Grant TS0220741301.

References

1. FAA System Safety Handbook, Chapter 17: Human Factors Principles & Practices (2003)
2. Stanton, N.A., Salmon, P.M., Walker, G.H., Baber, C., Jenkins, D.P.: Human Factors Methods: A Practical Guide for Engineering and Design. Ashgate (2005)
3. Farmer, E., Brownson, A.: QinetiQ: Review of Workload Measurement, Analysis and Interpretation Methods. European organization for the safety of air navigation (2003)
4. Cain, B.: A Review of the Mental Workload Literature. (Tech. Rep. RTO-TR-HFM-121-Part-II). Defence Research and Development Toronto, Toronto, Canada (2007)
5. Corwin, W.H., Sandry-Garza, D.L., Biferno, M.H., Boucek, G.P., Logan, A.L., Jonsson, J.E., Metalis, S.A.: Assessment of Crew Workload Measurement Methods, Techniques and Procedures. vol. I - Process, methods, and results (WRDC-TR-89-7006). Wright Patterson Air Force Base, OH (1989)
6. Harris, D.: Human Performance on the Flight Deck. Ashgate (2011)
7. Veltman, J.A., Gaillard, A.W.K.: Physiological Workload Reaction to Increase Level of Task Difficulty. Ergonomics 41, 656–669 (1998)
8. Wilson, G.F.: An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. The International Journal of Aviation Psychology 12(1), 3–18 (2002)
9. Wickens, C.D., Hollands, J.G.: Engineering Psychology and Human Performance, 3rd edn. Prentice-Hall Inc., Upper Saddle River (2000)
10. Van Orden, K.F., Limbert, W., Makeig, S., Jung, T.P.: Eye Activity Correlates of Workload during a Visuospatial Memory Task. Human Factors 43(1), 111–121 (2001)

Pilot Attention Allocation Modeling under Multiple Factors Condition

Xu Wu, Xiaoru Wanyan, and Damin Zhuang

School of Aeronautics Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China
e126126_19@126.com,
{wanyanxiaoru,dmzhuang}@buaa.edu.cn

Abstract. A new forecast model of attention allocation was built on four trade-off factors, including information priority, occurring probability, salient and effort, which involved both channels of information processing. To validate this model, sixteen participants from Beijing University of Aeronautics and Astronautics were recruited to perform an instrument monitoring task under different conditions. Participants were required to concentrate on monitoring the flight indicators presented on a simulation interface of head-up display and respond to abnormal information (anyone of the indicators went out of the normal range) by pressing the corresponding buttons on the keyboard. Fixation distribution was recorded as evaluation index of attention allocation using Smart Eye Pro 4.5 eye-movement tracking device. Simultaneously, reaction time and correct rate of key-press response were recorded by computer automatically as evaluation indices of behavior performance. The regression analysis revealed good agreement between fixation distribution and theoretical results.

Keywords: attention allocation, mathematical modeling, eye movement tracking, cognitive ergonomics.

1 Introduction

Early researchers found that a comprehensive, accurate and timely information collection is necessary for pilots to fulfill complicated flight tasks in the human-machine interaction system of cockpit display [1]. Therefore, analyzing pilot's attention allocation mechanism is able to provide ergonomics evidence for cockpit display interface design, thereby improving flight performance and safety. Attention was one of the most important mental modulatory mechanisms to endow perception with selective ability during information processing [2]. Generally speaking, there were three main methods to investigate attention allocation, including subjective evaluation, eye movement tracking and cognitive modeling. Recently, cognitive modeling instead of subjective evaluation becomes a primary manner to investigate the essence of attention mechanism and the method of eye movement tracking is usually used for experiment validation.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 212–221, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Senders introduced the concept of bandwidth and established one of first quantitative models of instrument monitoring behavior [3], which suggested the relationship between visual scan and instrument bandwidth. Wickens et al. considered that salient, effort, expectancy and value were four influencing factors of attention allocation and then put forward the SEEV model [4]. Meanwhile, another model called A-SA (attention-situation awareness) was built based on both SEEV modules and situation awareness measuring to predict pilot's error [5]. However, Miller pointed out the limitations of SEEV model in the technical report AHFD-04-17 /NASA-04-6 [6], considering that the model should be applied using ratio scale measurements and the weight of parameters should be deliberated when its application conditions changed.

Besides, fuzzy control models were adopted to express the uncertainty of psychological state and congitive activities. Such methods were used by Nobuyuki Matsui et al. and Lou Yan et al. to bulid their attention allocation models [7][8]. In addition, Wanyan et al. advanced their models by introducing hybird entropy and analyzing the influences of information importance, protential possibility as well as detection efficiency on pilot's attention allocation mechanism [9]. Since subjective evaluation of information importance was fuzzy and random, it became the core foundation of their model. Moreover, detection efficiency decipted the difference of processing time caused by various information coding features and the distinct requirement of processing depth, while the protential probability was optimized and calculated when the hybird entropy reached its max. Nevertheless, as the information processing time was measured beforehand, it reduced the predictable ability of the model.

Based on Wanyan's model, the present study started with a cognitive evaluation model of information priority. Then, a new forecast model of attention allocation was established on both channels of information processing: top-down was explained as information priority and occurring probability according to subjective expectancy utility theory; bottom-up was inherited as salient and effort of SEEV model to extend the concept of detection efficiency from [9]. In order to apply and validate this model, an ergonomics experiment was developed using a head-up display simulation interface. Combined with eye movement tracker, the actual attention allocation of participants were obtained and compared with the theoretical results of the model.

2 Mathematical Modeling

2.1 Cognitive Evaluation Model of Information Priority

In common human-machine interaction system, information was received from visual display terminal and the human-machine interface could be divided into n areas of interest (AOIs). Such visual information was captured and transferred to brain by neural system. Since the different stimulus of visual information and the inconstant physiological state, arousing level of brain ω was changing during this procedure. The whole ω consisted n dimensional spaceΩ. However, arousing level of brain ω was influenced by various obtained information, and the results of information processing were unable to predict. Therefore, information priority function V was introduced to reveal the cognitive evaluation results of information priority:

$$V = \int_{\Omega} \psi^*(\omega,t)\bar{V}\psi(\omega,t)\, d\omega \tag{1}$$

$\psi^*(\omega,t)$ was the Hermite conjugated with $\psi(\omega,t)$. The whole square integrable $\psi(\omega,t)$ consisted Hilbert space $L^2(\Omega)$ at moment t onΩ. Operator $\bar{V}$ on $L^2(\Omega)$ depicted the potential information priority. Define state vector $\psi(\omega,t)$ on $L^2(\Omega)$ as the information cognitive state vector, and thus $L^2(\Omega)$ was the information cognitive space. Therefore, there existed a group of basal state vectors in $L^2(\Omega)$ to constitute the standard orthogonal system, so that each state vector $\psi(\omega,t)$ in space $L^2(\Omega)$ could be expressed as:

$$\psi(\omega,t) = \sum_{j=1}^{\infty} c_j(t)\varphi_j(\omega) \tag{2}$$

$c_j(t)$ was the coefficient of the expansion of state vector at moment t. Group of base vectors $\{\varphi_j(\omega)\}$ were chosen by implicative state of arousing level ω and memory formed by human experience as well as potential state through learning. In other words, based on the acknowledgement of information implicative cognitive state and according to learning and memory, the results of information cognition were updated so that $\{\varphi_j(\omega)\}$ was considered as the ability of potential cognition of information processing system when the arousing level ω was inspired. Accordingly, each possible state vector under various arousing level ω constituted a huge function group $\{\psi_1(\omega,t),\dots,\psi_L(\omega,t)\}$, which reflected statistics characteristic of mental activities:

$$\psi_k(\omega,t) = \sum_{j=1}^{\infty} c_j^k(t)\varphi_j(\omega)\,, k = 1,2,\dots,L \tag{3}$$

According to Eq.1 to Eq.3:

$$V = \sum_{l,m}^{\infty} \frac{1}{L}\int_{\Omega} \sum_{k=1}^{L}\left[c_l^{k^*}(t)c_m^k(t)\varphi_l^*(\omega)\bar{V}\varphi_m(\omega)\right] d\omega, l,m = 1,2,\dots,\infty \tag{4}$$

In Hilbert spaceΩ, the order of integral and summation was interchangeable:

$$V = \sum_{l,m}^{\infty}\left[\frac{1}{L}\sum_{k=1}^{L} c_l^{k^*}(t)c_m^k(t)\right]\int_{\Omega} \varphi_l^*(\omega)\bar{V}\varphi_m(\omega)\, d\omega, l,m = 1,2,\dots,\infty \tag{5}$$

Assuming:

$$P_{l,m} = \frac{1}{L}\sum_{k=1}^{L} c_l^{k^*}(t)c_m^k(t)\,, Y_{l,m} = \int_{\Omega} \varphi_l^*(\omega)\bar{V}\varphi_m(\omega)\, d\omega \tag{6}$$

So that:

$$V = \sum_{l,m} P_{l,m}Y_{l,m} \tag{7}$$

As the base vectors of $L^2(\Omega)$ were orthogonal to each other, it was easy to find an appropriate group of base vectors $\{\varphi_j(\omega)\}$ which made $P_{l,m}$ and $Y_{l,m}$ to be diagonal matrixes:

$$P = diag[p_1, p_2, .., p_n], Y = diag[y_1, y_2, \dots, y_n] \quad (8)$$

Therefore, information priority function V could be expressed as:

$$V = tr(PY) \quad (9)$$

It was noteworthiness that:

$$p_i \geq 0, \sum_{i=1}^{n} p_i = 1 \quad (10)$$

Hence, the cognitive evaluation result of information priority of AOI$_i$ was:

$$V_i = p_i y_i, i = 1,2, \dots, n \quad (11)$$

p_i meant the possibility of which potential cognitive state became available. y_i meant the cognitive evaluation results of information priority. As y_i was fuzzy and random to realize, as well as attention allocated to visual information was influenced by the physical and psychological state, there was a certain possibility that even the most important information could be ignored and failed to draw attention during a visual search task. Therefore, the possibility was depicted as $p_i$which might be optimized and calculated according to Eq.12:

$$p_i^* = \frac{\exp h_i}{\sum_1^n \exp h_i} \quad (12)$$

Here, h_i was the binary fuzzy entropy ofχ_i:

$$h_i = -\chi_i ln\chi_i - (1 - \chi_i) ln(1 - \chi_i) \quad (13)$$

And, χ_i was the membership degree standardized into [0, 1] based on the cognitive evaluation results of information priority$\bar{V}$. Therefore, Eq.11 was rewritten as:

$$V_i = p_i^* \chi_i, i = 1,2, \dots, n \quad (14)$$

2.2 Attention Allocation Model under Multi-factor Condition

Assuming the attention resource allocated to n AOIs was expressed as vectors:

$$A = (A_1, A_2, \dots, A_n) \quad (15)$$

Generally, information processing was composed of two channels: top-down and bottom-up. The former one was kind of automatic search or selective control of visual system while the latter one was driven by visual features of information. Subjective expectancy utility theory and SEEV model were introduced to explain the two channels respectively. Based on SEU theory [10], attention allocation based on top-down channel was quantified as the product of information priority $V$and occurring probability$P$. V was the factor involved with task requirement that could be determined as Eq.14. P indicated the occurring frequency and the amount of obtained information referring to the bandwidth of Sender's model. On the other hand, the factor of bottom-up was described as detection efficiency in reference [9] which might be expanded as two factors inherited of SEEV model: effort $E$and salient S. Effort was paid out if eye movement or head movement was required to obtain visual information. It was measured by the distance between AOIs. Salient model was put forward by Itti [11] and adopted by Wickens in SEEV to depict bottom-up attention capture. The present model followed the original definition of salient, and considered S as the interaction results of color c, size s along with character type t in the viewpoint of multiple visual coding of flight information:

$$S_i = (c_i + s_i + t_i)/3 \tag{16}$$

For most human-machine interaction systems in aviation, four factors mentioned before simultaneously influenced and restricted pilot's attention allocation mechanism. However, only salient S was the factor that inhibited attention resource from transferring between distant visual information. Therefore, attention resource allocated to AOI_i could be determined as:

$$A_i = P_i V_i S_i E_i^{-1} \tag{17}$$

And the attention allocation proportion of AOI_i could be determined as:

$$F_i = (A_i / \sum_{i=1}^{n} A_i) \times 100\% \tag{18}$$

3 Method

3.1 Experiment Interface Design

To validate the present model in aviation, experiment interface was designed as head-up display simulation interface and was built by GL Studio software, as shown in Fig.1. The experiment interface was presented on a 19-inch Lenovo LCD with resolution of 1440×900. According to regulations of optical environment for head-up display, the average luminance of experiment interface was moderated at 120cd/m^2 and the average illumination of experiment environment was about 600Lx. Participants were required to make responses through a normal keyboard. Smart Eye Pro 4.5 was used to track eye-movement in a nature way.

Fig. 1. Experiment environment and simulation interface

3.2 Participants

16 participants from Beijing University of Aeronautics and Astronautics were recruited in this study. All participants (10males, 6females; ranging from 22 to 27 years old, mean age 23.9 years old) were familiar with basic computer operation and aeronautics knowledge. All participants were right-handed with normal or correct to normal vision.

3.3 Experiment Design

Five flight indicators were selected as the monitoring task target such as No.1 radar state; No.2 rolling angle; No.3 indicated airspeed; No.4 barometric altitude and No.5 flight symbol. Abnormal state of the five indicators appeared randomly by computer program and only one indicator displayed abnormal state simultaneously. Experiment variables were response score and occurring frequency of abnormal state. In order to simulate a real cognitive situation of flight information, response score was manipulated as the membership degree of information priority according to a certain ratio. In addition, the level of two variables must meet the requirements shown as followed:

$$\sum_{i=1}^{5} frequency_i = 100(times), \sum_{i=1}^{5}(frequency_i \times score_i) = 100(points) \quad (19)$$

Eq.19 meant that the appearance number of abnormal state was 100 times and the perfect response score was 100 points during a whole trial. To validate the present model, four trials were designed as shown in Fig.2. Obviously, both the score and frequency were constant between AOIs in trial A, which only depicted the influence of bottom-up channel. Trial B and trial C were introduced one of the top-down channels (occurring probability or information priority) on the basis of trial A. Trial D was designed to validate the whole factors corresponding to the present model.

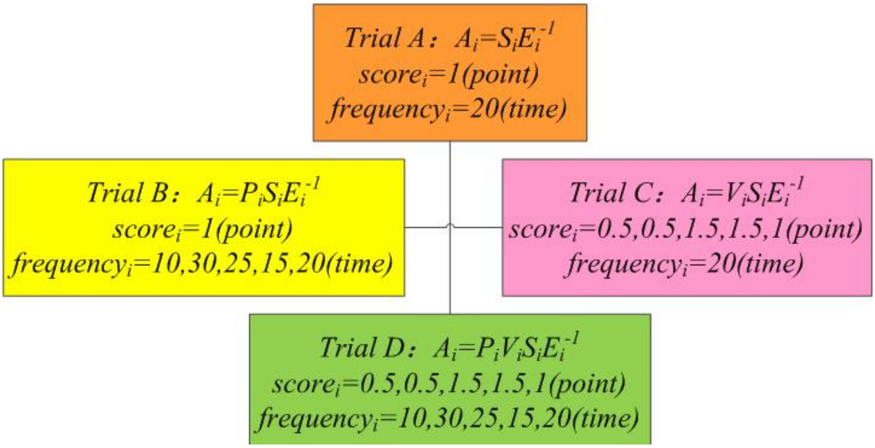


Fig. 2. Experiment trials design

3.4 Experiment Task

The monitoring task and keyboard response were performed in the experiment. Participants were required to allocate their attention resource according to different trials. As the experiment was expected to perform at a medium level of mental workload, participants were asked to make their response within 1s and the interval of continuous abnormal information was 2s [12]. Moreover, Latin Square design was adopted to avoid the errors caused by experiment order and fatigue. Each trial lasted for 4min with 2min break. The Smart Eye system kept real-time tracking during the whole trials.

4 Results

4.1 Theoretical Results of Mathematical Model

In trial A, both scores and frequencies were equal for all the AOIs. Hence, only salient and effort represented for the bottom-up channels were involved and were determined by the visual characteristic of simulation interface. Salient was measured by the performance results of multiple visual coding according to the former research [13] along with Eq.16 and effort was measured by the distance between AOIs, as shown in Tab.1. Values of all the factors were standardized into [0, 1] and the theoretical results of attention allocation were calculated as shown in Tab.2.

Table 1. Values of salient and effort

Area of interest	1	2	3	4	5
color	1.00	1.00	1.00	1.00	1.00
size	1.00	1.00	1.00	1.00	1.00
type	0.50	1.00	0.80	0.80	1.00
salient	0.83	1.00	0.93	0.93	1.00
distance(cm)	13.75	14.50	18.25	18.25	11.25
effort	0.75	0.80	1.00	1.00	0.60

Table 2. Theoretical results of the forecast model

Trial	Area of interest	1	2	3	4	5
A	S (10^{-2})	17.70	21.32	19.83	19.83	21.32
	E (10^{-2})	18.07	19.28	24.10	24.10	14.46
	Attention (%)	18.81	21.25	15.81	15.81	28.33
B	P (10^{-2})	10.00	30.00	25.00	15.00	20.00
	Attention (%)	9.29	31.49	19.52	11.71	27.99
C	V (10^{-2})	9.00	9.00	27.00	27.00	28.00
	Attention (%)	8.43	9.53	21.26	21.26	39.52
D	Attention (%)	4.19	14.21	26.43	15.86	39.30

Trial B and trial C added occurring probability and information priority on the basis of trial A, respectively. Probability was decided by the experiment design of frequency. According to the cognitive evaluation model mentioned in part 2.1, information priority was calculated by response score as shown in Tab.3. In addition, trial D was designed to validate all four factors from Eq.17 and its theoretical results were shown in Tab.2.

Table 3. Cognitive evalutaion of information priority

Area of interest	1	2	3	4	5
score	0.50	0.50	1.50	1.50	1.00
membership degree χ_i	0.25	0.25	0.75	0.75	0.50
potential possibility p_i^*	0.18	0.18	0.18	0.18	0.28
information priority V	0.05	0.05	0.14	0.14	0.14

4.2 Experimental Results of Eye-Movement Tracking

During the whole trials, the infrared images were transformed into digital images by a PCI frame grabber with the sampling rate of 60Hz. The experimental results of attention allocation proportion were defined as the ratio of fixation points for a certain AOI_i to the fixation points for all the AOIs, as shown in Tab.4.

Table 4. Attention allocation proportion results (%)

Trial	Area of interest				
	1	2	3	4	5
A	18.49±2.33	22.07±2.81	16.17±1.58	15.78±1.74	27.62±1.98
B	6.61±3.29	33.12±5.39	16.95±3.43	14.95±2.98	27.96±2.90
C	10.15±2.43	3.27±1.73	24.26±3.94	23.33±4.42	39.01±4.87
D	6.82±2.29	15.64±5.24	24.98±3.25	17.39±3.59	35.14±4.26

4.3 Statistical Analysis of Theoretical and Experimental Results

For all trials, the regression analysis showed good agreement between theoretical and experimental results of attention allocation ($r=0.91$, $R^2=0.83$, $p<0.01$). Correlation analysis revealed significant correlation between fixation distribution and the four factors of the model: $r_P=0.402$, $r_V=0.654$, $r_S=0.552$ and $r_E=-0.303$.

The model was established in a nonlinear method, it was quite different from linear weighting method of SEEV model. In order to make comparison of these two methods, another regression analysis was applied to validate the four factors in a linear weighting method, as shown in Tab.5. The results showed less satisfied correlation with fixation distribution ($R^2=0.80$).

Table 5. Linear regression results

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
(Constant)	-0.079	0.048		-1.628	0.105
P	0.759	0.063	0.381	12.100	0.000
V	1.116	0.043	0.711	26.159	0.000
S	0.692	0.25	0.092	2.762	0.006
E	-1.174	0.074	-0.436	-15.861	0.000

5 Discussion

It could be seen from Tab.2 and Tab.3 that the data obtained from our experiments were consistent with the theoretical results calculated from the attention allocation model. Trial A was designed to verify the bottom-up influences of salient and effort. Based on it, trial B and trial C respectively added one single top-down factor and finally trial D was designed to examine the whole model. Therefore, the effectiveness of the attention allocation model and the connections of its factors were confirmed. However, since the experiment was carried out at a medium level of mental workload, the behavior performance of keyboard response was considerably high with most accurate rate above 90% for all trials. In that case, it was not necessary to include the performance result here. In addition, the present model was built on four multiplicative factors which simplify and optimize the linear weighting method adopted in SEEV model according to the results shown in Tab.5.

The multiple factors from the present model involved both of the information processing channels of top-down and bottom-up. During the early attention capture phase, parallel manner based on bottom-up was used and considered as the fundamental characteristic of information processing [14]. It was depicted in the model that the salient information was easier to capture attention and the information requiring more effort cost was reluctant to observe. In the feature integration phase, the priority and bandwidth were considered to draw attention in the top-down way. Therefore, information of high occurring probability or high importance related to task was preferred to get more attention resource. Nevertheless, the predicted allocation of attention resource was balanced by these trade-off factors.

There were still drawbacks in the present study, the model was validated in a low fidelity simulation experiment, and the influence of task type and some extreme levels of mental workload were neglected. For example, it was unsure whether the model was still fit in an overload or underload condition. Further research would focus on the internal connection between mental workload and attention allocation, and it hoped that an ideal prediction of pilot's attention allocation under different mental workload conditions could be developed.

6 Conclusion

This paper introduced a mathematical model formed by information priority, occurring probability, salient and effort to predict pilot's attention allocation. The effectiveness of model was validated with a target monitoring task in head-up display simulation interface. The experiment results revealed that the model accounted for above 83% of the variance in visual attention allocation between trials. Moreover, the present study could provide ergonomics evidence for cockpit display interface design.

Acknowledgement. This study was financially supported by National Basic Research Program of China (Program Grant No.2010CB734104).

References

1. Guo, X., Liu, B., Ma, X., et al.: General information of display and its priority for advanced fighter cockpit. Chinese Journal of Aerospace Medicine 17, 260–263 (2006)
2. Wickens, C.D., McCarley, J.S.: Applied attention theory, pp. 41–58. CRC Press, Boca Raton
3. Senders, J.W.: The Human Operator as a Monitor and Controller of Multidegree of Freedom Systems. Human Factors in Electronics 5, 2–5 (1964)
4. Wickens, C.D., Goh, J., Helleberg, J., et al.: Attentional models of multi-task pilot performance using advanced display technology. Human Factors 45, 360–380 (2003)
5. Wickens, C.D., McCarley, J.S., Alexander, A., et al.: Attention-situation awareness (A-SA) model of pilot error. In: Foyle, D., Hooey, B. (eds.) Pilot Performance Models, pp. 213–240 (2007)
6. Miller, S.M., Kirlik, A., Kosorukoff, A., et al.: Ecological Validity as a Mediator of Visual Attention Allocation in Human-Machine Systems. NASA-04-6, pp. 12–14 (2004)
7. Matsui, N., Bamba, E.: Consideration of the attention allocation problem on the basis of fuzzy entropy. Association Symposium of Measurement and Automatic Control 22, 27–32 (1996)
8. Lou, Y., He, H., Lu, Y.: Attention allocation behavior modeling of virtual driver. Microcomputer Information 24, 274–276 (2008)
9. Wanyan, X., Zhuang, D., Wei, H., et al.: Pilot Attention Allocation Model Based on Fuzzy Theory. Computers & Mathematics with Plications 62, 2727–2735 (2011)
10. Karni, E.: Subjective expected utility theory with state-dependent preferences. Journal of Economics Theory 60, 428–438 (1993)
11. Itti, L., Koch, C.: A saliency-based search mechanism for overt and covert shifts of visual attention. Vision Research 40, 1489–1506 (2000)
12. Zeng, Q., Zhuang, D.: Target identification of different task weight under multi-interface and multi-task. Journal of Beijing University of Aeronautics and Astronautics 32, 499–502 (2006)
13. Wu, X., Wanyan, X., Zhuang, D., et al.: Ergonomics study of information coding of cockpit display under certain ambient illumination. In: The 4th International Conference on Image and Signal Processing, pp. 305–308. IEEE (2012)
14. Treisman, A., Gelade, G.A.: A feature integration theory of attention. Cognitive Psychology 12, 97–136 (1980)

A Coherent Assessment of Visual Ergonomics in Flight Deck Impacted by Color and Luminance

Ye Zhou[1], Wei Zhang[1], Baofeng Li[2], Jinhai Yu[2], and Zhi Ma[1]

[1] School of Aeronautics, Northwestern Polytechnical University, Xi'an, China
[2] Shanghai Aircraft Design and Research Institute, COMAC, Shanghai, China
zhouy060103@mail.nwpu.edu.cn

Abstract. This research proposed a coherent assessment method for evaluating visual ergonomics in simulated flight deck via evaluating psychological indices and performance during a series of experiments. A simulated flight deck environment was established according to the dimensions of a real commercial aircraft cockpit with back projection, and then a balanced sequence of pseudo random variates was generated with replicated Latin square design. After the experiment of 18 interior color levels and 2 luminance levels, a complex statistical analysis was conducted to examine the significance as well as correlations within different factors. The fluctuation of luminance can affects the results slightly, while the change of the interior color, both hue and saturation-intensity levels, can influence subjects' visual ergonomics significantly and interactively. This coherent assessment indicates that light blue was the best choice, whereas, vivid yellow was the last among the 18 colors.

Keywords: Visual ergonomics, psychological indices, performance, coherent assessment.

1 Introduction

According to the statistical results by Boeing Company, there were 1,085 fatal accidents involving commercial aircraft, world-wide, from 1950 through 2010 for which a specific cause was known [1]. Despite the continuing growth of worldwide commercial air traffic, accident rate persist nearly the same over the last 20 years. Therefore, the number of fatal accidents will increase. Mainly human factors, including weather related pilot errors, mechanical related pilot errors, air traffic controller errors, improper loading of aircraft, and other human errors, have been identified as the primary cause of air traffic accidents. Wrong decisions resting upon insufficient situation awareness are the main reason within that focus.

Flight deck environment, where pilots perform tasks, is the essential part, marking its safety of navigation. Thus, aircraft designers should transform their research focus into cognitive ergonomics related factors design of flight deck, which can affect interaction effectiveness subtly and remarkably, to guarantee pilots' adequate situation awareness (SA) and performance, rather than solely on the construction quality and morphological characteristics.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 222–230, 2013.
© Springer-Verlag Berlin Heidelberg 2013

This experiment being made by our research group aimed to investigate the influence of the ambient color and illumination of cockpit interior environment on pilots' psychological reactions and performance. The effects of colors of different hue, saturation and intensity levels and light or dark illuminance on visual ergonomics were examined.

1.1 Visual Ergonomic Assessment in Flight Deck

Numerous theories of ergonomic assessment of interior color and luminance have been developed [2], yet very little experimental research on the effects of interior color on pilots' visual ergonomics in the flight deck has been reported. In this special and small aircraft environment, ambient color plays an important role in enhancing pilots' psychological states, physiological status, performance, and even situation awareness by influencing pilots' visual ergonomics.

The constitution of SA is multifaceted and intricate, due to the most commonly used three-level model [3]. How the ambient environment affects pilots' SA is via fluctuating their anxiety, arousability, fatigue and laterality, which will further influence their attentiveness, memory, orientation, discrimination and capacity to withstand fatigue. Accordingly, visual ergonomics affecting SA should be assessed from different aspects simultaneously. Several ergonomics measurements for SA have been established [4], which highly suggests an integrated approach including human performance and subjective, physiological indices, self-report ratings or observer-rating techniques.

Two different ergonomics assessment tools were used coherently in this experiment to see if the different results converge into the same direction and how they can be integrated to evaluate the visual ergonomics in flight deck with different ambient environments.

1.2 Hypotheses

In terms of subjective rating and task performance, it was predicted that the interior color and ambient lighting conditions would affects the subjects main-effectively and interactively. In addition, the hue, saturation and intensity are predicted to have main effects on visual ergonomics.

Even though stimulus screening ability was considered as an individual differences which may interact with the interior environment significantly [5-8] and may affect experiment results, the aims of this experiment was to evaluate the visual ergonomics of a specific flight deck as well as the stability of the effects on subjects with different stimulus screening ability. Thus, the giant differences among pilots make the clarification of subjects unnecessary. The individual difference was assumed to be minor compared to the effects of ambient color and luminance in this within-subjects design.

2 Methods

2.1 Design and Subjects

This experiment employed a design with "ambient color" as a within-subject variable with 18 levels, "work plane illuminance" as a dichotomous within-subject variable.

Independent Variables. After discussed in experimenter groups, 18 color variable levels, with 6 hue levels and 3 different saturation-intensity levels were chosen and detailed in Table 1. The illuminance levels were set as "light" from 120 to 240 lx, and "dark" from 0 to 120 lx.

Table 1. Variable levels of simulator interior color with Munsell notations

Hue	Saturation-Intensity		
	High S.-Medium I.	Low S.-High I.	Low S.-Low I.
Red	5R 5/18	5R 7/6	5R 2/6
Orange	2.5YR 6/18	2.5YR 8/6	2.5YR 2/6
Yellow	10Y 9/12	10Y 9/4	10Y 3/4
Green	10GY 7/16	10GY 8/6	10GY 2/6
Blue	10B 6/16	10B 8/6	10B 2/6
Gray	N 6/	N 8/	N 2/

Dependent Variables. During performing tasks in this experiment, participants' physical workload, mental workload and performance were analyzed using subjective, self-report rating scales together with performance via following methods:

Subjective Rating Scale. The simple rating technique called Instantaneous Self-Assessment (ISA) rating scale [9], which comprised several on-line ratings of the pilots, was administered in the current research to measure the pilots' overview of the situation on that particular moment in the scenario. While, the NASA-TLX [10] was a multi-dimensional rating scale and applied to assess pilots' workload comprehensively. According to this, VEISA (Visual Ergonomics Instantaneous Self-Assessment) was devised to evaluate subjects' anxiety, arousability and fatigue every 10 minutes to provide an overview assessment during the trial process (1 being very high and 5 being very low). Besides, subjects were asked to respond to the VERQ (Visual Ergonomics Rating Questionnaire) to give their assessment in terms of luminance, glare and spaciousness of the scenario they exposed to.

Performance. The accuracy and response time together were recorded by E-Prime program for assessing pilots' performance, and thus situation awareness during the landolt ring differentiation task and numerical verification task.

Subjects. A total of 36 graduate students (30 males and 6 females), whose visual acuity and color vision were qualified by Ishihara tests and Monoyer's chart, participated in this study as unpaid volunteers. Their mean age was 23.7 years (SD = 2.61).

Sequential Effects. Each participant spent about 3.5 hours in the scenario either in the morning, afternoon, or evening on three different time periods with one week in between. Each time, participants were firstly exposed to a controlled condition, and then exposed to the colorful rooms with different illuminance levels randomly in a balanced sequence determined by a replicated Latin square design.

2.2 Laboratory Setup

There is a prevailing perception that different colors arouse people's anxiety, arousal, fatigue as well as their performances to different extent. However, few scientifically conclusive results can be drawn about the impact of color on human performance [11], as many studies have lacked the necessary experimental control and appropriate design measures. Specifically, poor resemblance of the experimental conditions to the real contexts, unchangeable real world settings and the lack of long-term assessment have severe confused influence on the experiment outcomes.

Simulator Environment. This research proposed a simulator experimental method, applying back projection to change the ambient color, and several light emitting diodes (LEDs) to provide different illumination intensity levels. The simulated environment was constructed by four pieces of translucent screen, illustrated in Fig. 1, in accordance with the dimensions of a real commercial aircraft cockpit, and a projector for each surface.

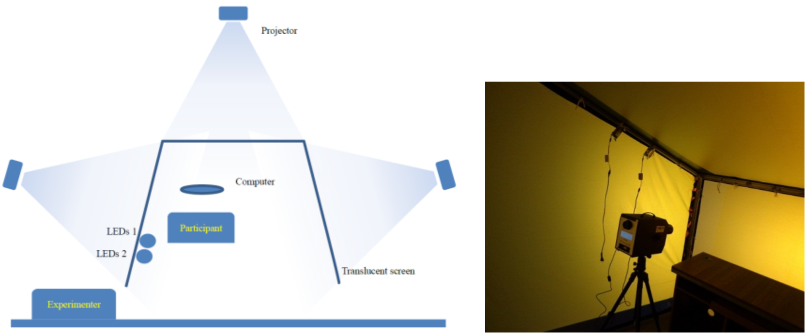


Fig. 1. Simulated experimental environment

Color and Illuminance Variable Levels Adjustment. The CIE (Y, x and y) equivalents of the color levels set in Munsell notations can be calculated assuming the standard Illuminant C as neutral origin, which approximates 6700°K. Therefore, the color variable levels were adjusted by Everfine CBM-8 color luminance meter, shown in table 2.

Table 2. Adjustment and measurement of color levels with CIE Yxy

Hue	High S.-Medium I.			Low S.-High I.			Low S.-Low I.		
	$Y/cd/m^2$	x	y	$Y/cd/m^2$	x	y	$Y/cd/m^2$	x	y
Red	55.6	0.49	0.25	220.7	0.34	0.36	30.8	0.43	0.32
Orange	117.4	0.50	0.43	188.9	0.37	0.38	30.2	0.43	0.38
Yellow	237.2	0.40	0.48	244.5	0.33	0.37	43.4	0.35	0.37
Green	171.7	0.30	0.54	194.2	0.31	0.36	32.7	0.27	0.47
Blue	147.6	0.21	0.28	216.3	0.27	0.32	44.0	0.22	0.26
Gray	101.4	0.30	0.33	166.5	0.30	0.33	36.3	0.30	0.33

Since lighting and color have a coherent impact on illuminance, the illuminance in each variable levels were measured by Everfine illuminance meter, shown in table 3. Most of them meet the setting value above.

Table 3. Measurement of illuminance levels in different color levels

Hue	High S.-Medium I.		Low S.-High I.		Low S.-Low I.	
	Light	Dark	Light	Dark	Light	Dark
Red	120.22	47.11	163.91	90	112.66	39
Orange	134.60	61.01	160.32	86.69	109.01	35.04
Yellow	215.13	142.45	218.08	145.41	115.75	41.75
Green	189.85	116.3	202.50	129.55	110.73	36.63
Blue	170.33	96.75	200.61	127.5	111.55	37.27
Gray	140.50	66.3	191.16	117.03	108.56	34.24

3 Results

3.1 Analysis of Variance of Color and Luminance

In terms of anxiety, a significant main effects of color, $F_{(17,1260)} = 95.20$, $p<0.01$, main effects of luminance, $F_{(1,1260)} = 14.73$, $p<0.01$, and interactive effects of color and luminance, $F_{(17,1260)} =10.51$, $p<0.01$, were found. For the accuracy during landolt ring differentiation task, a significant main effects of color, $F_{(17,1260)} = 15.43$, $p<0.01$, and interactive effects $F_{(17,1260)}= 13.11$, $p<0.01$, as well as minor significant main effects of luminance, $F_{(1,1260)}= 4.15$, $p<0.05$, were detected. For other 8 indices, the main effects of color and interaction were significant, while, the main effects of luminance were non-significant.

As we predicted, the trivial fluctuation of luminance can only affects the results slightly. By contrast, the change of the 18 interior color levels can significantly influence subjects' psychological status and task performance. Luminance was mostly applied to examine the stability of a color's effect in this experiment.

3.2 Analysis of Variance of Hue, Saturation and Intensity

In terms of anxiety, a significant main effects of hue, $F_{(5,630)} = 60.95$, $p<0.01$, and interactive effects of hue and saturation-intensity, $F_{(10,630)} = 7.331$, $p<0.01$, were found. While, the main effects of saturation-intensity, $F_{(2,630)} = 2.001$, was non-significant. For anxiety and arousability, the main effects of both hue and saturation-intensity as well as their interactions were significant. The simple effects were detailed in Fig.2.

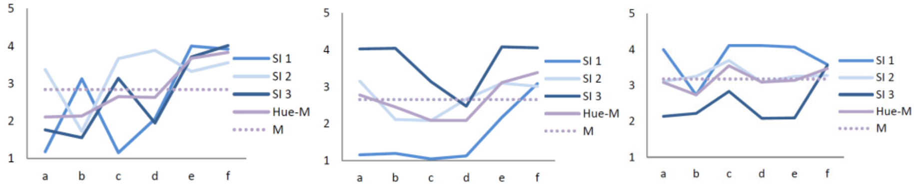


Fig. 2. Anxiety, arousability and fatigue by saturation-intensity levels over hue levels

In terms of luminance, glare and spaciousness, the main effects of both hue and saturation-intensity as well as their interactions were significant. The simple effects were detailed in Fig.3.

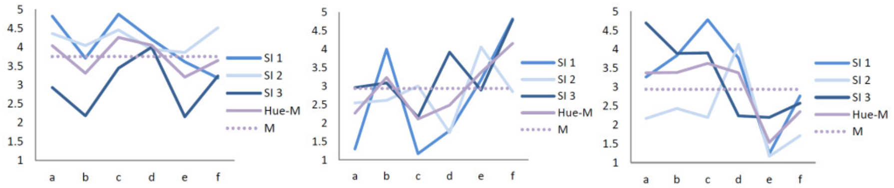


Fig. 3. Luminance, glare and spaciousness by saturation-intensity levels over hue levels

In terms of accuracy during numerical verification task, significant hue main effects, $F_{(5,630)} = 97.96$, $p<0.01$, saturation-intensity main effects, $F_{(2,630)} = 17.36$, $p<0.01$, and interactive effects of hue and saturation-intensity, $F_{(10,630)} = 7.331$, $p<0.01$, were found. While, in terms of response time, a minor significant interactive effects was found, $F_{(10,630)} = 2.098$, $p<0.05$, while, the hue main effects, $F_{(5,630)} = 1.378$, saturation-intensity main effects, $F_{(2,630)} = 0.8743$, were non-significant. The simple effects were detailed in Fig.4.

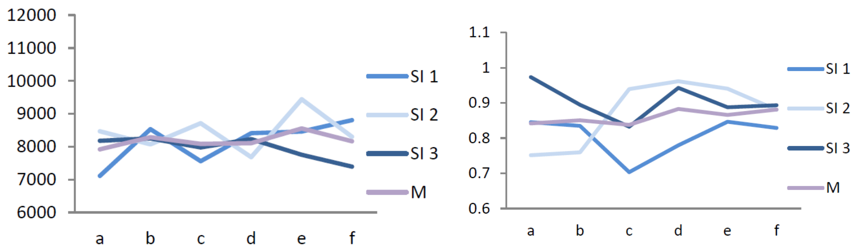


Fig. 4. Response time (ms) and accuracy (%) during numerical verification task by saturation-intensity levels over hue levels

In terms of accuracy during landolt ring differentiation task, significant hue main effects, $F_{(5,630)} = 2.359$, $p<0.01$, saturation-intensity main effects, $F_{(2,630)} = 16.65$, $p<0.01$, and interactive effects of hue and saturation-intensity, $F_{(10,630)} = 27.92$, $p<0.01$, were found. While, in terms of response time, a significant interactive effects was found, $F_{(10,630)} = 5.083$, $p<0.01$, whereas, the hue main effects, $F_{(5,630)} = 1.049$, saturation-intensity main effects, $F_{(2,630)} = 2.251$, were non-significant. The simple effects were detailed in Fig.5.

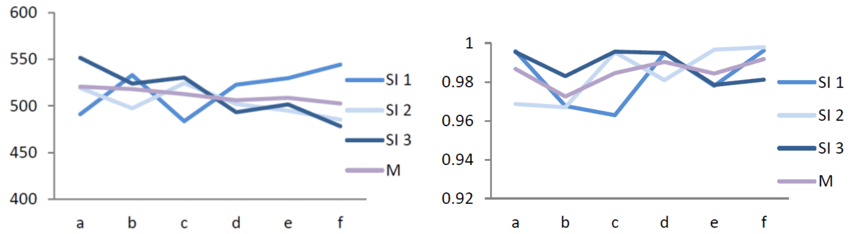


Fig. 5. Response time (ms) and accuracy (%) during the landolt ring differentiation task by saturation-intensity levels over hue levels

The obtained results of this experiment indicate significant main effects for flight deck ambient color on participants. Particularly, hue of ambient color significantly affected anxiety, arousability, luminance, glare and spaciousness; saturation-intensity variable significantly affected arousability, fatigue and luminance of subjects. The interactive effects of hue and saturation-intensity of ambient color is evidently significant in terms of task performance.

3.3 Coherent Assessment of Visual Ergonomics

The score of each color on 10 sub-indices was marked (-1 being bad and unacceptable, 0 being acceptable, 1 being good) via analyzing the data and diagrams above. The weights of 6 subjective sub-indices were measured in evaluation of the overall psychological effects according to subjects' 5 rate scale after their third session, as shown in table 4.

Table 4. The weights of six subjective sub-indices

Anxiety	Arousability	Fatigue	Luminance	Glare	Spaciousness
0.2248	0.2171	0.1405	0.1826	0.1341	0.1009

The 4 sub-indices of performance could hardly be assumed to be more important or less important than any other, thus, were assumed to be of same important. The normalized overall psychological ratings, performance ratings as well as their mean values and differences were detailed in table 5. The mean values and differences of the two indices should be considered coherently to evaluate the comprehensive visual ergonomics. The vivid blue-SI1 and light blue-SI2 were considered as to affect subjects' psychological ratings most positively, while vivid red-SI1, dark orange-SI3 and vivid yellow-SI1 were less positive; dark red-SI3 and light blue-SI2 facilitate task performance mostly, while vivid yellow-SI1 and dark gray-SI3 hampered subjects status from well performing tasks. Comprehensively, light blue-SI2 was the best choice among 18 flight deck interior colors in this experiment, followed by moderate gray-SI1, pale yellow-SI2 and light orange-SI2. By contrast, vivid yellow-SI1 was the last, followed by vivid red-SI1. Vivid blue-SI1and dark gray-SI3 were highly phrased subjectively, their effects on performance were relatively unacceptable, whereas dark red-SI3, dark orange- SI3 and light green-SI1 which were less preferred subjectively, stimulate subjects accuracy and response time when performing tasks.

In order to give a concise and coherent quantity to evaluate visual ergonomics, the equation 1 was proposed.

$$VE = \frac{c(w_1 Phy + w_1 Per)}{|Phy - Per| + c} \tag{1}$$

Where, w_1, w_2 and c are positive constants. Assuming the psychological effects and performance effects impacted by interior color were of the same importance, which means $w_1 = 0.5$, $w_2 = 0.5$, c was tested and set to 0.3, and thus the coherent value of visual ergonomics were calculated and detailed in table 5.

Table 5. Coherent assessment of visual ergonomics

	Phy.	Per.	Mean	|Phy.-Per.|	V.E.
Red-SI1	0.1910	0.25	0.2205	0.0592	0.1842
Red-SI2	0.7670	0.25	0.5085	0.5170	0.1867
Red-SI3	0.3167	0.875	0.59585	0.5583	0.2083
Orange-SI1	0.6820	0.375	0.5285	0.3070	0.2612
Orange-SI2	0.7248	0.5	0.6124	0.2248	0.3501
Orange-SI3	0.1341	0.625	0.3796	0.4909	0.1440
Yellow-SI1	0.1405	0	0.0703	0.1405	0.0478
Yellow-SI2	0.7670	0.625	0.6960	0.1420	0.4724
Yellow-SI3	0.7650	0.5	0.6325	0.2650	0.3358
Green-SI1	0.2746	0.75	0.5123	0.4754	0.1982
Green-SI2	0.7650	0.375	0.5700	0.3900	0.2478
Green-SI3	0.4502	0.625	0.5376	0.1749	0.3396
Blue-SI1	1	0.375	0.6875	0.6250	0.2230
Blue-SI2	1	0.875	0.9375	0.1250	0.6618
Blue-SI3	0.4094	0.5	0.4547	0.0907	0.3492
Gray-SI1	0.9496	0.75	0.8498	0.1996	0.5103
Gray-SI2	0.8174	0.5	0.6587	0.3174	0.3201
Gray-SI3	0.7325	0.125	0.4287	0.6075	0.1417

In conclusion, the result shows that two evaluation indices did not always converge into the same tendency, but can give an overview insight to subjects' physical and mental conditions and further indicate pilots' situation awareness. Participants' responses and evaluation of the simulations of real environments can be established to be adequately similar to people's responses to the actual environment, whereas color, light and other interior elements can be altered effortlessly and naturally. The experimental methods and ideas presented may have wider application in ergonomics analysis of interaction effectiveness for a specific scenario.

Acknowledgments. This research was supported by National Basic Research Program of China (973 Program) (Number: 2010CB734106).

References

1. Statistical Summary of Commercial Jet Airplane Accidents (Worldwide Operation 1959-2011), Boeing, US (2011)
2. Baumstarck, A., Park, N.: The Effects of Dressing Room Lighting on Consumers' Perceptions of Self and Environment. Journal of Interior Design 35, 37–49 (2010)
3. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors 37, 32–64 (1995)
4. Gawron, V.J.: Human performance, workload, situational awareness measures handbook, 2nd edn. CRC Press, Taylor & Francis Group, Boca Raton, Florida (2008)
5. Mehrabian, A.: Theory and evidence bearing on a scale of trait arousability. Current Psychology: Development, Learning, Personality, Social 14, 3–28 (1995)
6. Kwallek, N., Soon, K., Lewis, C.M.: Work Week Productivity, Visual Complexity, and Individual Environmental Sensitivity in Three Offices of Different Color Interiors. Journal of Color Research and Application 32, 130–143 (2007)
7. Dijkstra, K., Pieterse, M.E., Pruyn, A.: Individual Differences in Reactions towards Colour in Simulated Healthcare Environments: The Role of Stimulus Screening Ability. Environmental Psychology 28, 268–277 (2008)
8. Jalil, N.A., Yunus, R.M., Said, N.S.: Environmental Colour Impact upon Human Behaviour: A Review. Procedia - Social and Behavioral Sciences 35, 54–62 (2012)
9. Castle, H., Legatt, H.: Instantaneous self-assessment (ISA)-validity & reliabil-ity (JS 14865 Issue 1). BAE Systems, Bristol, United Kingdom (2002)
10. Hart, S.G., Staveland, L.E.: Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: Hancock, P., Meshkati, N. (eds.) Human Mental Workload, pp. 139–183. North Holland, Amsterdam (1988)
11. Beach, L.R., Wise, B.K., Wise, J.A.: The human factors of color in environ-mental design: Critical review. NASA technical report NCC 22-404, NASA/Ames Research Center, Moffett Field, CA (1987)

Digital Expression of Civil Pilot's Basic Operation

Zhuoyuan Jiang, Bin Chen, Quanxin Cao, and Yuandong Liang

Shanghai Aircraft Design and Research Institute, COMAC, Shanghai
973humanfactor@gmail.com, chenbin@comac.cc

Abstract . "Human Error" is one of the major reasons of aircraft accident and incident. In order to reduce the loss by "Human Error", the "human error" action sequence must be detected and provide warning to pilot or intelligent action when pilot violates the operation procedures or any actions that may cause potential accident or incident. To identify and manage the "Human Error" is becoming more and more important.

The response of unit varies with human's behavior. If the correct operation behavior can be digital described, then human's operation behavior will be partly quantized. Pilots guide the airplane mainly by operations manual, excluding facility failure, there exist operations regularity, that is time-ordered action sequence. Operating steps and therbligs can be intuitive quantitative by coding operations process. Fully considering behavioral characteristics of the pilot group, to optimize time sequence and action sequence parameters will make the operation behavior code be more accurate. The study in operation behavior coding and analysis will play an important role in effectively reducing the probability of flight accidents caused by human factors.

This paper designed and developed a set of effective behavior coding method on the basis of computer compilation principle, starting from unit operation behavior and description of abstraction, using mathematical method to analyze the connections between operation tasks.

Keywords: Human Factors, Human Error, Pilots' Operation Behavior, Coding System.

1 Introduction

Civil plane pilot assumes tasks of piloting and airplane management in the flight profile from taking off to landing. It may lead to serious incidents or even huge aeronautical disasters once there exist human errors in the operational processes of airplane equipment that requires high safety or well functions.

In recent years, with the development of automation of civil airplane, civil aviation incidents that caused by mechanical equipment and automation system decreased, meanwhile, proportion of human errors increased by years. Statistical analysis shows that the proportion of incidents caused by human errors is very high. Generally speaking, the proportion of incidents caused by human errors is more than 70%.

Therefore, to enhance the safe level of civil airplane heavily, inappropriate operational actions of civil airplane pilots should be in control and incline.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 231–240, 2013.
© Springer-Verlag Berlin Heidelberg 2013

It is known to all that everyone makes mistakes. The thought that flight safety can be guaranteed by expecting no human errors is unreal. We can only hope that mistake avoidance of the plane itself and function of fault-tolerance can lower the rate of incidents that caused by human errors.

So far, measures that used to anticipate operational errors in the role of system design mainly are:

1. Feedback of valid information of system equipment: to provide warning sign in the process of design, once operational errors occur, feedback information of the machine will remind operator to correct mistakes.
2. Interlock design: some operations are forbidden when the machine is in a certain state, control with proper electric circuit or mechanism to avoid breakdown caused by operational mistake.
3. Unique design: Only a certain state of the system operation can be accepted, the others are excluded, which eliminates wrong operations in essence.
4. Fault-tolerance design: The mainly reasons that lead to operational mistake are oblivion and muff. Fault-tolerance design means that it allows the existence of mistakes in prerequisite of no safety threat. For example, by the way of control-in-order (that is, the next process begins only when the present is done), operation system that is in charge may avoid the occurring of mistakes.
5. To improve the level of machine automation: a high level automation brings few and simple operations and processes, requires low technical competence and low possibility of mistakes.

The realization of the five points above can be summarized as one, which is, to establish digital coding bases of standard operational procedures, forming unique operation series; to monitor pilots' operation actions, forming real-time operation code. We can detect pilots' erroneous operations in time by comparing and then give a attention, avoiding the phenomenon that the pilot does not aware the wrong operations.

The prerequisite of this job is the quantitative description of civil aviation pilot's fundamental operational actions, i. e., to build code rules of civil aviation pilot's basic operational actions. For this reason, this project does a research on civil aviation pilot's basic operational actions, trying to code pilot's actions and realizing standard expression and quantitative description, providing a new settle thought for computer auto detection and identification of human errors, enriching present functions of warning and alarming, lowering the difficulty of finding mistakes, providing references for error prevention and risk aversion, improving flight safety level in essence.

2 Analysis of Civil Aviation Pilot's Basic Operational Actions

To analyze the realness and integrity of recorded whole operational procedure in the cockpit, as well as accuracy that related to the basic content of code system, we adopted operation flow chart and man-machine operations chart to record procedure and operations in detail while 5W1H is used to verify its reliability.

Six questions of 5W1H when doing the analysis:

1. What-what happened? (object)
2. Why-why do this? (purpose)
3. Where-where do this? (address)
4. When-when to start? When to finish? (time)
5. Who-who do this? (staff)
6. How-how to do? What measures? (method)

Four rules of ECRS when doing the analysis

1. Eliminate: "what is done? Is it necessary? Why?"
2. Combine: If the work or action cannot be cancelled, then can it be combined with other work?
3. Rearrange: Rearrange the work.
4. Simplify: To simplify work content and steps, same with action simplification, energy saving.

When analyzing, the six questions should be conducted systematically in a row, which is the base of successful procedure analysis. Considering more small questions again after getting the answers to those questions by applying this method, then we can do the work of elimination, combination, rearrangement and simplification, comprehensive analysis, then we can get the most accurate procedure and operational process.

3 Basic Operational Actions and Definition of Characteristic Actions

Although operations of pilots in flight process vary, they were built by several basic movements. Our code object is civil planes' basic operation action of pilot. We fixed 19 basic movements after six principles and ECRS analytic operational process and operational procedure, which is called 19 therbligs, see table1.

Table 1. Therbligs

Category	Name	Symbol	Definition
1	Reach	Re	Movement that approaches object
2	Grasp	G	Movement that holds object
3	Whirl	W	Movement that makes object pivot
4	Press	Ps	Movement that makes object move in the direction of force
5	Pull Out	PO	Movement that makes object move in the direction of source force
6	Tread	T	Movement that controls movement of object by foot
7	Release	Rl	Movement that departs object
8	Inspect	I	Movement that compares to standard
9	Report	Rp	Movement that expresses current situation of object
10	Search	Sh	Movement that fix the position of object

Table 1. *(Continued)*

11	Select	St	Movement that select object
12	Plan	Pn	Movement that delays for planning operational program
13	Hold	H	Movement that keeps the statement of object
14	Position	P	Movement that adjusts the position of object
15	Pro-Position	PP	Movement that places object to avoid 'position' movement when object is used
16	Rest	Rt	Movement for rest
17	Unavoidable Delay	UD	Inevitably halt
18	Avoidable Delay	AD	Evitable halt
19	Find	F	Movement that finds object

4 Design of Code of Basic Operational Actions

4.1 Principles of Design of Code

We should obey following principles to code pilot's basic operational actions:

1. Unique Although there are different names and descriptions for basic operational actions, it should be guaranteed that each basic operational action should correspond to a code, a value of code reflects a certain meaning of a basic operational action.
2. Openness The structure of code should adapt to the continue increasing demand of basic operational actions, leaving enough standby codes for new basic operational actions.
3. Terseness In the prerequisite of no influence on meaning expression and expandability of code, the digits of code should be as few as possible in order to reduce error rate, save time of computer processing and memorizing, easy to understand, detect and memorize.
4. Normative The format of code should be normalized to increase reliability. Code is done for basic operational actions in the flight process of pilot. All classifications should have the same code structure for a certain action in any flight phase of flight process holds the line.
5. Consistency Code should satisfy all existed standards and specifications as far as possible.
6. Suitability The code should be convenient to revise to adapt to possible changes of basic operational actions' features or attributes or interrelations.
7. Generalization The code should have enough meaning as far as possible. Code with more meaning may reflect more attributes and features of basic operational actions.
8. Stability It is unsuitable to change code often. It should be considered that the code changed, while the code system should be stable.

9. Detectable It helps to memory, easy to understand and use if the code reflects the meanings and features of basic operational actions as far as possible.
10. Maneuverability Code should be convenient for staff to operate and reduce operating time of machine.

Among these principles, there is a contradiction between terseness and generalization. If the code has strong meaning, the property of terseness will be affected; there is a subtle balance relation between suitability and stability, the control of them should be taken into account. Therefore, we should take all principles above into consideration in actual design process, realizing the optimization of design.

4.2 Dimensions of Code

Code is the only mark of code object, except for the functions that it can precisely define main part, movement, operation object and time, provide information on code object, distinctly reflect categories, attributes, features, etc. of code object, code of measures that are used to reduce human errors and redundancy code that is used to avoid miss of code in transmission are included. Six kinds of code structures are adopted now: hierarchy code, abbreviation code, sequence code, condition code, check code and compound code. We apply following code structure according to requirement of code content and every code structure's feature under code principles. Fig1 Code Structure of Pilot's Basic Operational Actions.

Abbreviate code + Sequence code + Condition code + Position Code + Time code + Check code

Fig. 1. Code Structure of Pilot's Basic Operational Actions

4.3 Code Structure

4.3.1 Abbreviate Code

Operational main part that involved in the process of pilot's actual operation-human's body regions are few, and there are only 19 pilot's basic operational therbligs, then we can use the way of mnemonic code, easily to grasp. Mnemonic code belongs to abbreviate code and is commonly used. It selects several critical letters from name and specification of code object to be code or part of code, with way of association to help memorize, easy to understand.

Fig2 are the expressions of operational main part and abbreviate code of basic action, two letters for body regions, 1~2 letters for basic action. See operational main part and abbreviate code of basic action in table2.

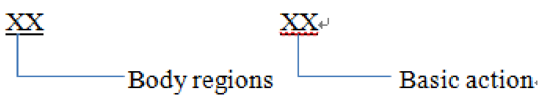


Fig. 2. Structure of Abbreviate Code

Table 2. Body Regions and Abbreviate Code of Basic Action

Body Regions						
Hand (H)		Foot (F)		Head (H)		Else(Es)
Left hand (LH)	Right hand (RH)	Left foot (LF)	Right foot (RF)	Eye (HE)	Mouth (HM)	Else(Es)
Basic Actions						
Reach (Re)	Grasp (G)	Whirl (W)	Press (Ps)	Pull Out (PO)	Tread (T)	Release (Rl)
Inspect (I)	Report (Rp)	Search (Sh)	Select (St)	Plan (Pn)	Hold (H)	Position (P)
Pre-position (PP)	Rest (Rt)	Unavoidable-delay (UD)	Avoidable-delay (AD)	Find (F)		

4.3.2 Sequence Code

In a certain phase of flight, it may involve operating a same object several times or repeating a basic action in the process of operating an object, thus, we should position the action by sequence code to describe the difference in the process. Sequence code is simple and commonly used. It put positive integers or alphabets to code objects. Simple to code, well to use, convenient to manage, easy to add, have no limits for the order of code object. However, it is hard to memory for the code itself does not provide any information about code object.

Fig3 is the expression of sequence code.

Fig. 3. Structure of Sequence Code

Among above, order number A-the M times operation for a same operational object in a certain flight phase, the span is 1~99;

Order number B-repeat a basic action N times in M times operation of operational object, the span is 1~99.

4.3.3 Condition Code

In order to keep the continuity of action, in case of action omission in the process of pilot's operation and reduction or avoidance of human errors, we classify the attributes of basic actions into 3 groups, which are denoted by condition code. To select the right code to satisfy the demand and combine them in a predefine order in use.

Condition code (condition combination code) is commonly used in the system of surface classification, entry in each surface is coded by its principle. Combining code in each surface as needed and order predefined in use. Code structure is flexible; it can be single or combination. It is convenient to change and expand for it is classified by surface. A condition code can reflect the whole recorded information features, as well as part features when it is partly used. It is elastic and specially suitable for dynamic compound quick inquiry, summing and other operations. Condition code has much more values in the design of information management system. However, the volume use rate of condition code is low.

The expression mode of condition code is shown as figure 4. Overlapping attribute of pilot's basic operational actions, the way of classification and concrete meaning of connect attribute are shown as table 3. If pilot's operational actions do not comply with the request of condition code, alarm can be used to draw pilot's attention.

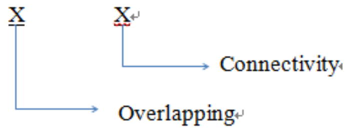


Fig. 4. Condition Code Structure

Table 3. Action attribute classification

	code	meaning
Overlapping	A	This action cannot overlap with other actions (except independent actions).
	B	This action can overlap with other actions.
	C	Totally independent, can overlap with any actions
Connectivity	1	Connect with AB category actions in required time
	2	Connect with any other overlapping actions in required time.

4.3.4 Position Code

As operational object is the object used in a specific flight phase, we should consider two aspects on operational object: flight phase and use object. Flight phases of Airbus are detailed and there is a lot of equipment in cockpit. The way of coding of all phases and equipment just start from 1 may increase the complexity of coding procedure. Therefore, we can apply layer code, breaking down from whole to part, to help code.

Layer code belongs to the kind that is commonly used for linear taxonomic hierarchies. It is a code that ordering by the relation of subordinate of object of classification and hierarchies. The code is cracked into several hierarchies, which is in accordance with the classification of code object. The code can definitely express categories of the classification objects, have strict subjection relations, and each hierarchy is meaningful. It is easy for computer to sum and summary as it has easy structure and high capacity. We should do some classifications before design, then to code and establish instruction of classification.

In addition, to indicate the shape features of equipment of cockpit from code and restrict action type, we can define the shape attribute by condition code; as a result, the position code of cockpit equipment consists of hierarchy code and condition code, which named compound code. Compound code includes two or more independent codes, flexible, easy to expand and mark part can be nimble used.

Therefore, we have a clear expression of the position of used operational object in the process of flight operation of pilot. The full expression is showed as Fig.5. Hierarchy code and operational equipment's condition code are expressed as table4;

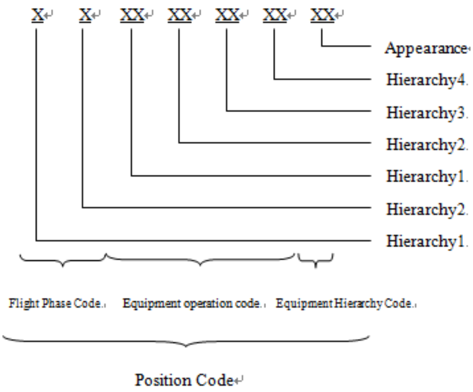


Fig. 5. Position Code Expression

Table 4. Flight Phase Layer Code

Flight phase							
Start up (1)	Taxi (2)	Take off (3)	Climb (4)	Cruise (5)	Descent (6)	Approach (7)	Landing (8)
Start up	Taxi	Take off	Climb	Cruise	Descent	Approach	Landing

Start up (1)	
Engine auto starts up (1)	Engine manual starts up (2)

Taxi (2)				
Propulsion push out (1)	Taxi and steering (2)	Brake check (3)	light control inspection (4)	Take-off fragmentary verification (5)

Take off (3)				
Propulsion set (1)	Taxi (2)	Rotation (3)	Minus propulsion altitude (4)	Acceleration altitude (5)

Climb (4)				
Climb monitor (1)	Speed change (2)	Acceleration climb (3)	Set barometer reference (4)	Terminate landing light (5)
Terminate seat belt indicator light (06)	EFIS options (07)	Check radio navigation page (08)	Transfer second flight plan (09)	Check best/highest altitude (10)

Cruise (5)			
Use of Flight Management System (1)	Gradient climb	Fuel monitor (3)	Approach preparation (4)

Table 4. *(Continued)*

Descent（6）				
Guide and monitor（1）		Mode transfer（2）		
Approach （7）				
ILS approach（1）	Imprecise approach（2）	Rotation approach（3）	Visual approach（4）	Precise approach（5）
Landing（8）				
Trim（1）	Call（2）	Dive（3）	Taxi（4）	Brake（5）

Table 3. Appearance Attribute Condition Code

Appearance Attribute				
Button（01）	Swtich(no cover) (02)	Cover switch（03）	Electirc switch (no cover) (04)	Electric switch（with cover）（05）
Electric switch(Spring attachment)（06）	Screw（07）	Select button（08）	selector（09）	keystoke（10）
Stick（11）	Wheel（12）	Handle（13）	Treadle（14）	Light（15）
Indicator light（16）	Monitor（17）	Indicator（18）	Hub（19）	Storage（20）
Pack（21）	Mask（22）			

If to adjust impulse force by using throttle lever in phase of climbing, throttle lever's position can be denoted as 4103100011. 8,9-bit code for 0 is the same with other classification median plus, no special significance. If transfer display mode in the phase of take-off taxi, then navigation mode button classification code is 3202031007.

4.3.5 Time Code

Generally speaking, the proportion of incidents that caused by human error in civil aviation incidents is more than 70%. Once a motion absent happened in the operational process, the consequence will be very serious, even huge aviation disastrous incidents. In case of omission of pilots' operational actions, the condition code requires pilot should complete a certain action at a regulatory time which may avoid action omission. The times involved in this process are the last time of action itself and time between this action and next action. We can use time code to denote time, that is, to add time property to code. The format of time code is: xxHxxMxxS, of which xx stands for number. Time code can be expressed as figure 6.

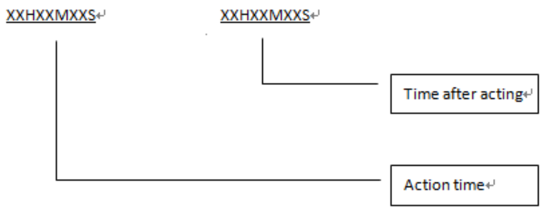


Fig. 4. Time Code Structure

4.3.6 Check Code

Generally, Check bit in the coding structure is the token of check code. In the process of structure designation of primary code, computing check bit attached to primary code with prescriptive mathematical method in advance, then the computer would calculate check bit in the same way when use the check code, it is definitely clear to see whether the input is right by comparing these two check bits. Cyclic redundancy code (CRC) is one of the most common used check bits. Cyclic redundancy code produced at sending terminal is sent to receive terminal by adding to the back of information bit. To check the received information with the same algorithm that produces cyclic redundancy code, if there is a mistake, then receive terminal makes a resend notice. CRC is good at error detection, low cost, and easily coded, while it is hard to detect errors made by 2 bits or more.

5 Conclusions

Empirical research is the main mode of human error research in the field of international civil aviation, which is well-directed but no universality. The proportion of accidents caused by human error still has not essentially dropped although there were a huge number of investments in the control of human error. The key factors should lie on the extremely sophisticated human factors involved, massy workload of preliminary obviation and exorbitant costs. Therefore, we can only do some afterward perfection and real-time diagnoses. However, real-time diagnosis requires that the computer should conduct real-time quantized comparison analysis while the scarcity of quantized description of pilot's fundamental operational actions restricted the speed of solving this issue.

To establish a standard of pilot's fundamental operational actions is the prerequisite to work out human error issue, which can be digitalized as well. As a result, this paper puts up with the opinion that we should make a study on features of civil airplane pilot's fundamental operational actions, trying to realize normalization expression and quantitative description by coding pilot's fundamental operational actions sequence, providing a new thought for computer automatic detection and distinguishment of human error, enriching existing warn and remind functions and offering references for error prevention and risk avoidance.

Part III

Military Applications

An Overview of Humans and Autonomy for Military Environments: Safety, Types of Autonomy, Agents, and User Interfaces

Michael J. Barnes[1], Jessie Y.C. Chen[1], Florian Jentsch[2], Elizabeth Redden[1], and Kenneth Light[1]

[1] U.S. Army Research Laboratory – Human Research & Engineering Directorate
Bldg 459, Aberdeen Proving Ground, MD 21005, USA
{michael.j.barnes,jessie.chen,kenneth.l.light}@us.army.mil
[2] University of Central Florida – Institute for Simulation & Training
Orlando, FL 32826, USA
florian.jentsch@ucf.edu

Abstract. The objective of this review is to extract design implications from multiyear US Army sponsored research investigating humans and autonomy. The programs covered diverse research paradigms: (a) effects of autonomy related to pedestrian safety during urban robotic missions, (b) supervision of multiple semi-autonomous robots assisted by an intelligent agent, (c) field investigations of advanced interfaces for hands- free and heads- up supervision of robots for dismounted missions and also investigations of telepresence, (d) effects of haptic control and stereovision for exploiting improvised explosive devices. Thirteen general design guidelines related to mixed initiative systems, pedestrian safety, telepresence, voice control and stereovision/haptic control are discussed.

Keywords: human-robot interaction, intelligent agent, military, autonomy.

1 Introduction

Modern combat is moving from the age of mechanized warfare, to the age of information, to the age of autonomy. Because of the increasing emphasis on the importance of using fewer operators to control multiple systems, some level of autonomy will be a necessity in future operations [1-3]. Autonomy covers a range of relationships between humans and intelligent systems - from systems that operate continuously without human intervention- to systems wherein specific behaviors can be performed without direct human control but humans must decide when to invoke the behaviors [4]. Autonomous systems planned for near-term military use are generally somewhere in between. The human maintains decision authority but there is only an occasional requirement for human intervention [1].

In this paper, we review human robot interaction (HRI) research funded by the Safe Operations for Unmanned Reconnaissance in Complex Environments

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 243–252, 2013.
© Springer-Verlag Berlin Heidelberg 2013

(SOURCE) Army Technology Objective (ATO). The purpose of the HRI research was to understand the effects of advanced interfaces and autonomy on the safety of humans operating in the same area as autonomous systems as well as to understand how autonomy affects overall mission effectiveness. The first two programs investigated different paradigms of human interactions with autonomy. The University of Central Florida [6-10] studied the effects of varying levels of autonomy (LOA) on safety and the Army Research Laboratory (ARL) researchers in Orlando investigated mixed-initiative autonomy using an intelligent agent [3, 5]. Both programs shared a common research goal of finding the sweet spots between human control and autonomous control.

We also review two additional ARL programs that investigated advanced interface concepts to improve soldier safety for dismounted operations. Researchers at Ft. Benning focused on field experiments evaluating interfaces that improved situation awareness (SA) and reduced workload for both autonomous and teleoperated conditions. These evaluations measured Soldier performance during field experiments using voice control and telepresence as means of maintaining decision authority while reducing control requirements [11-14]. ARL researchers at Ft. Leonard Wood investigated the utility of using a combination of stereovision and haptic arm manipulators to improve Soldier safety for defeating improvised explosive devices (IED) [16].

2 Level of Autonomy (LOA) and Soldier Safety

The research team led by Jentsch and involving Fincannon and others at the University of Central Florida (UCF) has a long history of supporting Army HRI programs. Some of their earlier work investigated the number of persons required to conduct reconnaissance missions for semi-autonomous robots, crew size for supervising robot to robot interactions, mixed unmanned aerial and ground vehicle operations, individual differences and effects of different training regimens [6, 7]. The two experiments summarized here directly addressed the question of Soldier safety and mission effectiveness as a function of LOA.

The initial experiment varied automation and degree of human involvement during simulations in a 1/35th scaled Iraqi city with similarly scaled robotic vehicles [10]. The researchers decomposed the robots' task into three components: (a) detect a possible significant object and making a decision to stop the robot, (b) identify the type of object, and (c) decide the type of action to be taken based on the current rules of engagement (ROEs). ROEs are command issued rules that permit Soldiers to conduct their missions under permissible guidelines; for the experiment, the ROEs were developed by the researchers and given to the participants before each session. In the *manual* condition, all tasks were performed by the human operator. In contrast, in the *autonomy* condition, even though all tasks were automated, operators were given the option of overriding the autonomy for tasks b and c. Finally, in the *collaborative* condition, task b was performed by the operator and tasks a and c were automated.

The collaborative condition took advantage of both the obstacle detection strengths of state-of-art autonomy and the human operator's perceptual strengths for target identification. UCF researchers referred to the latter as *perception by proxy*.

The most dramatic differences were evinced in task a, detection of possible targets and stopping the robot: 37% accuracy for manual, and 67% and 58% accuracy for the autonomy and collaborative conditions. This implies that an operator controlling a robot manually would find it very difficult to spot and react to unexpected events and that even imperfectly automated systems are safer than relying solely on the operator for this task. However, for synthesizing information (task b and c), a combined (collaborative) human and intelligent system decision was superior to either autonomous or manual control conditions except when the operator's workload was high. Thus, the experiment suggests that autonomy can enhance safety by detecting significant objects in the robot's path but that humans also can play an important role by being able to identify these objects (perception by proxy). However, choosing the correct ROE was best left to automation in this experiment. The results also suggest that the operator's role in overriding autonomy can be counter-productive and human intervention strategies for autonomy required further investigation.

The second experiment investigated LOAs in a similar tasking environment using the Mixed Initiative Experimental (MIX) computer simulation environment which allowed for more precise control of simulation parameters such as vehicular speed and pedestrian crossings [9]. Again the emphasis was on safety and LOA but the objective was to investigate the effectiveness of two operator intervention strategies. The test participants were told that pedestrians would transverse the robot's path under one of three LOAs conditions: fully autonomous (AU), management by consent (MBC) and management by exception (MBE) [2]. The AU system chose a response based on the ROEs (e.g., continue – intel suggests a dangerous situation) which were in turn based on the cover story for each simulated vignette. For the MBC conditions, the autonomy would always stop the robot and suggest a course of action based on the ROE which the operator had to consent to or change before executing. In contrast, in the MBE condition, participants could override the autonomous ROE but if they did not, then the autonomy-chosen ROE would be executed. The experimenters also varied autonomy reliability: either 60% correct or 90% correct ROEs, depending on the ROEs given to the operators for each vignette.

Overall, operators in the MBE conditions showed significantly superior performance (executing the correct ROE for safety), when compared to both AU and MBC. Figure 1 shows a significant interaction between reliability and LOA. MBE conditions allowed operators to take advantage of the accuracy of the AU conditions for high reliability conditions but also resulted in the operator being able to override poor AU decisions during low reliability mission segments. MBC operators showed a greater tendency to incorrectly second guess highly reliable autonomy. This interpretation is buttressed by the fact that operators tended to trust MBC conditions more than in AU ones, but their performance indicated that this trust was misplaced.

2.1 Design Implications

1. Autonomy can improve robotic safety by being able to respond to potentially dangerous situations better than humans in complex urban environments.
2. A possible strategy for overcoming autonomy limitations is developing hybrid systems that allow humans to do what they do best such as interpreting the significance of detected objects (perception by proxy).
3. Overall, in the UCF studies, the MBE LOA that allowed humans to override autonomous decisions was the most effective strategy compared to AU and MBC. MBE resulted in safer ROE decisions than did low reliability autonomy, but MBE showed only a minimal loss of decision accuracy when compared to highly reliable autonomy.
4. Trust as measured by the UCF subjective scale [9], was a poor predictor of performance; MBC was trusted more than autonomy but overall human performance was poorer during MBC mission segments than for either the MBE or AU conditions.

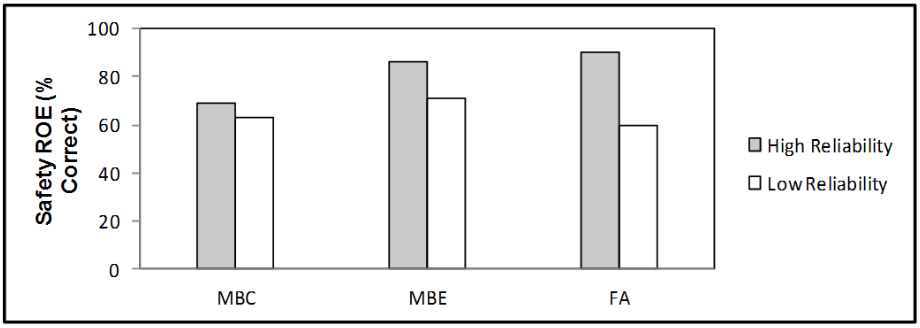


Fig. 1. Interaction between autonomy and reliability for predicting civilian safety; FA is same as AU in text and high is 90% and low is 60% reliability

3 Intelligent Agents for Supervisory Control: RoboLeader

Chen and colleagues simulated mounted combat situations wherein the operator was burdened with multitasking requirements as well as supervising multiple autonomous systems [3, 5]. Completely autonomous systems would not be practical in this environment because autonomy would limit tactical flexibility and pose safety risks while supervisory control of multiple autonomous systems introduced its own problems such as complacency effects and short-term memory limits [2]. Chen and her colleagues introduced the concept of employing an intelligent agent (RoboLeader).The agent would assess the current state of multiple systems, suggest algorithmic solutions, and execute them only when given permission by the operator. The advantage is that the operator could maintain SA and attend to other tasks and RoboLeader would act as subordinate crew member whose focus was the current state of the robotic assets.

The first experiment was a proof of concept for RoboLeader [5]. Humans working with RoboLeader were able to successfully re-route up to eight robots more rapidly than manual conditions when unexpected obstacles were encountered. In the second experiment, reliability of RoboLeader (60% and 90%) and type of errors (false alarm prone (FAP) and miss prone (MP)) were varied parametrically. Figure 2 shows the MIX simulation environment with a map display for robot rerouting, small windows showing views from the robots, a larger window for target identification, instruments panels, and a text window for RoboLeader to communicate with operators. Previous research indicated that high FAP alerts were more deleterious to overall performance than were MP weighted [2] systems because the cry wolf effect caused operators to lose faith in FAP alerts.

Fig. 2. Simulation scene showing the robot location map, four windows of robot camera views, and a larger display of scene as viewed by robot number 4

To the contrary, in the second study [5], the FAP conditions resulted in better overall scanning performance compared to the MP agents. Because of the layout of the embedded map displays, the locations of the robots could be checked easily for FP alerts in their experiment as opposed to previous research [5]. This made compliance to FP alerts efficient because of the relative ease of attentional switching. By way of contrast, the MP agent interfered with the operator's performance to a greater degree because participants in the MP conditions had to continually check data on the map thus drawing their attention away from the targeting displays. This conjuncture was supported by the SA measures indicating better performance on map related data for MP conditions again suggesting that operators focused on the map display to the detriment of their scanning performance. However, there were significant effects due to individual differences; for example, participants who were highly confident in their attentional control abilities had better overall MP performance. Also, higher levels of spatial ability and gaming experience had positive effects on performance.

In the third experiment, RoboLeader used more sophisticated algorithms to direct four robots to entrap a moving vehicular target. LOA was varied as well the addition of a visualization aid [3]. The purpose of this experiment was to assess the effectiveness of the RoboLeader agent for a more dynamic combat environment in which both the targets and the pursuing robots were moving. There were four LOA conditions: manual, hybrid, hybrid w. visualization, and fully automated w. visualization. For the hybrid condition, the human operator chose end-points for the pursuing robots and RoboLeader computed an optimal solution to entrap the moving target. The visualization aid showed how discrepant the robot's progress was from optimal solution to entrap the moving target. The full automation solution was correct 86% of the time whereas the hybrid solution (without visualization) was correct 96% of time, which although not statistically significant, suggested the possible advantages of human/autonomy collaboration found in the UCF studies. Visualization aiding had little impact on performance suggesting that even with partial autonomy, the raw data on the map display supplied sufficient information. Again, operators with higher levels of spatial abilities and more gaming experience showed improved performance. Both improved target acquisition while gamers were better at encapsulating the moving target [3, 5].

3.1 Design Implications

1. Intelligent agents acting as surrogate crewmembers are a potentially effective way of controlling multiple autonomous systems.
2. At a minimum, agent/human teams must have two characteristics (a) Operators must have final decision authority; (b) Agents must signal their intentions clearly.
3. Result in the above experiment [5] suggest that for agents that are not completely reliable, FAP (vs. MP) weighted alerts can be a relatively efficient means of alerting potential problems if the FA are easily checked and are not too numerous.
4. Individual differences in spatial abilities, attentional control, and gaming experience are important determinates of how well humans interact with autonomous systems.

4 Ft. Benning Field Experiments: Intuitive Interfaces

ARL researchers at Ft. Benning working closely with the Ft. Benning infantry school (and later the Maneuver Center of Excellence) evaluated dismounted HRI applications during realistic field experiments [11]. Their most recent research involved advanced interface designs to improve SA and free the soldier's hands and eyes for possible heads-up operations. Speech control of robotic assets has a number of distinct advantages. It is a natural way for Soldiers to interact with robots fostering a teaming relationship and it has the potential of hands- and eyes-free control. Redden and her colleagues conducted a number of studies evaluating the efficacy of voice for small robot control [12-13]. Their goal was to show that speech control could reduce the size of the controller by replacing the manual controller with a lighter, smaller speech system (Figure 3). The experiments were conducted using teleoperated robots but the results would transfers to operator interventions when necessary for semi-autonomous robots and for controlling miscellaneous functions such as menu selection. They found speech-based control exhibited the potential for benefits beyond controller size reduction. It decreased time and effort when performing multiple tasks simultaneously by allowing speech commands to be given for control of the robotic arm while at the same time maneuvering the robot using manual controls. However, the Soldiers had trouble with speech control if they had to control the pan and tilt of the robotic arm because the voice commands were discrete and lacked the fluid precise movements evinced by manual controllers.

Fig. 3. Earpiece used as a microphone

In the second experiment [13], they investigated the potential for using speech for multipurpose functions such as having the robot photograph IEDs or having the operator choose items on menu. When the operator was required to perform a secondary task, speech control improved multitasking performance because of the efficiency of speech for shared cognition. Similarly, when the operator had to access a menu related to taking a picture of a potential IED (e.g., "enlarge a picture"), speech control was significantly faster than manual control. However, actually taking a photo by maneuvering the robot was more efficient using manual control because maneuvering the robot is a continuous process. Also, the ARL researchers investigated intuitive vocabularies for the various tasks that Soldiers were asked to perform and the researchers developed a user centered lexicon for the experiment.

In a totally different domain, Elliott and her Ft. Benning colleagues [14] collaborated with researchers from the TNO laboratories in the Netherlands to evaluate telepresence techniques that give the feeling of actually being in the area the robot is viewing. The obvious advantage of telepresence is that an autonomous robot would be able to gather information for an area of interest (AO) without putting the Soldier in harm's way. Augmenting robot video is particularly important because previous research indicated that video feed from robots gives an impoverished view of the AO [15]. The telepresence augmentations included stereovision and a head mounted camera that the operator could use to scan the remote area in a fashion emulating actually being in the AO. In the first experiment, the tasking was relatively easy and target detection, SA or workload measurement differences were not significantly different from conventional interfaces. However, soldier participants preferred telepresence.

The second experiment contained more difficult detection tasks and the telepresence was augmented further using three-dimensional (3-D) audio cues to locate targets [14]. Telepresence was compared to a helmet mounted display (HMD) and a joystick to locate targets in a remote location. In addition, there was sound associated with each target. The 3-D audio augmentation resulted in improved performance compared to the HMD and joystick for workload reduction, speed of responses, and target identification. On the negative side, the telepresence equipment was bulky and not ideally suited for infantry operations.

4.1 Design Implications

1. It is extremely important to tailor speech commands to the target audience. Tailoring allows better retention and more efficient operation.
2. Speech control is quicker than manual control in situations that require secondary task accomplishment and also in situations in which the items that need to be accessed are embedded in menus.
3. Manual control is more effective than speech control for non-discrete tasks such as turning.
4. Although not currently configured for efficient infantry uses, telepresence has great potential for remote sensing of combat environments using robotic assets.

5 Ft. Leonard Wood: Soldier safety and IED Exploitation

The objective of this study was to explore the effectiveness of stereovision displays and haptic feedback for IED exploitation [16]. IEDs have proven to be a particularly deadly and difficult to detect weapon of terror that is being used against coalition troops and indigenous civilians. What make them doubly dangerous are the risks that Soldiers must take to defeat IEDs. Polaris Sensor Technologies and Harris Corporations working with ARL and the non-profit Leonard Wood Institute evaluated an interface suite to improve Soldier safety using a Talon ® robot to find, manipulate and destroy IEDs. Previous work by ARL had demonstrated the effectiveness and acceptance of stereovision for both navigation and arm manipulation for small robots IED operations. The current study incorporated not only a stereovision display but also a Harris controller that gave haptic feedback to the operator (Figure 4).

The nine participants performed navigation, search and arm manipulations tasks for scenarios that were indicative of US Army engineering, military police, and biochemical missions. There were statistically significant latency effects of view (3-D vs. 2-D) and non-significant trends for controller conditions favoring the 3-D-haptic combination. Similarly, there were significant effects for both these conditions for perceived workload reductions. The participants also endorsed both haptic and stereovision components individually and as a combined unit. In summary, the results indicated user acceptance as well as performance improvements for stereovision and haptic controllers especially when combined in the same interface.

Fig. 4. Harris Corporation's haptic controller

5.1 Design Implications

A combination of haptic feedback and stereovision shows promise for safely manipulating and defusing explosive and chemical devices using small robots.

6 Conclusions

We reviewed four experimental programs sponsored by the ARL whose purpose was to develop design guidelines for human roles in autonomy focusing on safety issues. The programs varied in their military context from safe operations for robots in an urban environment, to supervising multiple robots with assistance from an intelligent agent, to advanced interface evaluations at Ft. Benning, to IED exploitation at Ft. Leonard Wood. Two overarching trends manifested themselves: (a) Mixed–initiative systems where autonomous and human decisions making were shared but decision authority always remained with the human were superior to either full autonomous or manual control systems; (b) Enhanced sensors and interfaces such as 3-D visual/ audio, voice, and haptic systems can combine to give the human operator a realistic sense of immersive control resulting in improved safety and mission performance.

References

1. Barnes, M.J., Chen, J.Y.C.: Intelligent Agents as Supervisory Assets for Multiple Uninhabited Systems: RoboLeader. In: van Breda, L. (ed.) Supervisory Control of Multiple Uninhabited Systems-Methodologies and Enabling Human-Robot Interface Technologies (RTO-TR-HFM-170), pp. 16-1 – 16-14. NATO Science and Technology Organization, Neuilly sur Seine (2012)
2. Chen, J.Y.C., Barnes, M.J., Harper-Sciarini, M.: Supervisory Control of Multiple Robots: Human Performance Issues and User Interface Design. IEEE Trans. Sys., Man, & Cybern.–Part C: App. & Rev. 41, 435–454 (2011)

3. Chen, J.Y.C., Barnes, M.J.: Supervisory Control of Multiple Robots in Dynamic Tasking Environments. Ergonomics 55, 1043–1058 (2012)
4. Miller, C., Parasuraman, R.: Designing for Flexible Interaction between Humans and Automation: Delegation Interfaces for Supervisory Control. Human Factors 49, 57–75 (2007)
5. Chen, J.Y.C., Barnes, M.J.: Supervisory Control of Multiple Robots: Effects of Imperfect Automation and Individual Differences. Human Factors 54, 157–174 (2012)
6. Fincannon, T., Evans, A.W., Phillips, E., Jentsch, F., Keebler, J.R.: The Influence of Team Size and Communication Modality on Team Effectiveness with Unmanned Systems. In: Proc. 53rd Annual Meeting of the Human Factors and Ergonomics Society, pp. 419–423. Human Factors and Ergonomics Society, Santa Monica (2009)
7. Fincannon, T., Keebler, J., Jentsch, F., Evans, A.W.: Target Identification Support and Location Support among Teams of Unmanned System Operators. In: Proc. Army Sci. Conf. (2008), http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA504210&Location=U2&doc=GetTRDoc.pdf
8. Fincannon, T., Keebler, J.R., Jentsch, F., Phillips, E., Evans III, A.W.: Team Size, Team Role, Communication Modality, and Team Coordination in the Distributed Operation of Multiple Heterogeneous Unmanned Vehicles. J. Cogn. Eng. & Decision Making 5, 106–131 (2011)
9. Jentsch, F., Fincannon, T., Sellers, B., Keebler, J., Ososky, S., Phillips, E., Schuster, D.: Year 2 Final Report: Supervisory Control of Robots in Urban Operations ([Task Order 132] Safe Operation of Autonomous Robot Teams: Impact of Autonomy, Teaming, and Workload on Operator Trust and Performance). University of Central Florida, Orlando, FL (2012)
10. Jentsch, F., Barnes, M.J., Fincannon, T., Keebler, J., Ososky, S., Phillips, E., Schuster, D.: Safe operation of autonomous robot teams: Impact of autonomy, teaming, and workload on operator performance. In: Poster presented at the 2011 Army Research Laboratory Robotics Technical Assessment Board (TAB), Aberdeen Proving Ground, MD, August 08-09 (2011)
11. Redden, E.S., Elliott, L.R., Pettitt, R.A., Carstens, C.B.: Scaling Robot Systems for Dismounted Warfighters. J. Cogn. Eng. & Decision Making 5, 156–185 (2011)
12. Pettitt, R.A, Redden, E.S., Carsten, C.B.: Scalablity of Robotic Controllers: Speech-based Robotic Controller Evaluation (ARL-TR-4858). US Army Research Laboratory, Aberdeen Proving Ground, MD (2009)
13. Redden, E., Carstens, C., Pettitt, R.: Intuitive Speech Based Robotic Control (ARL-TR-5175). US Army Research Laboratory, Aberdeen Proving Ground, MD (2010)
14. Elliott, L.R., Jensen, C., Redden, E.S., Pettit, E.: Robotic Telepresence: Perception, Performance and User Experience (ARL-TR-5928). US Army Research Laboratory, Aberdeen Proving Ground, MD (2012)
15. Chen, J.Y.C., Haas, E.C., Barnes, M.J.: Human Performance Issues and User Interface Design for Teleoperated Robots. IEEE Transactions on Systems, Man, and Cybernetics–Part C: Applications and Reviews 37, 1231–1245 (2007)
16. Edmondson, R., Light, K., Bodenhamer, A., Bosscher, P., Wilkinson, L.: Enhanced Operator Perception through 3D Vision and Haptic Feedback. In: Proc. SPIE, Unmanned Systems Technology XIV, vol. 8373, pp. 25–27. SPIE, Baltimore (2012)

Autonomous Control in Military Logistics Vehicles: Trust and Safety Analysis

Nicole Gempton, Stefanos Skalistis, Jane Furness, Siraj Shaikh, and Dobrila Petrovic

Faculty of Engineering and Computing, Coventry University, Coventry, CV1 5FB, UK
{n.gempton,stefanos.skalistis,furness4,
s.shaikh,d.petrovic}@coventry.ac.uk

Abstract. Ground vehicles are increasingly designed to incorporate autonomous control for better performance, control and efficiency. Such control is particularly critical for military logistics vehicles where drivers are carrying sensitive loads through potentially threatening routes. It is imperative therefore to evaluate what role does autonomy play to help safety, and whether drivers trust autonomous control. In this paper we investigate the use of semi-autonomous vehicles used for military logistics and carry out human factors analysis to reflect on trust and safety issues that emerge from the driving of such vehicles.

Keywords: Military, Semi-autonomous vehicles, Logistics, Human Factors.

1 Introduction

Human failure is often a cause of accidents. Increasing the level of automation while useful in many cases, does not necessarily reduce the number of human failure related accidents. For such automation to be successful the human user must be aware of the automation and react to it appropriately. In some cases it is not possible to fully automate the desired behaviour and the system has to rely on humans exhibiting the right behaviour. Examples of such systems include Unmanned Aerial Vehicle (UAV) guidance [1], health care especially patient safety [2] and computer security [3].

Enhancing the driver experience of ground vehicles through increasing autonomy has been of interest for well over a decade now. Reduction in driver stress, freeing up limited attentional resources and improving road safety have been the major goals of this effort. However, autonomy brings with it a variety of other challenges that potentially risk road safety [4]. This could be due to sensor limitations, system design faults, error inducing design, or inadequate driver training; these certainly are some of the lessons learned from the introduction of autonomy in the aviation domain.

Adaptive cruise control (ACC) is an example of one mechanism introduced to provide safe distance control from a vehicle in front: once engaged, the vehicle operates in a typical cruise controlled fashion with the added feature of sensing the vehicle in front to adapt speed if it slows down or speeds up ensuring a minimal safe distance at

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 253–262, 2013.
© Springer-Verlag Berlin Heidelberg 2013

all times. Studies have demonstrated that such autonomy has the potential of causing delayed driver reaction [5], awkward handover and mode confusion, with up to a third of drivers having forgotten at some stage whether ACC was engaged or otherwise [6]. This has serious road safety risks and raises a question whether design of such mechanisms would ultimately be detrimental to the intended goal. In addition to the time on task effects, road conditions and terrain also significantly affect driver experience, and contribute to fatigue [7]; difficult terrains require more frequent driver interventions [8] in semi-autonomous vehicles.

In this paper we investigate the use of semi-autonomous vehicles used for military logistics and carry out human factors analysis to reflect on trust and safety issues that emerge from the driving of such vehicles. Section 2 describes some of the problems associated with this research. Section 3 describes our methodology in relevant detail. Section 4 presents the results of our experimental analysis. Section 5 presents a brief conclusion to the paper.

2 Motivation

There are potential economic, health and safety benefits of semi-autonomous vehicles in various industrial applications. Although the level of automation in mining is more advanced than many other domains, human oversight and control is still necessary given various factors such as legacy equipment, interoperability of hardware, and the ability to handle unforeseen circumstances. It is essential to use virtual engineering environments to model the vehicle and environment which can then be used to train drivers [9]. In addition to the known challenges, such as mode error where the driver cannot recall what state the system is in, there are particular challenges posed by semi-autonomous vehicles including

- handover between manual and automated control during a task [10, 11], which is critical as the driver needs to be able to judge when to reclaim control or otherwise,
- inadequate feedback from the vehicle to the driver [12], with the consequence that the system fails on drivers' expectations during a task and ultimately maximum benefit of the technology is not derived, and
- a fundamental change of task for the driver as their role changes from monitoring the situation to monitoring the situation and automation [6].

Most of the work done so far in this area has addressed such challenges in isolation and at a high abstract level [13-17], has studied vehicle sensor data [11], driver feedback [6] in a real or simulated environment, or performed physiological assessments [18,19]. The latter two strands of work entirely focus on driver perception and experience, borrowing from separate traditions of cognitive and physiological science. Our approach in this research is to conduct experiments involving master drivers (who are professionally trained to drive such vehicles) and analyse physiological and reaction time measures to assess how autonomy affects driver experience.

3 Experimental Setup

This section describes the methodology adopted. Section 3.1 describes the experimental design followed by Section 3.2 which gives the details of the virtual driving simulator implemented to carry out the experiments.

3.1 Experimental Design

The overall purpose is to assess the impact of autonomous control for drivers tasked with driving military logistics vehicles. Typical journeys are undertaken in convoys through hazardous and life threatening enemy territory. Such convoys could include a large number of vehicles, traveling over large distances at a slow speed, and may take up to 36 hours to complete a mission. The drivers are expected to keep an optimal distance between the vehicles. Autonomous control (in terms of cruise and lateral control) is expected to enhance convoy performance by maintaining an optimal speed, reducing fuel consumption, engine and brake wear, and reduce driver fatigue and cognitive load. This is particularly critical given drivers of driving through such journeys are likely to pose a difficult terrain, low visibility, high noise and roadside obstacles.

Experiments were designed as part of a virtual simulator where the drivers were asked to drive through a 3 hour journey and follow a vehicle in front as part of a convoy. Some experiments were designed to allow drivers to have manual control throughout the journey, whereas others were designed to (uniformly) incur periods of autonomous control (when control was explicitly taken over from the drivers) varying from 1 minute to up to 10 minutes. The journeys were designed to simulate ascending and descending routes, short periods of poor visibility and loud (bang) noises. Three drivers took part in a total of six experiments. The drivers had varying levels (7 to 24 years) of driving experience.

The total set of data collected from the experiments is given below

- Time (since start of experiment) in seconds
- Wheel input in terms analog turn of wheel
- Pedal input in terms of analog press of pedal
- Vehicle speed in metres per second
- Distance measured as the length of gap between two vehicles
- Time taken by the driver to attempt a stroop test
- Heart Rate (HR) in heart beats per minute (collected every 2 seconds)

The primary task performance measure is the reaction time of the drivers measured separately for cases of lateral track error and inter-vehicle gap in all scenario runs. This allows us to measure how quickly the driver is able to safely return the vehicle to the middle of the lane or within a safe distance of the vehicle in front. The distance travelled over the course of the experiment (3 hours in total for all experiments) is also evaluated.

Secondary task performance measures are influenced by the demands placed on the driver by the primary task of driving the vehicle (indicating the driver's spare cognitive capacity). The secondary task we used was a stroop test, as the test includes an implicit series of cognitive processes, including perception, attentional allocation, decision-making, and a motor response used to assess the drivers' cognitive load during the driving/monitoring task. The stroop test is one of the most widely used examples to study attention and cognitive control [20]. Our implementation of the stroop test used a body of text to pose a question (to judge colour-matching) displayed at a fixed location on the screen, occurring at regular intervals of 6 minutes asking the drivers to respond within three seconds (via paddles on the steering wheel).

To capture physiological responses, heart rate (HR) is a frequently used cardiovascular measure of mental workload in complex task environments [21]. Related to this is stress which is essentially a physiological response to the mental, emotional, or physical challenges that we encounter [22]. The drivers were asked to wear a HR monitor for 10 mins before commencing to get a baseline of their personal HR. A Garmin FR70 wireless HR monitor belt was used to capture readings at a 2s interval.

3.2 Driving Simulator

To develop a virtual driving simulator, the game engine Unity3D [23] was used to provide realism along with rapid development. The game was designed to simulate the cockpit of a heavy load vehicle following another similar vehicle at all times. As shown in Figure 1, a terrain containing a clear path was created. Both vehicles (being driven and followed) are similar in dimensions and capabilities. The vehicle interface provided the driver speed and temperature readings, along with warning signs to indicate proximity to the vehicle in front and autonomous control.

Fig. 1. Driver vehicle simulator screenshot

The vehicle shown in front was simulated to be driven autonomously and a similar controller was implemented for the vehicle being driven in front. The autonomous controller aimed to keep an average target speed of 40 Km/h using only 50% of the throttle in 0 degrees inclination/declination. While the percentage of throttle varied according to the degrees of inclination/declination, the average speed is maintained at 40Km/h at all times. In terms of lane alignment, it drove towards the centre of the lane predicting the vehicle's position 1 second ahead. The simulated terrain is a 4x4 (Km) circular plane containing hills and valleys. The driving lane runs throughout the terrain, and is 7m wide and approximately 50 Km long.

One challenge was to ensure that all data collected was optimized in terms of computational and memory usage. To address this, unnecessary rendering was avoided using thread-locking mechanisms. In order to achieve autonomous behaviour, in terms of lane alignment, KD-tree data structure was used keeping the computational effort as low as possible. The data structure provided the capability, given an arbitrary point, of finding the nearest stored point. This was used on the central points of the lane in conjunction with the vehicle's position. Throughout the experiments the drivers' input was monitored including the values generated by the wheel and pedals that the driver used to interface with the simulator. All data was collected at least every 200ms (except stroop tests, which were regularly scheduled).

To measure the deviation from the centre of the lane, a tolerance zone was defined to allow the driver to maneuver. From the centre (point) of the vehicle, deviation was defined as the centre of the lane and the projection of the centre of the vehicle on the 3D plane that defined the lane. The tolerance zone was defined as 30% of the total lane that is 15% left and right from the centre of the lane.

The physical properties of the vehicles are similar to a Tatra 8x8 vehicle and took into account dimensions, wheels (powered and steerable), engine torque curve, gear ratios and differential ratio.

Finally, a variety of development specific technologies were used to achieve a realistic simulation. The lane was developed using the Path tool [24] that allows the creation of an arbitrary path, in the form of single-columned triangular meshes, through the terrain. The autonomous system was developed using the UnitySteer library that provides an extensible framework to model low-level behaviours which can be combined to provide a sophisticated high-level behaviour. The low-visibility incident was implemented using the particle system and using textures from assets provided by Unity. All of the aforementioned tools are either provided by Unity3D or can be found on the Unity Store.

4 Analysis

We divide our attention to two different aspects of the results achieved from our experiments, including trust as an attribute of drivers' level of reliance and faith in autonomous control of the vehicle, and the impact of autonomous control on safety.

Of the total four experiments where drivers were asked to sit through a period of manual and autonomous driving, three out of four experiments showed that drivers had a slight increase in their average HR through periods of autonomous control. Figure 2 shows the average HR for the four experiments along with the baseline HR.

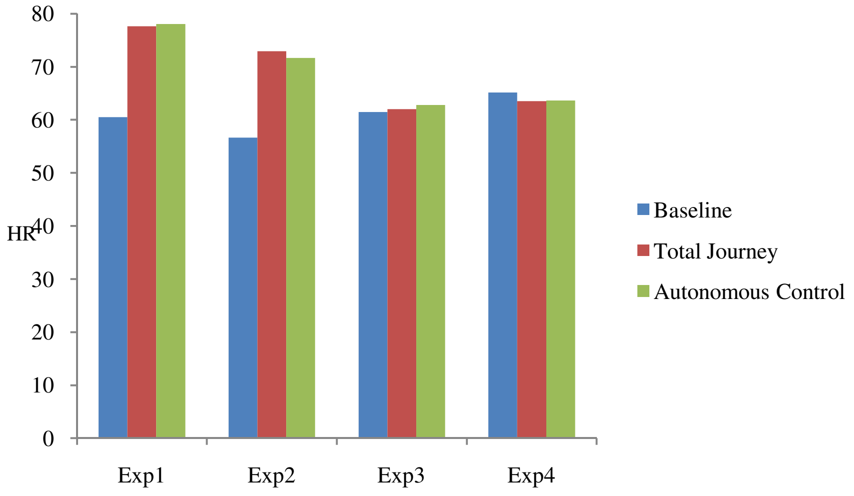


Fig. 2. The readings include baseline HR for drivers, along with their average HR measure for the total journey and autonomous control for the four mixed-mode experiments

While there are increases in the first and third experiment, the increase in the fourth experiment is negligible. When demonstrated for individual drivers, the impact of autonomous control on individual HR remains inconclusive.

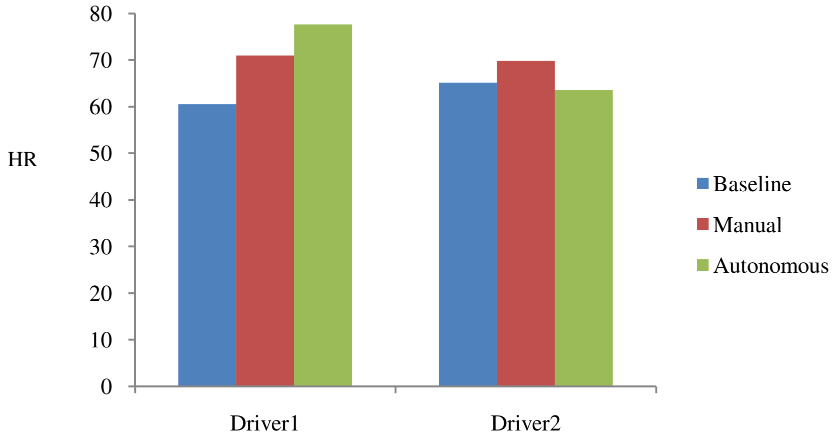


Fig. 3. Baseline HR for two drivers, along with their average HR measure for their manual (only) experiment and HR reading through autonomous for the mixed-mode experiments

Figure 3 above shows the difference in HR readings for the two drivers who sat through a pair of manual and mixed-mode experiments.

One aspect of this relationship worthy of note is the drivers' experience over a period of time of driving with autonomy. For the four mixed-mode experiments, the average HR readings were observed longitudinally for the autonomous only periods and the trends are shown in Figure 4 below.

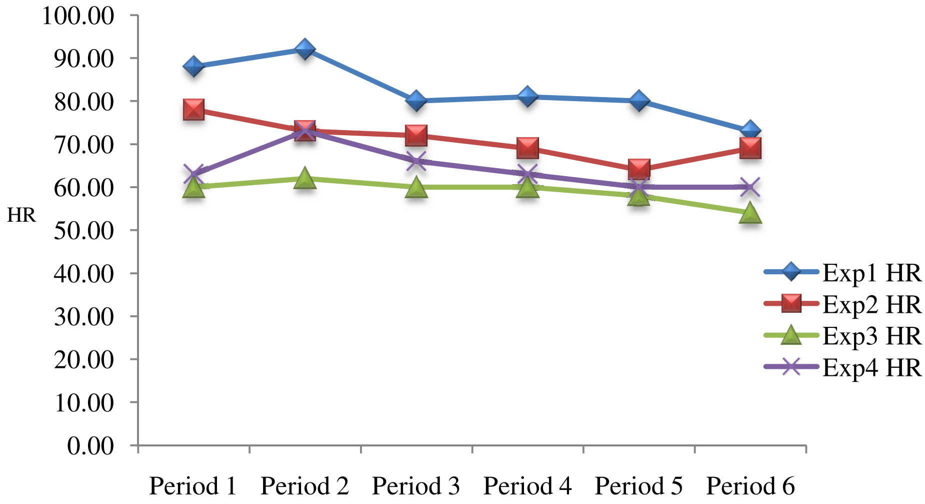


Fig. 4. There were six autonomous only periods during mixed-mode experiments. This graph shows the average HR readings longitudinally over the course of the experiment.

The four experiments show a notable decline in average HR readings as the experiment progress (except one experiment where in the last period the HR goes up). This reflects on drivers' increasing comfort in autonomous control as their experience of it progresses. Similar observation is found in the literature where drivers' trust in the vehicle autonomy is found to be improved over time [25-27].

An important part of safety is to ensure that any opportunity for driver error is avoided during the period of control transition. To assess how the drivers are affected in our study during periods of autonomous to manual handover, a handover period was identified to measure for HR. This included the period of 10 seconds before and after the actual handover (a total of 20 seconds) reflecting on how the HR readings change throughout this period. Figure 5 illustrates four experiments, of which three demonstrate an increase in average HR over the transition period. The maximum HR recorded in these three experiments is significantly higher than the average, demonstrating the high potential to cause errors during handover.

Moreover, the type of autonomy also influences the ease with which the operator could reclaim control. In a study of a shooting and planning style game, automation systems that removed the operator from task implementation, allowing the operator to plan ahead, proved the most disruptive when reclaiming control after automation failure [28]. The study shows that it was optimal to keep the operator in the loop with the system assisting in performing manual actions; removing the operator from manual control is detrimental to reclaiming control when needed [28].

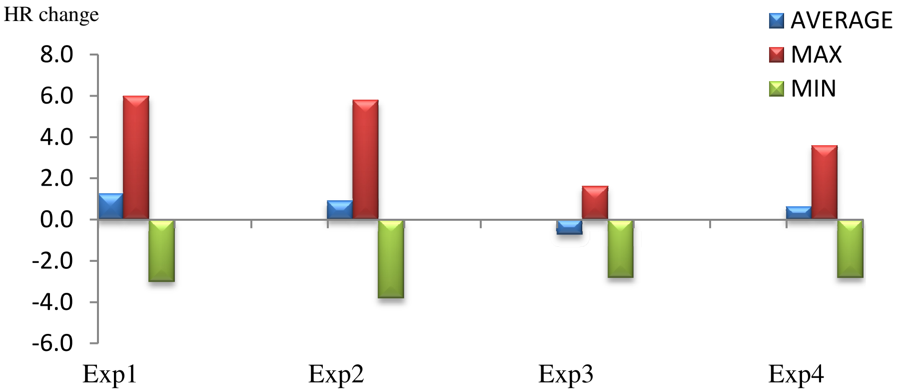


Fig. 5. For the four mixed-mode experiments average, high and low HR change is recorded throughout the handover period (10 seconds before and after the handover)

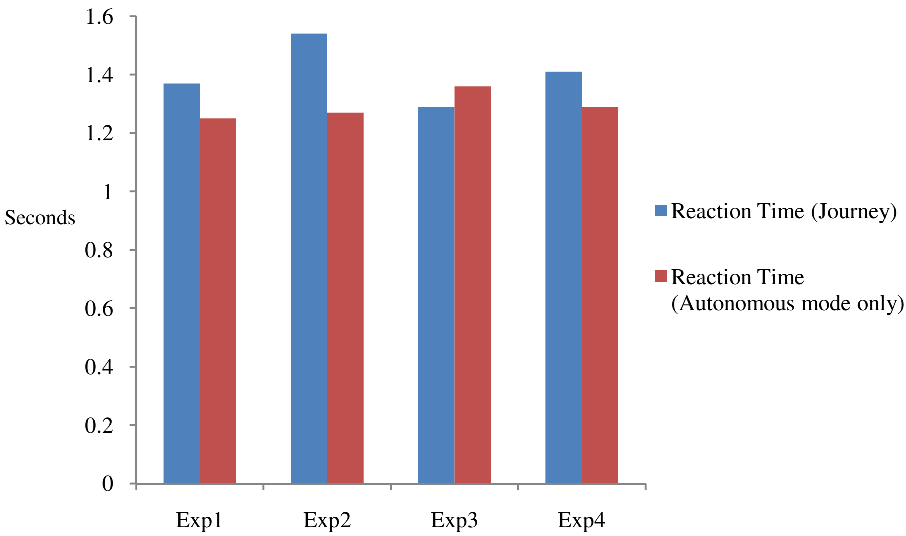


Fig. 6. In the four mixed-mode experiments, stroop tests were scheduled at regular intervals (during both manual and autonomous modes). The above readings show the reaction time for each experiment, both an average for the total journey and average reaction times for test conducted during autonomous mode only.

One factor that directly corresponds to safety is drivers' reaction time, which is critical if they are to respond to unexpected situations. Reaction time depends on the ability to process information within the driving environment, interpret that information to choose an action, and then react to the situation. Other studies show that when the driver role changes to a monitoring task, operator vigilance is reduced leading to loss of situational awareness and potential skill decay [29]. Our experiments however demonstrate a notable improvement in reaction time when drivers are driving through autonomous control mode. In Figure 6 above, three out of four drivers show a notable improvement over their average journey reaction times.

5 Conclusion

The use of actual master drivers from the military domain with field experience provides for valuable insight into how autonomy could play a significant role in establishing trust and improving on safety aspects of semi-autonomous vehicles. This work contributes to the field with empirical evidence obtained on the basis of carefully planned experiments. Further work is planned to use the empirical data to derive drive behavioural models for autonomous vehicles.

Acknowledgements. This work has been funded by the Ministry of Defence (UK) under the Unmanned Distribution Capability (UDC) initiative.

References

1. Chen, J.Y.C.: UAV-guided Navigation for Ground Robot Tele-operation in a Military Reconnaissance Environment. Ergonomics 53(8), 940–950 (2010)
2. Carayon, P. (ed.): Human Factors and Ergonomics in Health Care and Patient Safety. CRC Press (2011)
3. Cranor, L.: A framework for reasoning about the human in the loop. In: Proceedindgs of the 1st Conference on Usability, Psychology and Security, pp. 1–15. USENIX (2008)
4. Stanton, N.A., Marsden, P.: From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. Safety Science 24, 35–49 (1996)
5. Vollrath, M., Schleicher, S., Gelau, C.: The influence of cruise control and adaptive cruise control on driving behaviour a driving simulator study. Accident Analysis & Prevention 43(3), 1134–1139 (2011)
6. Larsson, A.F.: Driver usage and understanding of adaptive cruise control. Applied Ergonomics 43(3), 501–506 (2012)
7. Oron-Gilad, T., Ronen, A.: Road characteristics and driver fatigue: A simulator study. Traffic Injury Prevention 8(3), 281–289 (2007)
8. Scholtz, J., Antonishek, B., Young, J.: Operator interventions in autonomous off-road driving: effects of terrain. International Conference on Systems, Man and Cybernetics 3, 2797–2802 (2004)
9. Moore, P., Ujvari, S., Pu, J., Lundgren, J.O., Xie, C.: Intelligent semi-autonomous vehicles in materials handling. Mechatronics 9(7), 881–892 (1999)
10. Larsson, A.F.: Issues in Reclaiming Control from Advanced Driver Assistance Systems. In: Eu. Conf. on Human Centred Design for Intelligent Trans. Sys., vol. 2, pp. 557–564 (2010)
11. Viti, F., Hoogendoorn, S., Alkim, T., Bootsma, G.: Driving behavior interaction with acc: results from a field operational test in the netherlands. In: Intelligent Vehicles Symposium, pp. 745-750. IEEE (2008)

12. Norman, D. (ed.): The Design of Future Things: Author of The Design of Everyday Things. Basic Books (2007)
13. Rushby, J.: Using model checking to help discover mode confusions and other automation surprises. Reliability Engineering & System Safety 75(2), 167–177 (2002)
14. Curzon, P., Ruksenas, R., Blandford, A.: An approach to formal verification of human computer interaction. Formal Aspects of Computing 19(4), 513–550 (2007)
15. Shaikh, S.A., Krishnan, P., Cerone, A.: A Formal Approach to Human Error Recovery. In: Preproceedings of the 2nd Int. Workshop on Formal Methods for Interactive Systems (FMIS 2007), pp. 101–115, Technical Report RR-07-08, Queen Mary, Univ. of London (2007)
16. Cerone, A.: Closure and Attention Activation in Human Automatic Behaviour: A Framework for the Formal Analysis of Interactive Systems. In: Proceedings of the Fourth International Workshop on Formal Methods for Interactive Systems (FMIS), Electronic Communications of the EASST (ECEASST), vol. 45 (2011)
17. Shaikh, S.A., Krishnan, P.: Framework for Analysing Driver Interactions with Semi-Autonomous Vehicles. Electronic Proceedings in Theoretical Computer Science (EPTCS) 105, 85–99
18. Shelton-Rayner, G.K.: Quantifying exposure to psychological and physiological stress and automotive design. PhD thesis, Coventry University (2009)
19. Shelton-Rayner, G.K., Mian, R., Chandler, S., Robertson, D., Macdonald, D.W.: Quantitative physiological assessment of stress via altered immune functioning following interaction with differing automotive interface technologies. International Journal of Human-Computer Interaction 27(9), 900–919 (2011)
20. Gwizdka, J.: Using Stroop task to assess cognitive load. In: Proc. of the 28th European Conf. on Cognitive Ergonomics, pp. 219–222 (2010)
21. Nikolova, R., Collins, S.: Effect of Workload and Stress on Operator Functional State. NATO Science Series Sub Series I Life and Behavioural Sciences 355, 303–312 (2003)
22. Sun, F.-T., Kuo, C., Cheng, H.-T., Buthpitiya, S., Collins, P., Griss, M.: Activity-Aware Mental Stress Detection Using Physiological Sensors. In: Gris, M., Yang, G. (eds.) MobiCASE 2010. LNICST, vol. 76, pp. 282–301. Springer, Heidelberg (2012)
23. Unity – Game Engine, `http://www.unity3d.com`
24. Road/Path Tool, `http://www.sixtimesnothing.com/road-path-tool/`
25. Stanton, N.A., Young, M.S.: Driver behaviour with adaptive cruise control. Ergonomics 48(10), 1294–1313 (2005)
26. Kazi, T.A., Stanton, N.A., Walker, G.H., Young, M.S.: Designer driving: drivers' conceptual models and level of trust in adaptive cruise control. International Journal of Vehicle Design 45(3), 339–360 (2007)
27. Dickie, D.A., Boyle, L.N.: Drivers' Understanding of Adaptive Cruise Control Limitations. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 53 (23), pp. 1806–1810. SAGE Publications (2009)
28. Endsley, M.R.: Level of automation effects on performance, situation awareness and workload in a dynamic control task. Ergonomics 42(3), 462–492 (1999)
29. Endsley, M.R., Kiris, E.O.: The out-of-the-loop performance problem and level of control in automation. Human Factors: The Journal of the Human Factors and Ergonomics Society 37(2), 381–394 (1995)

Transparency of Military Threat Evaluation through Visualizing Uncertainty and System Rationale

Tove Helldin, Göran Falkman, Maria Riveiro, Anders Dahlbom, and Mikael Lebram

University of Skövde, Box 408, 541 28, Skövde, Sweden
firstname.lastname@his.se

Abstract. Threat evaluation (TE) is concerned with determining the intent, capability and opportunity of detected targets. To their aid, military operators use support systems that analyse incoming data and make inferences based on the active evaluation framework. Several interface and interaction guidelines have been proposed for the implementation of TE systems; however there is a lack of research regarding how to make these systems transparent to their operators. This paper presents the results from interviews conducted with TE operators focusing on the need for and possibilities of improving the transparency of TE systems through the visualization of uncertainty and the presentation of the system rationale.

Keywords: Automation transparency, threat evaluation, trust, uncertainty visualization, system rationale.

1 Introduction

In military settings, it is crucial that the decision makers, such as fighter pilots and command and control operators, have a good view of the current and evolving situation to be able to defend important assets, civilians and military personnel. As such, TE is an ongoing process that involves the determination of the intent, capability and opportunity of detected targets, i.e. what the targets intend to do, if they have sufficient resources to inflict harm and whether the context makes it possible for the targets to carry out their tasks [1-3]. Further, the TE process includes a classification of potential threats into categories (such as high/medium/low threat) along with a prioritization of these according to how much threat they pose to the defended asset(s) [2, 4]. However, in data-intensive and time-critical situations, military operators might find it difficult to perform these threat evaluation activities fast and with high quality, and incidents have occurred. For example, in 1988 the US Navy Cruiser USS Vincennes launched two missiles against what they believed was a hostile Iranian Air Force military aircraft in attack mode, when it was in fact an Iranian passenger airliner [5]. Another incident took place in 2003 where one British Royal Air Force Tornado GR4A was misclassified as hostile and shoot down by a US Army Patriot Surface-to-Air-Missile [6]. Several causes have been listed as contributing to these incidents – inexperienced crew, lack of time, insufficient data quality, failure of the

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 263–272, 2013.
© Springer-Verlag Berlin Heidelberg 2013

battle management system, classification criteria, rules of engagement and malfunction of the identification system [6, 7]. To avoid additional incidents, further development of computerized support has been recognized as crucial for aiding military operators perform their tasks (see for instance [2], [8, 9]). However, reports from various domains indicate that there are positive and negative effects of implementing computerized support systems, such as improved situation awareness and decreased workload on the one hand, but skill degradation and complacent behaviour on the other (see for instance [10-12]). As such, it is important to develop systems that provide appropriate support for the operator, that are easy to use and that the operator can trust. As discussed by [13], the possible severe consequences of making a wrong decision related to threat evaluation can result in that the operator becomes overly concerned with the risks associated with a course of action and, as such, will unlikely accept a system recommendation if he/she does not fully understand it or if the recommendation is different from the ones already considered by the operator. Thus, it is of utmost importance that the system generates and visualizes high quality recommendations along with a rationale for the recommendations generated [13]. This characteristic is termed *system transparency*, which implies that the system operators are able to easily use and understand how a system works (see [14, 15]). Important to note is that system transparency can take many forms – it can for example be associated with the underlying reasoning models used by the system, the input/output data, how to use the system as well as the system performance. To achieve improved system transparency, several researchers have investigated the effects of presenting additional meta-information regarding the inferences performed by the support systems on operator performance, such as the uncertainty associated with any system generated recommendation or by providing explanations to the behaviour of the system (see for instance [16-18]). For example, in the study presented in [18], it was revealed that the presentation of system confidence information aided the operators to more appropriately calibrate their trust in the system used. In the TE domain, where military operators have to rapidly analyse large amounts of possibly uncertain and contradictory data from multiple sources, we believe that the visualization of uncertainty and the system rationale can be of great importance for improved operator trust in the system used and better operator performance.

As a first step towards improving the transparency of TE systems, we argue that TE systems must include visualizations of the uncertainty associated with relevant pieces of threat information, as well as provide a rationale for the threat values suggested by the system. We believe that this is a challenging domain for such visualizations due to the massive amounts of information available and the fast-paced decision situation for the operator, making it imperative to investigate suitable information visualization representations. Further, it is believed that the importance of uncertainty visualization within this domain will become crucial due to improvements in sensor capabilities, making it possible to detect objects at farther distances, however with varying degrees of information quality, which must be conveyed to the operator. We anticipate that the visualizations of uncertainty and recommendation rationale will positively affect the performance of operators of TE systems, as well as lay a foundation for better trust in the TE systems used. To investigate this, the current paper presents interviews conducted with operators of TE systems where their perceived needs for system transparency were collected and analysed. The paper is organized as follows: section 2 provides background material related to threat evaluation and support system transparency. Section 3 presents the study setup and results, whereas section 4 summarizes and discusses the results obtained. Finally, section 5 outlines the conclusions drawn from the study as well as our ideas for future work.

2 Background

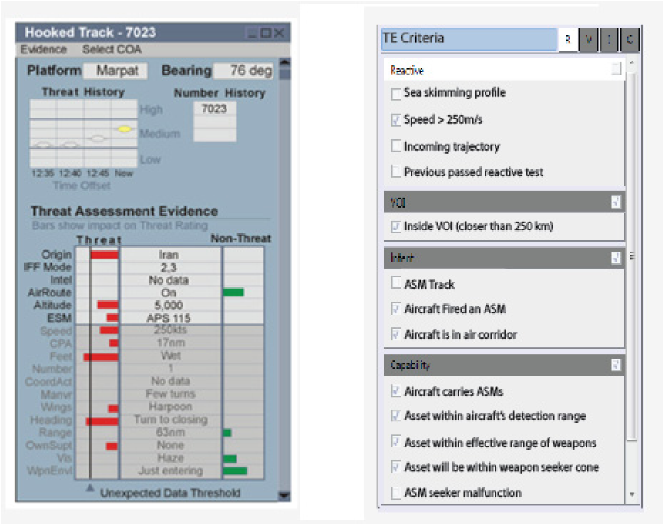
To evaluate threats, military operators have to analyze large amounts of data from multiple sources, which might be uncertain and contradictory as well as valid for only a short period of time. Uncertainty is prevalent due to, for example, the unpredictability of the behavior of the threats as well as due to imperfection of the information sources available. Time is also an important factor in the TE process due to the fast-paced tempo of the decision and action activities that must be performed. As stated by [8], the process of TE is cognitively challenging under usual conditions, and possibly worse under extreme conditions due to factors such as time, stress, short-term memory requirements and multi-tasking demands. In [8], the results from a task analysis performed together with military operators working in the maritime domain regarding how they perform their threat evaluation tasks is presented. Their results indicate that during a threat evaluation process, the operators formulate a hypothesis regarding the observed object by activating a threat profile that corresponds to the type of object that has been observed. Depending on the type of object and its behaviour, the operators can draw conclusions regarding the object's intent, capability and opportunity to harm the defended asset(s). During this evaluation process, it was found that the operator searches for evidence that confirms to the activated threat profile, but often fails to accommodate evidence that contradict the hypothesis. This is in line with research regarding naturalistic decision making (see [19]), where studies of human decision making in stressful, dynamic and uncertain environments have been conducted and where it has been found that decision makers perform poorly when it comes to recognizing contradictory information. Furthermore, due to the challenging decision situation where military operators must make decisions fast, the operators are prone to make errors and are often susceptive to biases [2]. As such, it is believed that transparency of TE support systems in terms of the reliability of the recommendations generated will have a positive effect on operator performance [2].

A feeling of system transparency can be achieved during operator training with the system, but also during system usage through explaining the inferences made and the recommendations generated. Such explanatory capacity of a system has been highlighted as crucial for appropriate confidence in and usage of a system (see for example [20, 21]). However, explanatory functions are seldom incorporated into automated systems. In [18], it is argued that support systems often present the operator with a diagnosis or solution to a problem with little or no explanation or qualification, which places the operator in a position where he/she must fully accept the system's advice or perform the entire decision making process on his/her own, often under time pressure and in critical situations. Yet, there are exceptions. For example, in the study presented in [22], it was revealed that the presentation of information regarding the performance of the system used to aid novice and expert command and control operators respond to battlefield threats through resource allocation indeed positively influenced the operators' trust and usage of the system. When indicated that the system was not performing reliably, the operators in the study appropriately adjusted their trust in the system through discarding the system recommendations. However, just like trust, there is a balancing act between too much and too little system transparency. A fully transparent system can result in operator information overload, causing high workload and reduced situation awareness [23]. However, at the other extreme, a system that does not provide information or adequate feedback regarding the system behaviour may lead to reduced workload, but at the cost of transparency which in turn can result in diminished situation awareness [23]. Hence, the result of not analysing appropriate ways of achieving system transparency might have severe effects on operator performance and trust.

Two studies were found during a literature search where interface and interaction guidelines for TE systems are discussed (see [2, 8] for more information). Amongst these guidelines, the importance of providing the TE operators with the cues used in the automatic reasoning of the system together with an explanation of the system generated recommendations is highlighted. For example, in the prototype presented in [8], evidence and counter-evidence with regard to the estimated threat level of an object is graphically represented and in [2], the TE criteria fulfilled by a detected object is presented to the operators (see Fig. 1). However, neither of these studies investigated ways of visualizing uncertainty or how to improve the transparency of the TE system used through providing a more detailed system rationale in terms of the degree to which a detected object fulfills the TE criteria. We argue that such visualizations could further improve the transparency of TE support systems. The next section presents results from interviews conducted together with operators of TE systems, focusing on the needs for improving the transparency of TE systems as well as possible ways of achieving such transparency.

3 Study

To investigate possible ways of making TE support systems transparent, semi-structured individual interviews were conducted with four active operators of air defense TE systems. The participants were between 25–35 years old and were all well accustomed to using TE systems. In this paper, we report on the findings regarding the necessity and willingness of the participating TE operators to recognize and interact with the uncertainty associated with the data presented and the rationale behind the system recommendations.



a) b)

Fig. 1. Parts of TE screenshots displaying a) evidence and counter-evidence of a threat rating of an object and b) the criteria fulfilled in the threat evaluation (from [8] and [2])

3.1 Dealing with Uncertainty

The participants in the study argued that they constantly perform actions to decrease the uncertainty associated with the current and evolving situation, such as uncertainties regarding the type of object detected and its altitude. However, such uncertainty is not displayed to the operators. Instead, the operators have to constantly analyze incoming data, redirect the sensors used and communicate within the military teams to create a good situational picture with little/no uncertainty. Yet, uncertainty is often prevalent due to countermeasures used by the adversary, making the reliability of the sensor data non-sufficient for performing evaluations with adequate quality. To more easily identify the sources of uncertainty, the participants in the study were positive towards being presented with the uncertain parameters, especially those that have a large impact on the calculated threat value, in close proximity to the target symbol on the map view of the TE system. In the same spirit, the participants argued that being presented with a summary of the variables that fulfill/do not fulfill the evaluation criteria would aid them to make better estimations of the evaluations performed. This is in line with the research presented in [8] where the importance of highlighting evidence and counter- evidence associated with the threat level of an object was reported. However, to not overwhelm the operators with too many variables presented simultaneously, the participants argued that the visualization of uncertain parameters with non-crucial impact on the generated threat value should not be default, but be easily extracted through navigating the menus of the TE system. As such, the participants argued for the default presentation of both an overview of the uncertainty associated with a calculated threat value on the map view of the TE system, including the uncertain parameters that have a large impact on the calculated threat value, as well as an optional, easily accessible detailed presentation of the uncertainty associated with the participating variables in the threat evaluations.

During the interviews, it became evident that it is not only large uncertainties that are of importance when making decisions in the TE domain, but also smaller ones. If the uncertainty regarding an important parameter used in the threat evaluation is too large, the operators will most often be forced to perform time-consuming data collections and analyses before making a decision due to the risk of fratricides. However, in the case of small uncertainties, the operators will have the opportunity to rely on their own experience and analytical expertise in order to deduce the most appropriate decision and action. As such, the visualization of small uncertainties must be further examined in the TE domain.

3.2 Transparency of TE Model

There are various models that can be used for TE (see [24] for more information). A common technique is the one based on rules where threats are prioritized according to their level of rule and parameter threshold fulfillment, for example "if SPEED=X and ALTITUDE=Y then THREAT=Z". Further, these rules and parameters can be weighted differently according to their presumed importance for the threat level of an object in different situations. In [2], the fulfillment of TE criteria used as a base for the evaluations performed is indicated by marking the corresponding "criteria fulfillment box". However, to which degree these criteria influence the threat evaluation is not presented – for instance, are the criteria met by far or just by passing the threshold value and how much does a single parameter influence the resulting threat value? During the interviews performed, the participants argued that such indication would aid them to more appropriately calibrate their trust in the system used as well as improve their understanding of the inner workings of the system. Furthermore, the participants in the study argued that the indication of rule/criteria threshold fulfillment and knowledge of different parameter weights in the evaluation process would also be useful when testing the sensitivity/insensitivity of different TE rule and parameter setups, which would further improve their understanding and usage of the TE system. As argued by one of the participants, knowledge of this kind could, for example, aid TE operators to understand quick fluctuations in threat values as well as identify where uncertainties exist in order for the operators to make their own estimation of the evaluations performed.

4 Discussion

The TE operators interviewed were positive towards having interactive representations of the uncertainty associated with the data as well as being presented with more detailed information regarding the system rationale behind the threat values automatically assigned to the targets. As such, the interviews have provided initial results regarding how the transparency of TE systems can be improved based on the systems

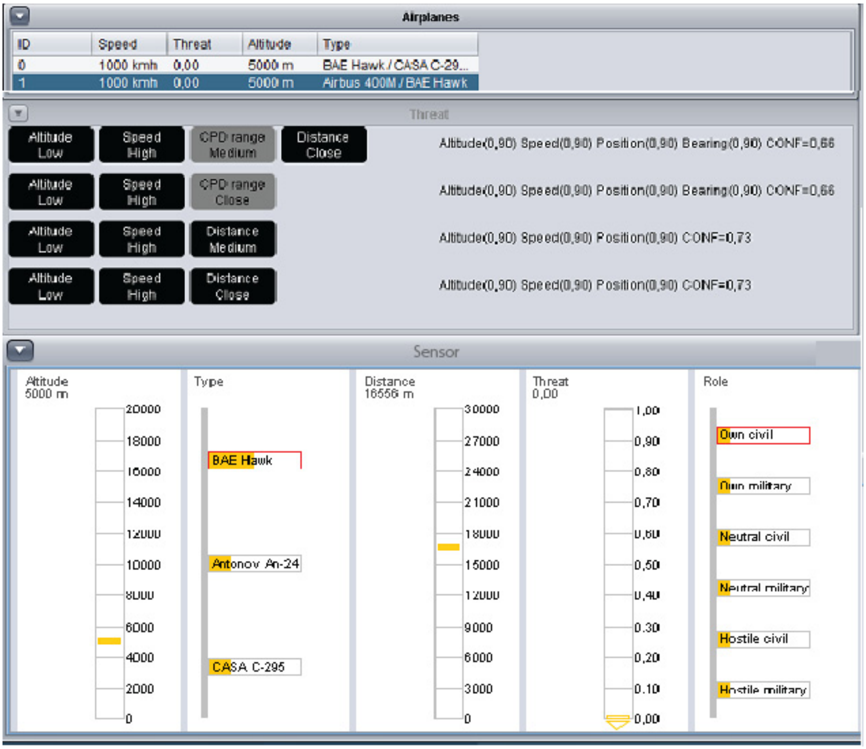


Fig. 2. A first version of the TE prototype, depicting possible in-depth visualizations of uncertainty through using intervals as well as TE criteria fulfillment with associated confidence measures

described in [8] and [2]. However, more research is needed regarding suitable ways of representing the uncertainty associated with the automatic evaluations performed. Several approaches toward visualizing uncertainty have been proposed in various domains. In [25], different techniques are listed, such as to use colors, edge crispness and transparency of graphical variables. However, as argued by [26], these techniques have not been empirically evaluated and it is not obvious how to select between them or how to integrate them. As such, it is imperative to investigate different representations in the domain of interest to be able to choose a visualization that appropriately reflects the uncertainty associated with the data used and that aids the operators make sense of the data. One representation of uncertainty that seems promising in the TE domain is the one based on intervals. Intervals were used to convey uncertainty within the engineering domain in a study performed by [17], with good results. We believe that this representation can be suitable in the military domain as well due to the importance of making it possible for the operators to receive a quick overview of the data collected, as well as to be able to quickly understand (on a more detailed level) where the uncertainty stems from, due to the importance of having a good awareness of the situation.

Further investigations are also needed in relation to finding suitable interaction formats between the TE operators and the models used for TE. Several TE models have been proposed, such as rule-based, fuzzy logic and graphical models (see [27]), which all have different possibilities and delimitations in their interactive capabilities. As such, empirical investigations together with TE operators are needed in order to establish appropriate interactions with and visualizations of the TE models. For example, should threshold values for incorporated TE criteria be determined by establishing fixed values that must be met, or by having gradually overlapping values, such as in the case with fuzzy logic? Or is it the cause-and-effect relations between the variables that provide the most valuable information about the TE process to the operators, as possibly represented through graphical models? It may also be the case that the suitable interaction and visualization formats depend on the planned usage of the TE system, for example during training sessions, system setups and as well as during battles.

As a first step towards investigating suitable visualization and interaction formats requested by TE operators, a TE prototype has been implemented. Fig. 2 depicts the first version of this prototype where the suggested interval representation for conveying the uncertainty associated with the individual parameters used in the TE process is illustrated. Furthermore, the figure shows one possible way of indicating which TE criteria that have been fulfilled as well as the confidence associated with these criteria.

5 Conclusions and Future Work

The results from the interviews carried out together with TE operators indicate that the transparency of TE systems can be improved in several ways. This paper has focused on increasing the TE system transparency through the visualization of uncertainty and through proving the operators with additional details regarding the rules and parameters used in the evaluations. Future work will include a series of evaluations of different ways of visualizing the uncertainty associated with the results generated by the TE support system. These evaluations will focus on which types of uncertainties should be presented to the operators, at which level of abstraction and in which situations. Future work will also include an evaluation of different recognized TE models in terms of their interactive and transparent capabilities and delimitations. Further developments of the TE prototype will be carried out and used in simulator settings together with TE operators in order to evaluate the effects of different visualizations and interactions on operator performance, decision making and trust in the TE system. A similar study will also be conducted in the fighter aircraft domain, where the impact of different visualizations of uncertainties and system rationale on the fighter pilots' trust in the system and performance will be evaluated.

Acknowledgement. This research has been supported by Vinnova through the National Aviation Engineering Research Program (NFFP5- 2009-01315), Saab AB, the University of Skövde and the Swedish Knowledge Foundation under grant 2010/0320

(UMIF). We would like to thank Håkan Warston (Saab AB, Electronic Defence Systems) and Klas Wallenius and Kjell Davstad (Saab AB, Security and Defence Solutions) for their valuable feedback. We would also like to thank the participants in the study.

References

1. Steinberg, A.N.: An Approach to Threat Assessment. In: Shahbazian, E., et al. (eds.) Harbour Protection Through Data Fusion Technologies, pp. 95–108. Springer, Netherlands (2009)
2. Irandoust, H., et al.: A mixed-initiative advisory system for threat evaluation. In: 15th International Command and Control Research and Technology Symposium: The Evolution of C2, Santa Monica, California, USA (2010)
3. Roux, J.N., van Vuuren, J.H.: Threat evaluation and weapon assignment decision support: A review of the state of the art. ORiON 23, 151–187 (2007)
4. Paradis, S., et al.: Threat evaluation and weapons allocation in network-centric warfare. In: 8th International Conference on Information Fusion, Philadelphia, Pennsylvania, USA (2005)
5. Smith, C., et al.: Decision Support for Air Warfare: Detection of Deceptive Threats. Group Decision and Negotiation 13, 129–148 (2004)
6. British Ministry of Defence: Military Aircraft Accident Summary, Aircraft accident to Royal Air Force Tornado GR MK4A ZG710 (2004)
7. Fisher, C.W., Kingma, B.R.: Criticality of data quality as exemplified in two disasters. Information & Management 39, 109–116 (2001)
8. Liebhaber, M., Feher, B.: Air threat assessment: Research, model, and display guidelines. In: Proceedings of the 2002 Command and Control Research and Technology Symposium (2002)
9. Nguyen, X.T.: Threat assessment in tactical airborne environments. In: 5th International Conference on Information Fusion, Annapolis, Maryland, USA, pp. 1300–1307 (2002)
10. Lee, J., See, K.: Trust in automation: designing for appropriate reliance. Human Factors: The Journal of the Human Factors and Ergonomics Society 46, 50–80 (2004)
11. Parasuraman, R., et al.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans 30, 286–297 (2000)
12. Rovira, E., et al.: Effects of imperfect automation on decision making in a simulated command and control task. Human Factors: The Journal of the Human Factors and Ergonomics Society 49, 76–87 (2007)
13. Irandoust, H., Benaskeur, A.: Argumentation-based Decision Support in Naval Command and Control. In: The ECAI-2006 Workshop on Computational Models of Natural Argument (CMNA VI), Riva del Garda, Italy, pp. 26–30 (2006)
14. Mark, G., Kobsa, A.: The Effects of Collaboration and System Transparency. Presence 14, 60–80 (2005)
15. Preece, J., et al.: Interaction Design. Beyond Human-Computer Interaction. Wiley, New York (2002)
16. Pfautz, J., et al.: The Role of Meta-Information in C2 Decision-Support Systems. In: Command and Control Research and Technology Symposium, San Diego, California, USA (2006)

17. Dong, X., Hayes, C.C.: Uncertainty Visualizations: Helping Decision Makers Become More Aware of Uncertainty and Its Implications. Journal of Cognitive Engineering and Decision Making 6, 30–56 (2012)
18. McGuirl, J., Sarter, N.: Supporting Trust Calibration and the Effective Use of Decision Aids by Presenting Dynamic System Confidence Information. The Journal of the Human Factors and Ergonomics Society 48, 656–665 (2006)
19. Zsambok, C.E., Klein, G.A.: Naturalistic decision making. Lawrence Erlbaum (1997)
20. Jensen, F.V., Aldenryd, S.H., Jensen, K.B.: Sensitivity analysis in Bayesian networks. In: Froidevaux, C., Kohlas, J. (eds.) ECSQARU 1995. LNCS (LNAI), vol. 946, pp. 243–250. Springer, Heidelberg (1995)
21. Lacave, C., Díez, F.: A Review of Explanation Methods for Bayesian Networks. Knowledge Engineering Review 17, 107–127 (2002)
22. Fan, X., et al.: The Influence of Agent Reliability on Trust in Human-Agent Collaboration. In: 15th European Conference on Cognitive Ergonomics: The Ergonomics of Cool Interaction, Madeira, Portugal (2008)
23. Duggan, G.B., et al.: Too Much, Too Little or Just Right: Designing Data Fusion for Situation Awareness. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 48, pp. 528–532 (2004)
24. Dahlbom, A., Helldin, T.: Supporting Threat Evaluation through Visual Analytics. In: 3rd IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), San Diego, California, USA (2013)
25. Griethe, H., Schumann, H.: Visualizing uncertainty for improved decision making. In: 4th International Conference on Perspectives in Business Informatics Research, Skövde, Sweden (2005)
26. Skeels, M., et al.: Revealing uncertainty for information visualization. Information Visualization 9, 70–81 (2010)
27. Johansson, F.: Evaluating the performance of TEWA systems. Doctorial thesis, School of Science and Technology, Örebro University, Örebro, Sweden (2010)

Design of a Guided Missile Operator Assistant System for High-Tempo Intervention Support

Tobias Kloss and Axel Schulte

Universität der Bundeswehr München (UBM), Department of Aerospace Engineering
Institute of Flight Systems (LRT-13), 85577 Neubiberg, Germany
{tobias.kloss,axel.schulte}@unibw.de

Abstract. Controlling a short-range missile with in-flight reconfiguration capabilities places high demands on the design of the missile operators' control station and automated functions. To enable the missile operator to react fast, reliable and in a responsible manner to unforeseen events, e.g. high risk for collateral damage, an automated decision support system is investigated in this article. A common approach to reduce the high time demands of the operator is to transfer more functions from the human to the machine. Such emerging high levels of automation introduce ethical problems as well as new issues in human-automation-interaction to be resolved. At the Institute of Flight Systems we follow a well-established approach of human-automation cogency to assist human operators while keeping them fully involved in decision processes, i.e. "dual-mode cognitive automation, DMCA". This article presents first steps towards the application to a high-tempo mission with minimal information on the task and the tactical environment being available to the automated system.

We present an approach to relieve the human from the time critical task to enter suchlike information into the system, thereby freeing cognitive resources for mission critical decisions. At the same time the assistant system observes the actions of the missile operator, infers his/her most likely intents, and adapts support functions accordingly. Comparing the human's control actions with intention related task models, the assistant system shall identify errors and suggest alternative actions or possible solutions. The operator remains in full control of all functions and decides whether to accept or decline the assistant systems advises. This article provides an overview over the main conceptual ideas and the current status of prototype implementation.

Keywords: missile operator, intervention support, levels of automation, assistant system, intent recognition, adaptive automation.

1 Introduction

At the Institute of Flight Systems (IFS) at the University of the Bundeswehr Munich (UBM), the characteristics of a short range guided missile mission are subject of research. State of the art so-called fire-and-forget missiles are capable to engage the designated target with a high degree of reliability. Before launch, the missile is

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 273–281, 2013.
© Springer-Verlag Berlin Heidelberg 2013

programmed with cruise waypoints and the target signature. The onboard navigation and seeker head enable the missile to automatically find its way to the target without the need or possibility for human intervention after launch. However, in case of an unforeseen event occurring at the target area, for example the detection of uninvolved civilians, the fire-and-forget strategy can cause massive and totally unacceptable collateral damage. This is why today the requirement arises to enable a human operator to intervene even at very late stages of the missile flight [1].

In this article, the characteristics of controlling a missile with reconfiguration capabilities in flight are examined. Therefore, we look at a missile which is controlled by a human missile operator from a ground control station outside the battlefield. Targeting information will be provided by an infantry soldier in the field, the so-called spotter. Upon fire request the tasks of the missile operator are to plan the flight trajectory for the missile and to configure the seeker. After launch, the missile will be monitored and controlled via data link. This technically allows the operator to react to unforeseen events in the target area.

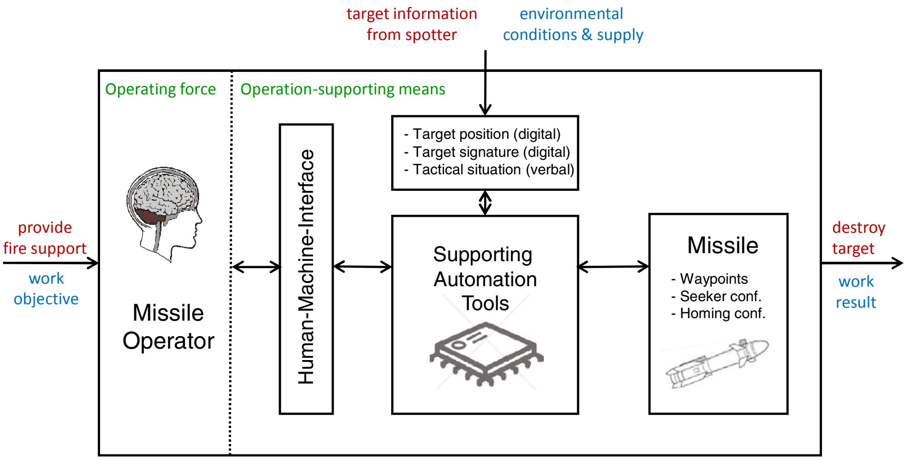


Fig. 1. Work System of Missile Operator

To better understand the connectivity between missile operator, spotter and missile, the corresponding work system is shown in Fig. 1. A work system represents the components which accommodate the functions of the underlying work process [2]. The operator constitutes the operating force. His/her work objective is to provide fire support by planning the missile mission, executing and adapting it to possible changing demands. He/she can interact with the missile through an HMI and the help of supporting automation tools. Information about the target is received from the spotter in the field. This information consists of digital data which can be processed directly by the supporting automation tools like the target position coordinates and the target signature. Information about the tactical situation is only transmitted verbally to the missile operator, though.

The flight duration of the considered short-range missile is usually not much longer than one minute. In case of an intervention request, the missile operator has only very limited time to adapt mission parameters. Besides commanding a mission abortion, the operator has the option to adapt the flight trajectory, change to an alternative target, or modify the target approach configuration (azimuth and elevation angles). This high-tempo, high-workload environment reaches the limits of human sensory and cognitive abilities, driving the need for automated support systems [3].

The following chapter will dwell upon the fundamental problems in defining the right aggregation level of interaction to be chosen for the human-machine-interface design. This question is frequently oddly associated with the term level of automation (e.g. [4]).

2 Analysis of Interaction Level

In cognitive ergonomics the design of automation functions and human-machine-interface should follow the well-adjusted allocation of functions between the human and the automation. Especially in the application we are looking at, which is heavily driven by high-tempo lethal decisions, ethical as well as cognitive performance aspects are the driving factors.

Not only from an ethical stance it is recommended to always keep the human in the decision loop. The missile operator is responsible for a successful yet justifiable mission outcome. Therefore, he/she should be aware of the situation and capable of acting at any given point in time. Only with complete knowledge about the situation, the operator can project decisions in time and decide when and how to intervene. Regarding the fact that there is only very little time remaining for this, once a missile has been launched, high levels of automation, in its extremes “full autonomy” might look like a tempting choice, although corrupting aforementioned principles. Lower levels of automation, i.e. in our case levels of abstraction in the interaction concept with automated functions may result in the necessity of a vast number of control actions of the human with the system. In its extreme the operator would be demanded to enter each single trajectory point and parameter into the system manually. The advantage of such low level of automation is that it provides full control over and direct access to all functions and sub-functions. On the other hand, given the high-tempo environment excess high mental and even physical workload would be the consequence.

Higher levels of automation and abstraction can reduce and simplify interaction with the downside of possible opacity of the background processes of the automation. The operator may easily get out of the decision loop by not fully understanding the automated processes. Not in all cases it can be assured that a highly aggregated macro-function offered by the automation (e.g. “abort mission”) will result in what the user actually intended or expected.

In case of an intervention request caused by an unforeseen event, the task of the missile operator is to adapt the mission accordingly. Therefore, the system has to provide appropriate functions covering all possible occurring use cases. Taking into

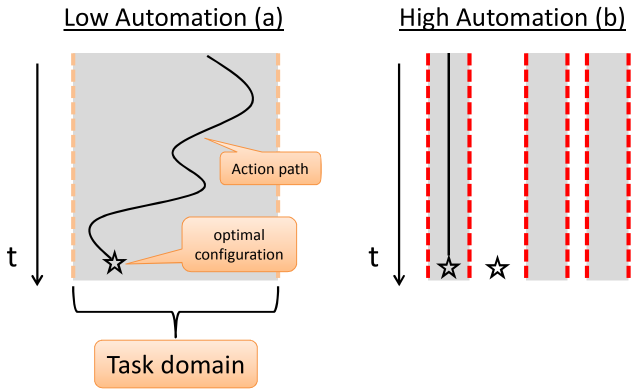


Fig. 2. Comparison of different levels of automation abstractions

consideration the previous discussion, the question arises which level of abstraction is the right one to choose? Fig. 2 illustrates a possible course of action for the missile re-tasking given different levels automation abstraction.

As depicted in Fig. 2 (a), with a low automated system, the operator can choose from a wide range of actions alternatives to achieve the desired outcome. In this setup, the operator should always be able to define an appropriate action sequence. But taking into account the high-tempo environment, too much freedom of choices can prevent finding the shortest command sequence and produce unnecessary delays, or the number of required interactions to do the job is just too high to be accomplished within the given time frame.

On the other hand, with a highly automated system Fig. 2 (b), the missile operator may trigger one of very few predefined action sequences, which have been configured during design time of the system to achieve a certain outcome. Most likely it would be the case that the very situation would demand for slightly deviating adjustments, which are not possible to be entered into the system. In this case the operators' work demands may be massively below the ones described before with the downside of cutting back his/her authority. Once initiated, the action sequence will be executed automatically, prohibiting the operator of further interventions. In case the situation demand for a solution not considered at design time, the automation fails to offer appropriate functions. Additionally, a high level of automation switches the operator's activities to monitoring automated carried out tasks. Too much confidence in the automation can lead to complacency, skill degradation and loss of situation awareness [3]. The operator cannot fulfil his responsibility for the mission by just accepting and not fully understanding the impact of the automated task processing.

A first approach to reduce the disadvantages of the configurations described above is to choose a moderate level of automation abstraction - high enough to limit the required number of interactions, low enough to permit deviation and to retain authority of the system. In the considered missile mission with changing demands over time, the optimal level of automation strongly depends on the current work objective of the operator. In this environment a static design of automation does not look suitable to provide optimal support.

In [5], an alternative approach is proposed by adapting the function allocation between human and machine to the current needs. Therefore, "Adaptive Automation" needs to know the work objective before being able to adjust the supplied supporting tools. In our application this could be realized by implementing a dialog system to enter the missing information. The resulting additional interaction step for the operator would lead to an increased workload, though, which is not desirable during an already stressful situation. Additionally, by integrating adapting function allocation, the operator is confronted with suddenly changing operation procedures.

In this article, an alternative approach is proposed. A low level of automation abstraction is chosen to allow the operator to retain full control. In addition to this conventional automation, an assistant system is introduced to adaptively support the operator during high demanding situations. The current work objective of the operator is recognized by observing the actions of the operator, using an intent recognition algorithm. Thus, the described additional dialog system can be avoided. The probability of inferring the correct work objective is increasing with each interaction step executed by the operator. After reaching a minimum threshold probability, the assistant system can support the operator achieving his/her objective by comparing and evaluating the upcoming action path to an appropriate underlying human action model.

The design of a missile operator assistant system is described in the following chapter.

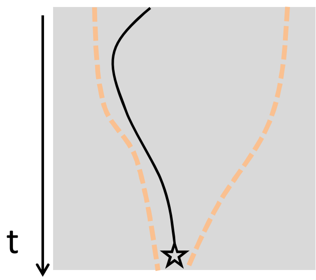


Fig. 3. Assistant System – principle of operation

Fig. 3 summarizes the principle of operation of the missile assistant system. The action alternatives offered by the system remain unrestricted in the first place, i.e. the operator may carry out any desired low-level task at any time, allowing him/her to keep full control of the situation. With each step of the observed user interactions, the actual task can be narrowed down, depicted by the yellow dashed lines, allowing to guide the operator in finding the shortest command sequence to reach the designated work objective.

3 Missile Operator Assistant System

Fig. 4 depicts how the assistant system is embedded into the missile operator work system. To illustrate the difference to conventional automation being part of the operation-supporting means, the assistant system is placed next to the missile operator as part of the operating force.

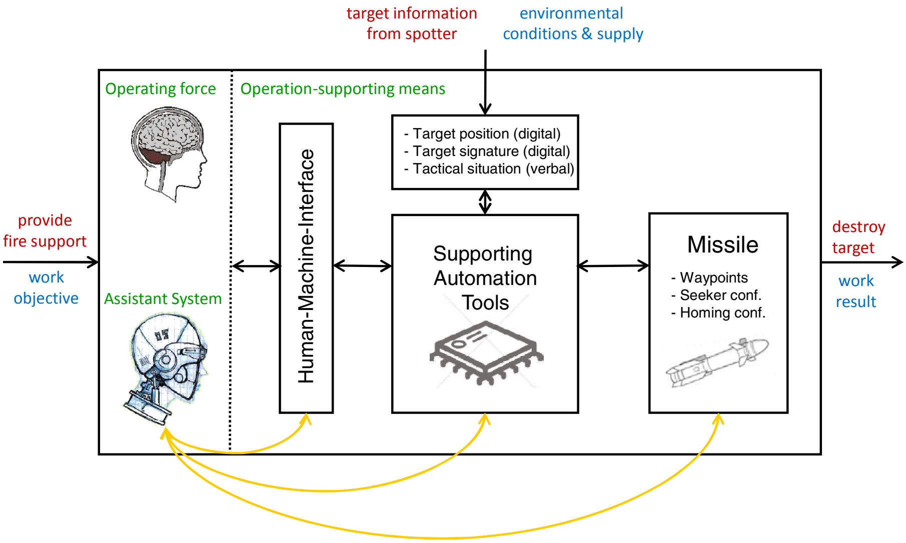


Fig. 4. Work System of Missile Operator and Assistant System

The assistant system's job can be compared to that of a co-pilot in an airplane: The co-pilot knows the current work objective, is aware of the environment and can exchange information with the pilot in command. Situational task demands, for example a request to change the landing runway, can force the pilot to hand over autopilot functions. Now, the co-pilot can assist by monitoring the pilot's actions and informing about faulty operation, making alternative proposals or even taking over control of the aircraft. Since a machine overtaking control from the human in a missile mission is undesirable due to ethical aspects, the proposed intervention level of the missile assistant system will be restricted to associative and alerting functions.

The supporting automation tools as part of the operation supporting means still exist to support carrying out closed tasks like automated route planning and flight duration calculations. Illustrated by the yellow arrows in Fig. 4, the assistant system can interact with the missile operator through the HMI. It also has access to the supporting automation tools to develop and present alternative action recommendations to the missile operator.

As described, the assistant system is based on recognizing the actual task by observing the operator's actions in a non-intrusive way. To be able to offer punctual and feasible support, a reliable estimation of the operator's objective has to be found as fast as possible. As described in [6], activity recognition in human computer interaction can be based on three main aspects: the *interaction set*, the *task model* and the *inference technique*.

The *interaction set* consists of the observed user interactions by the operator like keystrokes, mouse clicks and software events. Interactions can be embedded into the current context to increase the value of information. For example, the new position of a waypoint moved by the operator can be compared to the previous position to determine the most likely intention of the modification to the flight path. Moving and zooming the view can be interpreted as preparing changes either of the target configuration or the cruise flight path depending on the new visible view section.

The *task model* represents the sequential structure of a task. It can be generated by using task analysis tools (for example Hierarchical Task Analysis [7]) or by analysing the observed interaction sequences of test subjects during experiments. The resulting model should represent the optimal way to achieve the corresponding task, acting as a base for the supporting functions of the assistant system.

The *inference technique* links an observed interaction sequence with the task model [6]. A certain task can be achieved in various ways, using different available action alternatives or by altering the order of actions. The resulting, varying action sequences differ from the task model and can reduce the reliability of the intent recognition algorithm.

Commonly used predictive statistical models are presented in [8]. Hidden Markov Models have a rich history in sequence data modelling (for example in speech recognition systems [9]) and a promising approach for the missile operator assistant system.

At the time of identifying a matching task model, the assistant system can start to support the operator to achieve his/her objective. The implemented supporting functions include presenting single hints and alerts up to suggesting alternative action plans to the operator. The assistant system does not automatically execute or override any actions, because the operator should always be in charge of his own decisions. The resulting implemented *level of automation of decision and action selection* as defined by [4] ranges from level 1 ("The computer offers no assistance") in case of no deviations being detected between task model and human actions up to level 5 ("Executes […] suggestions if the human approves").

Additionally, based on the remaining uncertainty in identifying the correct task model, a higher level of automation could result in faulty decision support, corrupting the work of the operator. By restraining the maximum intervention level to 5, the operator can choose to ignore the proposed actions.

Possible supporting functions of the assistant system in the defined range can be

- informing about possible negative implications of the current task,
- highlighting the most imminent task,
- proposing pending tasks to reach the desired configuration with the option of automatic execution, and
- suggesting adjustments of already executed tasks to increase mission performance.

4 Experimental Method

At the UBM, a simulation environment for the introduced guided missile mission has been designed and implemented. The simulation allows human-in-the-loop experiments with test subjects, acting as missile operator. The operator can be confronted with different scenarios to force stressful, high workload situations as described above. After an incoming request for fire support, the planning phase begins. Now, the operator has to set up the missile configuration by planning the flight path, defining the target and configuring the final approach. The chosen configuration has to meet the following constraints:

- Limitation of flight duration and missile agility,
- time-over-target requirement,
- avoidance of terrain, structure and restricted areas, and
- ensuring the visibility of the target to the seeker during final approach.

Automation tools are implemented to support the missile operator by offering route planning, flight duration calculation and presenting information about violated constraints. After missile launch, the operator has to monitor the mission progress and wait for an incoming intervention request from the spotter. In the considered missile mission, possible requests can be

- mission abort,
- adjusting the time-over-target,
- modifying the final approach configuration (final leg length and approaching direction),
- switching to an alternative target, or
- updating the position of a moving target.

Each of the listed objectives requires the operator to execute a sequence of actions. As an example, to modify the approaching configuration, the operator has to move and zoom his view to be able to reach the corresponding waypoint. Additionally it may be necessary to adjust other cruise waypoints to avoid unfeasible or ineffective flight manoeuvres or to adjust the flight duration. Observable actions of the operator can be

- panning and zooming the map view,
- selecting an interaction mode,
- inserting, moving and deleting waypoints and
- setting and updating the target position.

The proposed missile operator assistant system based on intent recognition will be integrated into the simulation. By confronting the test subjects with various demanding scenarios, the resulting mission performance is analyzed while offering different types of supporting automation tools. It will be investigated if the assistant system can enhance the operator's performance in comparison to conventional automation.

5 Conclusion and Future Work

This paper introduces the design of a missile operator assistant system adapted to the special characteristics of controlling a short range missile with reconfiguring capabilities in flight. Pros and cons of integrating low and high levels of conventional automation support are described. The proposed assistant system acts on top of an environment of low level of automation allowing full control to the operator at any point in time. It is designed to adaptively support the operator in critical situations without requiring additional interaction steps. To offer reasonable support, the assistant system obtains knowledge about the current work objective by observing the operator's actions, inferring his/her most likely intention.

To examine the effectiveness of the proposed assistant system, a missile simulation environment is being developed at the UBM. A test subject acting as missile operator can be confronted with stressful, error-prone situations. Different types of supporting tools will be integrated to confirm the described negative effects of conventional automation. An assistant system based on non-intrusive intent recognition functionality will be implemented and evaluated.

References

1. Erwin, S.I.: Air Force wants missiles redirected in flight: weapons with two-way communications links could help avoid fratricide. National Defense (Magazine/Journal) 87, 28(2) (2003)
2. Onken, R., Schulte, A.: System-Ergonomic Design of Cognitive Automation. SCI, vol. 235. Springer, Heidelberg (2010)
3. Johnson, K., Ren, L., Kuchar, J., Oman, C.: Interaction of automation and time pressure in a route replanning task. In: International Conference on Human-Computer Interaction in Aeronautics (2002)
4. Parasuraman, R., Sheridan, T.B., Wickens, C.D.: A model for types and levels of human interaction with automation. IEEE Transactions on Systems, Man, and Cybernetics. Part A: Systems and Humans 30, 286–297 (2000)
5. Inagaki, T.: Adaptive automation: Sharing and trading of control. In: Handbook of Cognitive Task Design, pp. 147–169 (2003)
6. Courtemanche, F., Aïmeur, E., Dufresne, A., Najjar, M., Mpondo, F.: Activity recognition using eye-gaze movements and traditional interactions. Interacting with Computers 23, 202–213 (2011)
7. Crystal, A., Ellington, B.: Task analysis and human-computer interaction: approaches, techniques, and levels of analysis. In: Proceedings of the Tenth Americas Conference on Information Systems (2004)
8. Nazemi, K., Stab, C., Fellner, D.W.: Interaction analysis: An algorithm for interaction prediction and activity recognition in adaptive systems. In: IEEE International Conference on Intelligent Computing and Intelligent Systems, ICIS (2010)
9. Rabiner, L.: A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)

Enabling Dynamic Delegation Interactions with Multiple Unmanned Vehicles; Flexibility from Top to Bottom

Christopher A. Miller[1], Mark Draper[2], Joshua D. Hamell[1], Gloria Calhoun[2], Timothy Barry[3], and Heath Ruff[3]

[1] Smart Information Flow Technologies (SIFT), 211 First St. N. #300, Minneapolis, MN USA 55401
[2] Supervisory Control and Cognition Branch, 711th Human Performance Wing, Air Force Research Laboratory, 2255 H Street, Wright-Patterson Air Force Base, Ohio 45433-7022, USA
[3] Ball Aerospace and Technologies Corporation, Fairborn, OH 45324, USA
{cmiller,jhamell}@sift.net, {mark.draper,gloria.calhoun,tim.barry.ctr, heath.ruff.ctr}@wpafb.af.mil

Abstract. A "delegation approach" to human interaction with automation should strive to achieve all of the flexibility that a human supervisor has in instructing, managing, redirecting and overriding well-trained human subordinates. But in the absence of human-like, natural-language understanding "androids", what would such interaction look like? This multi-year design and evaluation project explores such interaction concepts for pilot control of multiple remotely piloted systems. This paper details the underlying philosophy of delegation and presents many design innovations developed to date.

Keywords: flexible automation, adaptive/adaptable automation, playbook, delegation, UAVs, UASs, remotely piloted aircraft, multi-modal interaction.

1 Introduction

Humans do not "control" other humans in the sense that we control aircraft, automobiles, or power plants. Instead, we "delegate" actions, tasks, goals or responsibilities to "subordinates", and we "supervise" their performance. We have been advocating a "delegation approach" to human-automation interaction [1] that strives for all the flexibility that a human supervisor has in instructing, managing, and overriding well-trained human subordinates. But machines are not human, and at least most proposed automation for interaction with pilots does not have a human-like appearance, much less share natural language understanding or apply "common sense". So, what should delegation interactions look like with pilot-centric automation and how should it behave? A multi-year design project is exploring this question for pilot control of multiple semi-autonomous systems with the U.S. Air Force Research Laboratory (AFRL).

The application domain is control of multiple next-generation Air Force Remotely Piloted Aircraft (RPAs) in likely missions. This poses a challenging range of

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 282–291, 2013.
© Springer-Verlag Berlin Heidelberg 2013

interaction needs from precision or emergency reaction manual flight control to aggregate, multiple vehicle participation in precise integrated activities such as coordinated attack. Future needs require that a single operator be able to manage multiple, concurrent vehicle missions with no RPAs behaving in an unproductive or problematic fashion. The human pilots must remain in full charge of all capabilities of the vehicles they manage, even when they are focused on, perhaps even manually flying, one of them. In practice, this has meant that the operator must be able to intervene and demand any activity (including manually-controlled flight) at any time without impinging on the vehicle automation's ability to take control back when it is delegated. Pilots must be able to adapt automation's preplanned missions and behaviors to the uncertainties that will inevitably exist in military operations, or override them *effectively* to ensure that human intent is accomplished even in the "fog of war."

The overall AFRL program, called FLEX-IT (for Flexible Levels of Execution-Interface Technologies) has developed a series of designs and demonstrations to illustrate these goals. Below, we first describe the work domain, then the FLEX-IT control and delegation methods along with the delegation rationale for their selection and implementation. Next we present some specific design innovations, and summarize pilot feedback received to date. Finally, we discuss directions for future development.

2 The Multiple RPA Control Domain

Many current RPAs require at least 2-3 individuals during operations. A widespread goal is to at least reverse that ratio—enabling a single pilot to productively operate 3 or more RPAs [2]. Such missions would certainly include intelligence, surveillance and reconnaissance and might involve force protection, resupply, and/or attack and combat roles. Missions might be independent (searching for different, unrelated targets), or coordinated and loosely coupled (e.g., searching portions of a broad area) or cooperative and tightly coupled (e.g., coordinated attack). Present and likely future Air Force missions will require the ability to for pilots to exert precise control over flight paths, sensor operations, etc. Automation plays a key role in achieving the desired operator-to-vehicle ratio, but precise, expert human input will still be required.

This necessitates highly flexible human-automation interaction which spans, supports and integrates levels and styles of control ranging from full manual flight to rapid and comprehensive full-mission automation. It also mandates a wide variety of command input methods from the familiar throttle and stick, through keyboard and mouse to potentially speech, touch and sketch inputs—all to facilitate easy, natural yet efficient and flexible RPA control. System information output is similarly diverse, demanding use of the traditional map- and sensor-view screens along novel modalities and configurations to provide rapid but comprehensive awareness.

Our overall goal is to create designs that enable the flexible delegation a human supervisor has with intelligent subordinates in a well-understood operational domain. Thus, our designs take a "delegation approach" to multi-RPA control—preserving the functionality, if not always the modality, of instructions that a pilot commanding a multiple aircraft mission would employ if interacting with human subordinates.

3 FLEX-IT Control and Interaction Methods

In human organizations, supervisors can communicate with trained subordinates with a presumption of shared understanding of the common plans, goals, and methods in the work domain, of the common, culturally-determined priorities and limitations on appropriate behavior, etc. They can talk about activities in detailed natural language, using all modalities for nuanced intent, urgency, uncertainty, etc. that humans share. Moreover, they will likely share specialized vocabulary—a jargon or prescribed protocol to reduce errors and improve speed. These factors enable delegation with great flexibility. Supervisors can activate complex, multi-agent behaviors with a word, but they can also shape and refine those behaviors, create new ones, can task at various time points before and during behavior execution, and can even jump in and physically perform or illustrate behaviors for subordinates if desired. This is the flexibility we are striving to achieve in human-automation interaction.

3.1 Delegation within a Hierarchical Task Decomposition

We have argued [1, 3] that delegation can be viewed as taking place within a "space" of possible domain tasks and this space can be characterized using a hierarchical decomposition [as in 4] where alternate performance methods are represented as "or" branches, and sequential dependencies, looping, conditional branching, etc. are captured as a directed graph. Automation changes the nature of tasks to be performed [5,6], but these alternatives can be regarded as alternate branches. The selection of a branch to take is, in essence, a function allocation decision—which is to say, a delegation decision. It can be made either at design time or left to execution time. If left for execution, this enables adaptive function allocation, but leaves open the question of who gets to make the allocation decision. If automation makes it, then the system can be called adaptive; if the decision is left to a human supervisor, then the system is adaptable [7]. Adaptable systems, especially those that offer flexibility in how delegation actions are made and communicated, have demonstrated advantages in terms of workload, predictability and user acceptance [1,8].

Figure 1 shows a decomposition for a single task (monitoring a target). Resource allocation alternatives have been suppressed (assuming a single RPA equipped with some kind of sensor) for ease of presentation; including them would entail many alternate branches. The parent task (Monitor Target) is decomposed into sequentially ordered subtasks (Ingress, Approach, Inspect, etc.). Most of these can be skipped (as depicted by the dotted forward arrows), but the Inspect subtask is required for a valid instance of Monitor Target—though one can iterate through Inspect multiple times (note the dotted backward arrow). Each subtask is further decomposed into one or more "Fly Maneuver" subsubtasks and "Inspect" also requires a concurrent "Maintain Sensor Focus" subsubtask. A range of possible maneuvers is shown as the next de composition of the first "Fly Maneuver"--the branching lines convey that any maneuver is potentially viable. Each maneuver is decomposed into specific control actions.

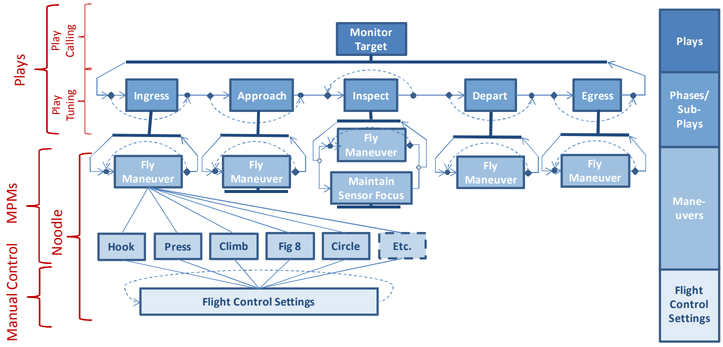


Fig. 1. Hierarchical decomposition of a single task along with illustrations of how FLEX-IT's control modes cover the range of delegation interactions within this hierarchy

A supervisor interacting with human subordinates could choose to do this entire task alone or could delegate it to a subordinate to plan and execute simply by invoking the name and providing a target. S/he could also choose to delegate parts (e.g., the Ingress portion) but retain control of the rest, and could constrain or stipulate how tasks are to be performed (e.g., specific routes to take or avoid, etc.). S/he could offer instructions either holistically (as a full mission plan) or piecemeal, and could jump in to revise or override as needed. S/he could also assemble lower-level tasks to achieve a new and different goal—something that would not be a Monitor Target mission, but would nevertheless use some of the same behaviors. Delegation need not occur at any fixed level within the hierarchy—an entire mission might be assigned to one subordinate, while only a maneuver flying task to another. Decisions are revisable: after initially delegating full mission execution, the supervisor might later decide to constrain a subtask—or even to fly portions personally. It is precisely this level of flexibility and ease of control input that we are striving to provide in FLEX-IT.

3.2 FLEX-IT Control Interaction Methods and Modalities

Figure 2 illustrates a current FLEX-IT workstation. The different control approaches described above are implemented and available in a variety of supporting input and output modalities including speech, touch, Hands On Throttle And Stick (HOTAS), and keyboard and mouse controls. This also parallels the flexibility of human-human delegation since humans communicate in multiple modalities—usually concurrently, but sometimes preferring one or more due to contextual factors.

FLEX-IT's control modalities and pilot interactions are described in more detail in [9] but will be summarized here. The left screen shows a map with RPAs, targets, routes, etc. with the bottom pane showing auxiliary details about plays and vehicles. The right screen is devoted to sensor feeds from each RPA. Besides HOTAS (manual control) and keyboard/mouse (most other inputs), FLEX-IT has also used a touchscreen to duplicate or extend some input capabilities. Headphones and a boom microphone enable auditory output and speech input.

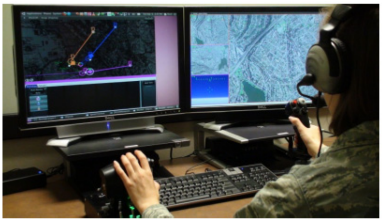


Fig. 2. The FLEX-IT workstation

Fig. 3. A chain of noodles representing a complex 4D flight path achievable by this RPA

FLEX-IT's Manual Control is available through vehicle selection and HOTAS activation that controls flight and also sensor movements (as well as "noodle" inputs as described below). This requires a sophisticated HOTAS controller with multiple auxiliary buttons. HOTAS control has evolved to be the most convenient and natural method for controlling the immediate flight of a vehicle. As a design principle, pilots should be able to manually control any vehicle at any time regardless of other previously delegated activities. This is akin to a supervisor saying "Stop what you're doing and do this instead."

The "Noodle" (Figure 3) is a unique control approach in FLEX-IT. It provides a convenient and natural way for pilots to tactically "pre-fly" the airplane from its current position. Using the current flight profile (position, altitude, airspeed, heading, etc.) it takes inputs of hypothetical flight control movements, via HOTAS, and then shows the results of sustaining those control actions for a variable time. This "noodle" (named for its appearance on the map) is pushed in front of the RPA icon and constantly updated until accepted for execution. Then, the RPA flies the noodle to its end. Subsequent flight control actions can be appended to initial ones to "chain noodles" as much as 15 minutes into the future.

Fig. 4. A completed Monitor Target plan, ready for pilot review, editing and/or acceptance

Noodle control differs from a waypoint-based, autopilot-flown route in two important ways. First, it is generally much faster and more natural to command since it uses the same modality used for flight—the HOTAS—rather than the typical keypad and cursor for inputs. Second, most existing autopilots treat waypoints as "goal points" and do not necessarily show the route used to achieve them. Especially for points very near a current position or placed very close together, the aircraft may not fly a straight line to achieve the waypoints. This is potentially dangerous if the actual route enters a restricted zone or crosses a national border. Since the noodle projects flight control inputs, it shows the vehicle's precise flight path barring disturbances.

Noodle control is similar to waypoint control in one important aspect: it allows delegating a flight path ahead of time. This can save the pilot time and/or enable offsetting when precise directions are given. If a pilot needs to steer one RPA through restricted zones and concurrently use HOTAS sensor controls to lock onto a ground target from another RPA, the noodle can dictate the flight path, allowing attention to be directed to the sensor task while the first RPA flies it.

Micro-Play Maneuvers (MPMs) are limited duration, fine-grained tasks. MPMs are primarily intended for quick verbal interactions, though we also provide methods for them to be "called" by mouse. Similar flight behaviors could be achieved by a small set of flight control inputs. Examples include: climb or descend to an altitude or at a specified rate, change to a specified heading or at a specified bank rate, circle and figure eight loiter patterns, hook (left or right), and press (that is, fly at) or separate from a specified target. The voice inputs to activate an MPM are defined from Air Force usage and radio practices—so while the pilot must remember specific verbal strings, these are natural, expected ways of discussing these behaviors. Feedback is provided on the map to show what the system has heard and will execute.

MPMs are suited for quick commands or corrections. Their verbal modality is ideal for use when controlling another vehicle—akin to a supervisor engaged in other work who can nevertheless tell a subordinate "Do this!" to maintain productivity over short periods of reduced attention. There are relatively minor consequences to invoking an incorrect MPM as long as it is noticed and overridden quickly—especially easy via the verbal input modality. Thus, MPMs are initiated immediately upon speech recognition with no further authorization from the pilot. They also persist for a limited time before either completing and entering into a default holding pattern or persist indefinitely with occasional queries as to whether this behavior should continue.

Plays are more complex means of commanding RPA(s) over longer time spans. FLEX-IT makes use of SIFT's Playbook® approach to adaptable automation, described elsewhere [1]. Plays are templates representing an agreed-upon goal and range of acceptable methods of achieving it. For example, in our Monitor Target play (Figure 1), the goal is achieving sensor images of a stationary target and acceptable methods involve getting a sensor-equipped RPA into position to provide that imagery and then depart. Other methods (such as satellite imagery, or even imagery provided by multiple vehicles) are not acceptable exemplars of *this* play. This is strictly a definitional convention, but it is precisely the kind of conventions human organizations evolve to facilitate rapid communication and shared understanding.

Play calling is flexible as well. Plays can be called (verbally or by touch/mouse) by invoking the play name along with required parameters—e.g., saying "Monitor Target Alpha" (a pre-designated target), or "Monitor Target here" with a pointing gesture. This represents maximal delegation since detailed planning (what route, inspect formation, etc.) is left to automation using pre-stipulated defaults and preferences. The user can command more specific methods, though, by issuing verbal commands either at initial play call (e.g., "Monitor Target Alpha, altitude, 14000" to stipulate an altitude of 14000' throughout the play), or subsequently (e.g., "Change Inspect altitude, 15000" to revise only the Inspect phase).

Since plays generally involve more substantial planning that may be left to automation, the plan for a given play is reviewed by the pilot prior to acceptance and execution. A tentative plan is shown with a dashed line route plan in Figure 4. Sub-play phases are indicated by chevrons denoting their start points ("I" for Ingress, etc.) Sub-plays can be independently edited, as can the play as a whole, either verbally or via touch or mouse-controlled menus or by dragging and dropping waypoints. Once the play is accepted, FLEX-IT executes it with ongoing automated monitoring and revisions within the constraints of the play. The pilot can continue to revise it by manipulating the route or other details manually or through subsequent verbal "change" commands.

4 Design Challenges and Innovations

The governing heuristics we have used in our design may be stated as follows:

1. The pilot should interact with the FLEX-IT automation, as much as possible analogous to how s/he would interact with an intelligent subordinate.
2. Everything must be doable via multiple modalities
3. Pilot must be able to interact via any of the control approaches at any time
4. Pilot can seize control at any point with automation "backing off"
5. Detailed, longer duration plans must be preserved and intelligently resumable
6. Pilot must approve detailed and/or risky automation behaviors before enacting
7. Pilot must not be forced to approve predictable, limited, recoverable and/or non-risky behaviors before enacting
8. Predictability should generally be preferred over presumed workload saving

Achieving flexible interactions has led to several innovations. Our ultimate goal in FLEX-IT is both "top down" and "bottom up" play construction. We have largely achieved top down play calling and refinement—using a simple label or action to call a complex, high level play and having automation interactively develop a plan in context to achieve it. We have not yet fully achieved "bottom up" construction — using any and all of the control approaches, at any time and via multiple convenient modalities, to construct and revise innovative sequences actions as desired. We are working toward it, however, with innovations such as:

Play Suspension, Resumption and Staleness: Plays generally represent longer term intentions and will take longer to develop and refine. Allowing the pilot to seize the controls at any point could negate the effort spent in laying out an initial plan for automation. By contrast, a human supervisor could indicate that an interruption was temporary and or, alternatively, that a whole new plan was needed. We provide this flexibility by not merely enabling pilots to interrupt a play to take manual control, but also to "pause" that play and make it available for later resumption.

When the pilot grabs the HOTAS or calls an MPM disrupting an ongoing play, the vehicle responds, but four changes appear on the display (as shown collectively in Figure 5e): (1) the leading "whisker" off the nose of the RPA icon shows a symbol indicating the vehicle's type of control. During play (or waypoint following) control, the whisker has a white circle (see Figure 5a), but under manual control, this icon is removed, (2) a pause icon is shown on the RPA symbol (see Figure 5b) indicating a paused play is available, (3) a moving white diamond appears on the route indicating where the RPA will resume unless the pilot dictates otherwise (Figure 5c), and (4) a "staleness" meter conveys information about how "out of date" the paused play is becoming (Figure 5d). The pilot can issue a verbal or menu-based mouse command to "Resume" at any point. The resumption point (cf. Figure 5c) is generally not where the vehicle left the route, but rather a point on the route as near as possible to the vehicle's current location without passing a subplay phase boundary. This provides a conservative, yet more intelligent resumption behavior. Mouse or touch gestures can also be used to say "Resume here" at any desired point along the planned path.

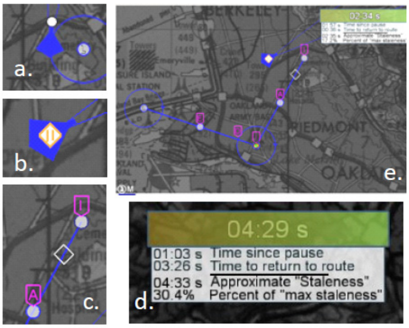


Fig. 5. Symbology for a paused play for an RPA off plan

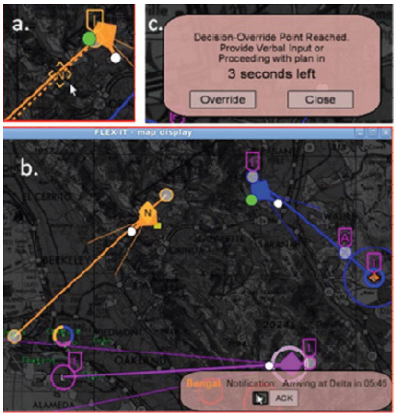


Fig. 6. Notification and Decision point behaviors showing (a) insertion of a notification point, (b) activation of that notification point and the resulting message, and (c) a similar message for a decision point

Our current staleness indicator simply sums the time since pause with that required to return to the computed resumption point. The expanded meter (shown in Figure 5d) shows these details, but this would normally be collapsed to show just the upper portion: a count-up timer that gradually shades from green to brown over 15 minutes. The pilot's true need is for knowledge of whether the vehicle's time away from plan threatens its ability to accomplish goals, requiring extraordinarily complex reasoning. This representative meter is a stand in for a more complex decision aid, but near term re-designs may go a step further to provide, for time-based plays, an indication as to how much, if any "slack time" exists before the initial play can no longer be resumed and still meet initial deadlines.

Notification and Decision Points: Another form of flexibility that human teams have is the ability to issue contingent instructions and/or instructions to report and/or ask for guidance. We have provided these in FLEX-IT by use of Decision and Notification Points. These are simple points (currently tied to route locations, but also linkable to times or events) that, when reached, trigger an interface action such as a popup notification or window providing a selectable choice. Both decision and notification points come "pre-packaged" with some plays (e.g., a "Transit" play moves a vehicle from its current location to an endpoint and notifies the pilot when almost there) but pilot would have the ability to lay these down at any point. Figure 6a illustrates a notification point being placed along a route, while Figure 6b shows that point, having been reached, triggering a message ("voice notes" in the form of audio recorded messages can also be provided). Figure 6c shows a decision point message provided upon reaching a number of orbits in the Inspection phase of Monitor Target. This point took the form of an "Override or Proceed" decision; if the pilot does nothing, the system will proceed with the plan. A complex taxonomy of decision points and resulting actions is viable, but we have only illustrated the concept to date.

These are only a few examples of innovations that have been developed during FLEX-IT within the overall objective of attempting to replicate human delegation flexibility.

5 Feedback and Future Directions

FLEX-IT has been reviewed with 19 active duty RPA pilots to date. Extensive results are reported in [10] but will be summarized here. Ratings of the overall concept have generally been very high. In our most formal review to date, six RPA pilots saw the demonstration described above, and all rated the overall flexible interaction design as either a 4 or 5 (where 5 corresponded to "Great Aid" and 1 to "No Aid"; mean rating=4.7). All six pilots gave the high-level play calling implementation a rating of 5, and all raters said the system would be a "Great Aid" to reducing workload and increasing situation awareness in future, multi-RPA operations. All other design concepts (noodle, MPMs, tailoring plays, transition between control methods, decision points, etc.) had mean ratings ranging from 4.2 to 4.7 with no rating lower than 3. Specific feedback from this and other exercises has helped us tune many aspects of the system including the format of speech commands, the review and approval process for MPMs and plays and the addition of notification points.

Work on FLEX-IT is ongoing. We are currently transitioning from our initial implementation in a low-fidelity simulation to a higher fidelity control environment used for simulation, user testing, and for control of live RPAs (AFRL's Vigilant Spirit Control Station). Core current design challenges include:

- Maintaining user awareness and expectations about the vehicle designated to receive commands in the various modalities available (a non-trivial issue in a multi-actor, multi-modal control system striving for human-level naturalness),

- Migrating to a play- or activity-centered awareness of the current situation rather than (or in addition to) a vehicle-centered awareness,
- Maintaining robust speech interaction vocabulary as our suite of possible verbal interactions is beginning to increase exponentially, and
- Achieving completely flexible assembly of interaction modes at the user's initiation—including intermixing of control actions at all levels and time frames to construct unique activities according to the pilot's needs.

FLEX-IT's powerful yet integrated and natural suite of interaction modes have clearly indicated that there is promise to using human-like flexible and adaptable delegation as a guide to interactions with complex automation. Flexible human-centered delegation will continue to serve as a guide as we proceed.

Acknowledgments. This research was supported by AF/A2Q ISR Innovations Division and managed by AFRL/RHCI at Wright-Patterson Air Force Base. Support from Ball Aerospace & Technologies Corporation and Smart Information Flow Technologies personnel was provided under Air Force Contract FA8650-08-D-6801.

References

1. Miller, C.A., Parasuraman, R.: Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. Human Factors 40(1), 57–75 (2007)
2. Eggers, J.W., Draper, M.H.: Multi-UAV Control for Tactical Reconnaissance and Close Air Support Missions: Operator Perspectives and Design Challenges, In: Proc. NATO RTO Human Factors and Medicine Symp. HFM-135. NATO TRO, Neuilly-sur-Siene, CEDEX, Biarritz, France (2006)
3. Miller, C., Parasuraman, R.: Beyond levels of automation: An architecture for more flexible human-automation collaboration. In: Proc. 47th Human Factors and Ergonomics Society, pp. 182–186. HFES, Santa Monica (2003)
4. Diaper, D., Stanton, N. (eds.): The handbook of task analysis for human-computer interaction. Erlbaum, Mahweh (2004)
5. Billings, C.: Aviation automation. Erlbaum, Mahweh (1997)
6. Woods, D.: Decomposing automation: Apparent simplicity, real complexity. In: Parasuraman, R., Mouloua, M. (eds.) Automation and Human Performance: Theory and Applications, pp. 1–16. Erlbaum, Mahweh (1996)
7. Opperman, R.: Adaptive User Support. Erlbaum, Hillsdale (1994)
8. Layton, C., Smith, P., McCoy, C.: Design of a Cooperative Problem-solving System for Enroute Flight Planning. Human Factors 36(1), 94–119 (1994)
9. Miller, C., Hamell, J., Barry, T., Ruff, H., Draper, M., Calhoun, G.: Adaptable operator-automation interface for future unmanned aerial systems control: Development of a highly flexible delegation concept demonstration. In: Proc. Infotech@Aerospace Conf., AIAA, Reston, VA (2012)
10. Calhoun, G.L., Draper, M., Ruff, H., Miller, C., Hamell, J.: Future unmanned aerial systems control: Feedback on highly flexible operator-automation delegation interface concept. In: Proc. Infotech@Aerospace Conf., AIAA, Reston, VA (2012)

"Person to Purpose" Manpower Architecture Applied to a Highly Autonomous UAS Cloud

Jon Platts[1], Scott Findlay[2], Andrew Berry[2], and Helen Keirl[3]

[1] Muretex Ltd, Suite 1, Robinswood, Middle Street, Lincoln, LN1 2RB, UK
jtplatts@muretex.com
[2] QinetiQ, Cody Technology Park, Ively Road, Farnborough, Hampshire, GU14 0LX, UK
{sjfindlay1,ajberry}@QinetiQ.com
[3] Defence Science and Technology Laboratory, Headquarters, Porton Down, Salisbury, Wiltshire, SP4 0JQ, UK
hjkeirl@dstl.gov.uk

Abstract. Studies indicate that “cloud" based concepts will provide benefits by maximising the availability of capability, reducing redundancy and permitting efficiencies in operation and deployment of effect. To deploy the cloud will require many problems to be solved. This paper examines automation applied to the cloud and builds on substantial work looking at command abstraction of users and consumers interacting with systems. The work retains the absolute authority of the human supervisor. Data is presented of a recent trial which immersed serving military personnel, exercising both manned and unmanned systems within a synthetic environment, whilst divorcing operators from platform ownership and concentrating instead on task ownership (thus linking person to purpose). Baseline systems were compared with systems possessing higher degrees of automation and tool functionality. The results are discussed and the key conclusions show clear benefits to operating in the person to purpose manner.

Keywords: Automation, Autonomy, Command and Control, Decision Making and Decision Support, Display design, Human Factors/System Integration, Mental workload, Requirements analysis, HCI.

1 Introduction

Current trends in defence spending in the UK, coupled with the desire for agile, flexible and potent responses to emerging threats, generate the need to do more with less and to do everything better.

Studies in the UK [1] had led to the conclusion that “cloud" based concepts will provide benefits by maximising the availability of capability, reducing redundancy in systems and permitting efficiencies in operation and deployment of effect.

To develop and deploy, the cloud will require many problems to be solved. This paper focuses on the application of automation to the cloud and builds on substantial work looking at raising the level of abstraction with which users and consumers

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 292–301, 2013.
© Springer-Verlag Berlin Heidelberg 2013

interact within systems. Moreover, it addresses the need for retaining the absolute authority of the human supervisor whilst endowing the system with sufficient intelligence with which to execute its mission.

The key area addressed by this paper is that of the role of the operator. Control of Unmanned Air System (UAS - the collection of UAVs and their control stations) is traditionally devolved to anUnmanned Air Vehicle (UAV) “pilot". This term owes much to the current view of linking a human authority to a single platform. In virtually all cases of the operation of air vehicles, there is ultimately a single human authority or “Captain” responsible for all aspects of the operation and airmanship of the air vehicle. This thinking has largely been mapped from the operation of manned platforms to those of unmanned. Furthermore, in the deployment of air platforms against tactical aims, thinking still revolves around the platform as the smallest atomic unit of capability. Within the work reported here, a term used for this is linking “person to platform”. It is argued that a key enabler for cloud based operations is to replace person to platform with “person to purpose” with the latter abbreviated to P2P. Following the adoption of P2P [2], users and consumers would become task focused and demand services from the cloud as required, remaining agnostic to platform and thus contribute to realising the aims of doing more with less and better.

The QinetiQ team have been engaged with developing automation technologies for unmanned systems since the mid 1980's and using the approach described here since 1998, [3, 4]. A notable achievement was the UAS Surrogate work [8] undertaken for the UK MoD. Recently the team have focussed on automation technologies for Intelligence Surveillance Targeting and Reconnaissance (ISTAR) roles, involving a formation of heterogeneous UAVs that constitute the Autonomous UAS. This paper briefly addresses the initial results from the third in a series of 4 trials (known as *Trial Caucasus*) which took place in November 2012. Previous trials addressed both technology development and rehearsal of the approach to analysis adopted in *Trial Caucasus*.

2 Objective

The objective of the work reported here is to show that by linking people to purpose rather than platform more can be done better with less.

Thiscontinuing work is significant because of the deep insight offered into a maturing concept. The approach has developed fully functional consoles for roles such as image analysts, airborne formation leader, Joint Tactical Air Controller (JTAC), and tactical commander,thus allowing these roles to be played out within realistic missions,operating in richly populated Synthetic Environments (SE). Experienced military *Players* are exercised through a detailed and credible storyline and their interactions with the technology and each other analysed by a team of analysts. The construction of the storyline had been meticulously crafted to excite the technology being tested as well as being militarily authentic.

Underpinning the workstations used in the trial is mature machine based decision-making technology, coupled with a sophisticated Human Machine Interface (HMI), asone of the aims of the overarching programmeis to mature such technology. This UK Ministry of Defence (MOD) funded work is seeking to inform the generation of requirements for future UK capability by building on other research/work undertaken to improve the HMI of, and decision support for, manned air platforms and ground-based planning & control systems. Furthermore, execution and demonstration of the trials work serves to illustrate novel concepts to a wide stakeholder audience and to raise the level of debate about such concepts and their impact on future operations. Ultimately the work serves to preserve the government "intelligent customer" status with regard to the desirability and affordability of such advanced concepts.

3 Description of Experiment

The trials have been designed around a classic SE model. Military Players are equipped and immersed in a sophisticated SE. In the command sense,a high command (known as the White Force) execute all roles not formally being assessed in the trial. They execute and manage the Master Events List (MEL)which acts as stage direction. The MEL is designed to inject into the SE various events that cause the mission to proceed. The injections will occur by directly manipulating the SE (e.g. by initiating an enemy attack) or, communicated by chat or radio, various mission directives. The injected events are designed to exercise key aspects of the technology in question. For *Trial Caucasus*, 12 runs were carried outeach of 85 minutes duration. The run matrix allowed the control of learning effects, provided for sufficient novelty in terms of events experienced by the players, and exercised the technology and concepts sufficiently to gather data for meaningful analysis. Players received 3 days of training in the week prior to the event. Training was also provided for analysts and observers.

See Fig. 1. A baseline configuration (Box A), where both manned and unmanned players had the minimal toolset. Box B, where the UAS had a much higher degree of autonomy and tools, whilst the manned players retained the baseline capability. Box C, where the UAS retained the baseline capability and the manned elements had a greater level of decision support. Finally, Box D, where both manned and UAS had all the available tools and decision support aids.Hypotheses were proposed and tested according to this methodology. The following is a typical hypothesis:

> *CAUCASUS [Trial name] Concept[Box D] – 'Integration Manned-Unmanned Teaming'; improves efficiency and mission task performance by more than the sum of the efficiency and mission task performance improvement found between 'Current Ops'[Box A] and 'UAS Person to Purpose'[Box B] and 'CurrentOps' and 'Networked Manned-Unmanned Teaming' in isolation (i.e. the integration value is greater than the sum of the two parts) (box A-D vs. (box (A-B) + (A-C))*

There were two themes to the analysis. One was focused on the key interactions between Players and the functions team-work plays on mission performance. The second was focussed on the effectiveness of the tools and technology, their functionality and the viability of the P2P approach using logged quantitative data and rich post-run verbal debriefs.

[2]The Dstl [1] CAPTEAM methodology explored the four box configurations.Collaborative Adaptability Proficiency Test Evaluation Assessment Methodology (CAPTEAM) is a validated independent assessment provided by Dstl, designed to estimate mission efficiency by the metrication of Reward and Effort associated with critical mission events and decision processes. The metrics measured under CAPTEAM are as follows:

- Workload Time Pressure, Mental Effort and Stress;
- Replan Task Load Decision and Action;
- Situational Awareness Demand, Supply and Understanding;
- Decision Quality Confidence, Survivability, Effectiveness and Timeliness;
- Teamwork Collaboration, Communication, Shared Situational Awareness, Leadership, Support, Team Workload and Influence Power;
- Task Performance, Adaptability Proficiency and Probability of Mission Success;
- Tools Utility and Usability;
- Technical System Reliability, Confidence and Usability.

The metrics are split into Taskwork, scored by the Player, and Teamwork, scored by a Subject Matter Expert (SME) Observer.

[2]Combining both themes of analysis, a real-time scorecard was developed. The scorecard was completed by all the Players. All participants (SME and Players) were asked to score their effectiveness at specific tasks throughout each run.

The analysis was supported by a team of analysts who shadowed each Player within the trial. Analysts gathered rich data during trial runs and recorded the timelines of key events and observations. Following each run, a questionnaire was completed by each Player which contained a bespoke set of elements covering issues such as trust and usability, as well as serving the needs of the first theme of analysis.

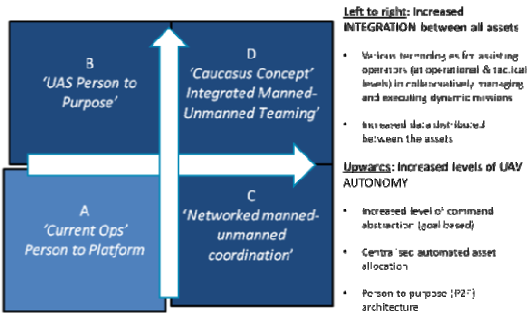


Fig. 1. The "Four Box" model of experimental design

---

[1] Dstl – The Defence Science and Technology Laboratory, part of the UK MoD.

[2] These paragraphs © Crown copyright 2013. Published with the permission of the Defence Science and Technology Laboratory on behalf of the Controller of HMSO

Fig. 2. Composite image showing White Force stations (top left), UAV Operator station (top right), Fast Jet station (bottom left) and Mission Commander station (bottom right)

4 Description of Technology

The descriptions below give an overview of the *with-tools* cases, together with the supporting technology. The technology utilised can be split, for convenience, into two broad themes: supporting infrastructure; and focus technology. This section deals briefly with the former but will discuss the latter in greater detail.

4.1 Supporting Infrastructure

A Virtual Battle Space 2 (VBS2) simulated world was provided, representing a geo-specific contemporary operational area on a 100km by 100km tile. Within this area, the terrain was a mix of mountainous and flat land, interspersed with civilian settlements of varying density. VBS2 scripts were constructed to exercise "canned" behaviours by entities such as people and vehicles. These entities and their actions were designed to stimulate actions from the Players and to provide background clutter and "confusers" to make the Player tasks more complex.

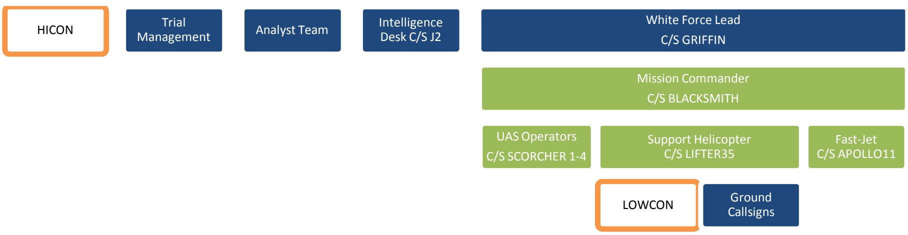


Fig. 3. TrialCaucasus Organization and C2 Structure showing 7 *Players* in the lighter shading with both high control (HICON) and low control (LOWCON) in the darker shading

Dynamic models were provided for all the air assets (UAVs, Fast Jet, Rotary Wing Support Helicopter) within the SE. The UAS comprised a pool of 3 Tactical UAVs (TUAV) and one Operational UAV (OUAV). The principal difference being that the OUAV had greater performance and was armed. Fig. 3. Shows the organisation of the trial and indicates the Player roles.

4.2 Focus Technology

Sitting between the Players and the simulated world was an array of decision making technology. This will be described in terms of that devoted to the UAVs, the Manned Players and the Human Computer Interface (HCI) or operator workstation.

UAVs. Automation of the UAVs is principally achieved through a QinetiQ developed technology known as the Task Execution Framework (TEF),[5]. This framework is a multiple, agent-based framework that views the overall automation task as the aggregation of atomic problems. It is constructed on the principle that any mission requires each atomic problem to be solved in an *ad hoc*manner in accordance with unfolding events - problems present themselves "randomly" as a function of external events or at the behest of the operator.Tools to solve or handle these problems and events will vary according to the nature of the problem. For example, the algorithms to optimally route a UAV around some controlled airspace are different from those required to schedule a series of asynchronous tasks of varying location and priority. Moreover, generated plans are subject to change following newly arrived events. The TEF provides a suitable mechanism to contain and initiate a range of planners, to execute plans, police those plans and to re-plan in the event of changes.

The military mission used to exercise the technology was focused on a long endurance ISTAR tasking. This acted as the motivation for deploying the Task Scheduler(TS). This planner, when given a series of Named Areas of Interest (NAIs) and a nominal priority,each of which constitutes a UAV task, will distribute those tasks in an optimal fashion. The aim of the planner is to reduce the dynamic planning burden on the Mission Commander and the UAV operators whilst maximizing coverage and minimizing redundancy and addressing the respective priority or weighting of the NAI. An intelligence report, or IntRep, was completed for each NAI and its completion signalled the end of the task. To eliminate system redundancy, whenever the imaging sensors were idle theywere utilised to service the "ISTAR Deck" by taking still shots of deck locations. For brevity, the collection of subject UAS technologies described here will be referred to as TEF.

A key aspect of the TEF design is its expectation that the operator will interact with plans as well as initiate them. This interaction is governed by the Pilot Authority and Control of Tasks (PACT) framework [4, 6,7].

Manned Players. The manned elements comprised both a single-seat Fast Jet (FJ) for Combat over watch and a Support Helicopter (SH). Tools for these players were principally focused on greater connectivity. Present Position Location Indication (PPLI) showed where the UAVs, FJ and SH were at all times, to represent the implementation of future network enabled mission co-ordination, as developed under a QinetiQ-led programme for the UK MoD. All platforms had a fused picture of ground entities, blue forces with common identifiers across the enterprise. Common reference points could be shared across the enterprise. Other tools were available to the manned platforms such as decision-support for airborne route planning and the ability to receive and mark up imagery from all entities.

Human Computer Interface. A key facet of realising the capability represented by the UAS TEF and associated planners is the means by which the user interacts with them. Too often such technology is limited by the "canned" nature of the plans and behaviours. Often the requirements to which the planning and behaviour technology is designed neglects the true needs of the user. The users' needs are often nuanced and almost inevitably, for the question "How would you do X?"the answer is, "Well it depends!"Therefore the approach adopted here concentrates on advancing the maturity of both planners and HCI in tandem and on a path informed by feedback and data obtained in trials such as that described here.

By way of example, a key part of the work undertaken in preparation for this latest trial was the continued refinement of the HCI elements that support the TS. An early version of the panel is shown at Fig. 4. This panel shows a series of planned tasks shown on a time chart stretching into the future. The numbers in each task refer to specific needs of that task. The tasks would be allocated to a specific operator. In keeping with the P2P philosophy this particular UAV platform (Rogue 1-4) is allocated to an operator – and this allocation would change throughout the trial run. As events unfold or as tasks are executed late, then the schedule would be updated and the tasks re-dealt to the platforms. Both operators and platforms are theoretically interchangeable with any operator not being bound to a particular platform. It is the

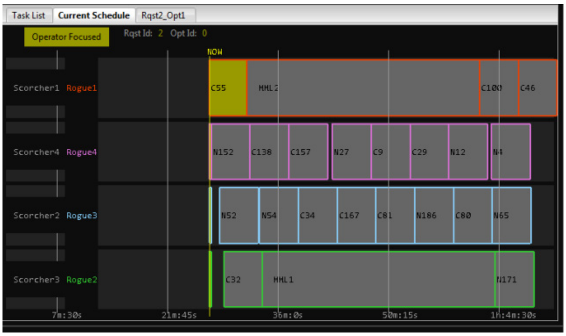


Fig. 4. Early Task Scheduler (TS) panel. Tabs at the top allow the user to specify additional *ad hoc* tasks and view a newly generated schedule prior to adoption.

impact of real-life mission constraints, such as that of only one armed asset or the need to maintain an individual's situation awareness for example,that drive the need to encompass additional controls over the TEF through HCI action.

5 Discussion of Results

Much data was collected addressing both of the analysis themes, P2P and CAPTEAM tools.

Initial analysis of the CAPTEAM protocol ratings, for only the UAS participants show benefits for systems B&D over systems A&C. The major differences between systems D and A were for the following metrics: Time Pressure, Decision Quality Confidence, Effectiveness and Timeliness, as well as the Performance metrics of Tools Utility, Adaptability Proficiency and Probability of Mission Success with D being the best configuration. It should be caveated that the above CAPTEAM results are all from observed trends and no statistical analysis has yet been performed. Therefore whether these differences are significant or not is unknown at the time of writing.

Turning to the P2P results and with respect to the prime tasking of servicing NAIs, a 40% increase was seen in the weighted coverage in boxes B and D. In these boxes the system allocated resources to task (TS), positioned assets and pointedthe sensor leaving the operator to interpret and report. See Fig. 5.

In terms of the background ISTAR Deck, where system harvests imagery automatically,an average76% increase in coverage was achieved over the baseline cases, where the operator had to manually service the deck. See Fig. 5.

Thequantitative scorecard ratings of task effectiveness (for both ISTAR and Combat tasks) showed clear benefits of systems B&D over A&C.For example,a 260% increase in the number of images taken for the ISTAR Deck. However, the actualoperational value of this increase remains to be evaluated.

Qualitative data gathered from the post run interviews supports the perceived routes towards adding value (reduced workload thus allowing greater mission focus & improved Situation Awareness(SA)).

"When [Task]*Scheduler works well it frees up lots of capacity"* (MC)

"*Wonders of autonomy - can just sit here now and just check what is going on. Like being in the 'passenger seat' and listen to everything that is going on. So have 'complete' SA."* (SCORCHER-1)

[after the final System A run] *"went ok but it's harder work than the morning run* [System D] *as have to make more decisions, have to monitor things more closely"* [anything you'd like to change?] *"Bring back system D"*(SCORCHER-3).

Some P2P asset confusion did occasionally occur, but not to the degree seen in previous trials. These issues will be alleviated by both additional system support in terms of improved HCI design and additional training.

Confidence in the perceived benefits of systems B & D was further increased by cross referencing all of the above results across different data sources.

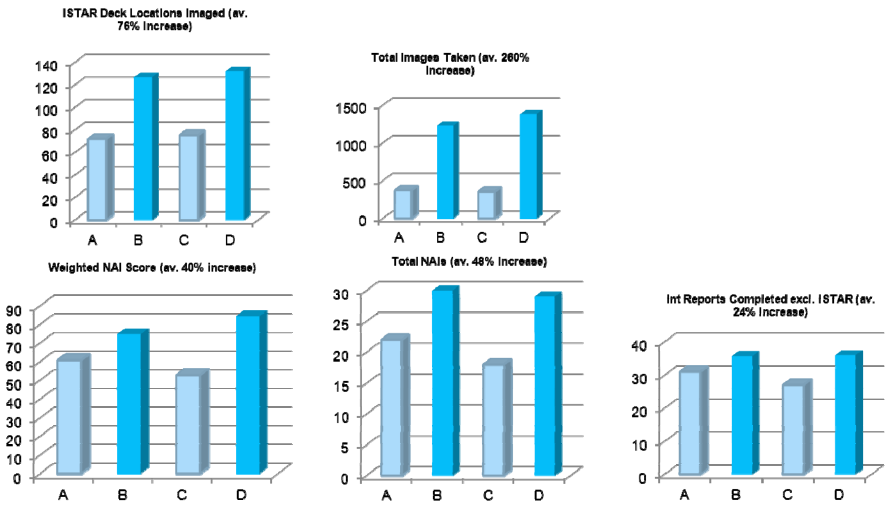


Fig. 5. Various metrics showing differences across each box configuration (A-D)

6 Conclusions

This paper gives a very brief insight into a large and multi-faceted trial event. One of the aspects addressed by the trial is the P2P concept where the operator is divorced from platform ownership and remains task focussed. Early indications are that overall system productivity and effectiveness was increased by the introduction of the P2P mode of operation and the supporting tools. However, deeper analysis continues.

Whilst the results do indicate tantalising benefits, enthusiasm must be tempered with the need to carefully calibrate the true extent of such benefits. For example, the ISTAR Deck automation clearly produces many more stills of areas of interest on the ground. Quantity of imagery is but one metric and simple automation in this case improved this measure. Clearly smarter processing could allow for fewer but better images to be collected and this,together with work guided by wider findings, remains the focus of future work.

Of particular interest to the P2P concept was the TS. This clearly demonstrated advantages to the operators and received many favourable comments backed up by quantitative evidence. However, there were also usability issues that mean the HCI will continue to be refined.

Acknowledgements. The authors are grateful to Dstl and, in particular, the UAS R&D Pipeline, for their continued funding of this work. The lead author is grateful to QinetiQ for the opportunity to work on this programme.

References

1. UAS Capability Investigation, Main Report, UK Ministry of Defence (April 2009)
2. Flight International (February 3, 2011), `http://www.flightglobal.com/news/articles/us-unmanned-systems-have-more-than-budget-problems-to-overcome-352717/`
3. Smith, P.R.S., Mayo, E., O'Hara, J., Griffith, D.: Combat UAV Real-Time SEAD Mission Simulation. In: AIAA Flight Mechanics Conference, AIAA-99-4185, Baltimore, USA (1999)
4. Platts, J.T.: Application of a Variable Autonomy Framework to the Control of Multiple Air Launched UAVs, AUVSI North America (2002)
5. Baxter, J.W., Horn, G.S.: Controlling Teams of Uninhabited Air Vehicles. In: 4th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2005), Utrecht, The Netherlands, July 25-29. ACM (2005) ISBN:1-59593-093-0: 27-33
6. Taylor, R.M., Abdi, S., Dru-Drury, R., Bonner, M.C.: Cognitive cockpit systems: information requirements analysis for pilot control of cockpit automation. In: Harris, D. (ed.) Engineering Psychology and Cognitive Ergonomics. Aerospace and transportation systems, vol. 5, ch. 10, pp. 81–88. Ashgate, Aldershot (2001)
7. Bonner, M., Diethe, T., Mathews, S.: Scoping Study for the insertion of Cognitive Cockpit adaptive automation taxonomy and control system into RTAVS for control of UCAV autonomy, QinetiQ internal study report (August 2001)
8. Platts, J.T., Kemsley, J.A., Richards, D.: Fast Jet Control of Multiple UAVs. In: Proceedings of the AUVSI North America Conference 2008, San Diego, CA (June 2008)

Human Factors and the Human Domain: Exploring Aspects of Human Geography and Human Terrain in a Military Context

Alex Stedmon[1], Brendan Ryan[2], Pat Fryer[3], Anneley McMillan[3], Nick Sutherland[3], and Alyson Langley[2]

[1] Cultural Communications and Computing Research Institute (C3RI), Sheffield Hallam University, UK
a.stedmon@shu.ac.uk
[2] Human Factors Research Group, The University of Nottingham, UK
{brendan.ryan,alyson.langley}@nottingham.ac.uk
[3] Helyx SIS Ltd, UK
{p.fryer,a.mcmillan,n.sutherland}@helyx.co.uk

Abstract. This paper introduces the concept of the Human Domain within military operations and considers how it has evolved from Cultural Geography into more specific sub-components of Human Geography and Human Terrain. At a high level, Human Geography and Human Terrain map across to strategic and tactical decision-making respectively. However, there is a confusing array of terminology and definitions surrounding these factors. Given this complexity, what might have originally been considered a Human Domain continuum from a strategic level down to a tactical level may be better represented as overlapping constructs on a spectrum of understanding, each with their own approaches to data capture and analysis.

Keywords: Human Domain, Data Models, Visualisation, Human Factors.

1 Introduction

Military operations are shaped by the characteristics of the environment in which they occur. Consequently, the scale, tempo, and complexity of military endeavours are linked to the physical and cultural landscapes of the regions in which they are conducted [1]. Globalisation and the influence of mainstream and social media within civilian society now combine to bring modern warfare, military manoeuvres, and operations other than war (OOW) into people's homes on a daily basis. In recent uprisings such as Arab Spring 2011 new communication technologies and social media were considered critical in coordinating the actions of the local populations [2]. In some ways this may be true, and military events at a tactical level now have the potential to shape public opinion. However, media usage and its impact on local populations, requires both access to the technology and a degree of literacy [3]. This highlights the need to understand underlying socio-demographic factors that are

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 302–311, 2013.
© Springer-Verlag Berlin Heidelberg 2013

essential in planning military interventions. Knowledge of the local area has a profound effect on military strategy as decision-makers seek to influence perceptions, understand sentiment and manage expectations amongst local populations. At the same time a sophisticated and increasingly media-aware and media-enabled enemy will try to exploit public opinion in order to destabilise and alienate populations from the ideals of their own governments and military forces. Understanding the population amongst whom the military must operate, and more specifically their culture, religion, perceptions and expectations, is now so important that it permeates through every aspect of military decision-making and military doctrine.

A former commander of coalition forces in Afghanistan highlighted the importance of understanding the local population with reference to concepts of Human Geography and the Human Terrain in relation to conducting counter insurgency (COIN) operations. Specific areas of difficulty are considered to be:

- the creation of location-specific, geo-social models and patterns of life that reflect spatio-temporal social, cultural, economic, political, religious, ethnic and other dynamic factors;
- the ability to assess local norms and detect deviations through modelling human behaviour and understand common patterns of life;
- understanding and capturing the effects of scale (e.g. nation, province, city, neighbourhood, sect, tribe, individual) and relating these to temporal aspects over different timescales;
- visualising motives, intentions, reactions, interactions between local populations and military forces;
- achieving deep levels of understanding to enable insight and provide opportunities to influence and support decision-making.

From these issues, the aim is to be predictive and understand possible trajectories of behaviour and plan for different ‘courses of action’ (COAs). The concepts of Human Geography and Human Terrain have been encompassed within the Human Domain that is now a critical element of conflict prevention and crisis management. In many ways this concept is still in its infancy although in reality the British military is re-learning lessons from its colonial past as well as the challenges it faced in the more recent troubles in Northern Ireland. Military doctrine is being re-written to take account of these lessons and is now being applied to future conflicts and embedded within information gathering procedures and information analysis structures [4].

With reference to the forward application of military knowledge, the future character of conflict (FCOC) seeks to identify the nature of military activities through to 2029 in order to guide the planning of organisational structures and operational activities. As global dynamics shift, key threats to stability include climate change, demographic trends, globalisation, access to declining energy resources, failed and failing states, and ideology. With these considerations in mind, the Human Domain is central to understanding and decision-making in military activities.

2 Terminology and Definitions

There are a number of technical challenges in understanding the Human Domain, not least in the proliferation of terminology and definitions that surround it and the concepts it encapsulates. Within the academic and military community, terms such as 'Human Geography', 'Human Terrain', 'Human Dimension', 'Cultural Geography', 'Human Dynamics', 'Patterns of Life' and even 'Human Factors' have evolved with very specific meanings and have been used interchangeably within various academic, doctrinal and operational contexts (see: [5]). Furthermore, many of the above terms have more than one meaning that may conflict with definitions for other concepts in other contexts. The landscape is therefore extremely complex and presents a challenge in determining a common understanding based on a lingua franca.

Human Geography is based on analysing the interconnections between people and places, including the spatial and temporal patterns of human activities in the context of the environment. Human Geography reaches from detailed local studies up to global trends. What gives Human Geography its disciplinary identity is its basis in spatial and temporal relationships. There are three primary themes in Human Geography, each taking account of tome and space [6]:

- economic geography - systems of production, exchange and distribution
- political geography - systems of class/group conflict
- social geography - systems of human interaction

In some ways definitions for Human Terrain seem to overlap with that of Human Geography, but focuses more on the characterisation of cultural, anthropological, and ethnographic information about a human population and their interactions within a joint operational area. Other interpretations include the notion of the soldier 'in the system' and this relates well to systems approaches within Human Factors (e.g. joint cognitive systems, socio-technical systems, etc) where people and technology interact beyond traditional and singular 'human-machine interfaces' (see: [7]). In relation to an environment, among which soldiers operate, Human Terrain must consider the impact and influence that a person has in a given system where others are already interacting.

To a certain extent, the proliferation in terminology and associated definitions is due to similar terms developing through different perspectives (for instance a geographical perspective versus an anthropological or social science perspective). Other differences occur due to an academic concept being applied to a military context. In some cases, context specific terminology may increase understanding within a particular domain or discipline as the terminology is related to other specific ideas and concepts. However terminology specialisation might confuse a common understanding between different domains or disciplines. This is particularly important within the context of FCOC, where concepts such as the centrality of influence, dynamic understanding, adaptive approaches and increased partnerships with allies, all require a common understanding in order to work in an agile and efficient manner.

3 De-constructing the Human Domain: Human Geography and Human Terrain

The Human Domain can be understood in terms of higher-level strategic concepts that relate to aspects of Human Geography, down to lower-level tactical concepts that relate to aspects of Human Terrain. In many ways these two factors represent a continuum but given their levels of complexity it may be more accurate to consider them as overlapping constructs on a spectrum of understanding and analysis, each with their own approaches, some of which may be in common (Figure 1).

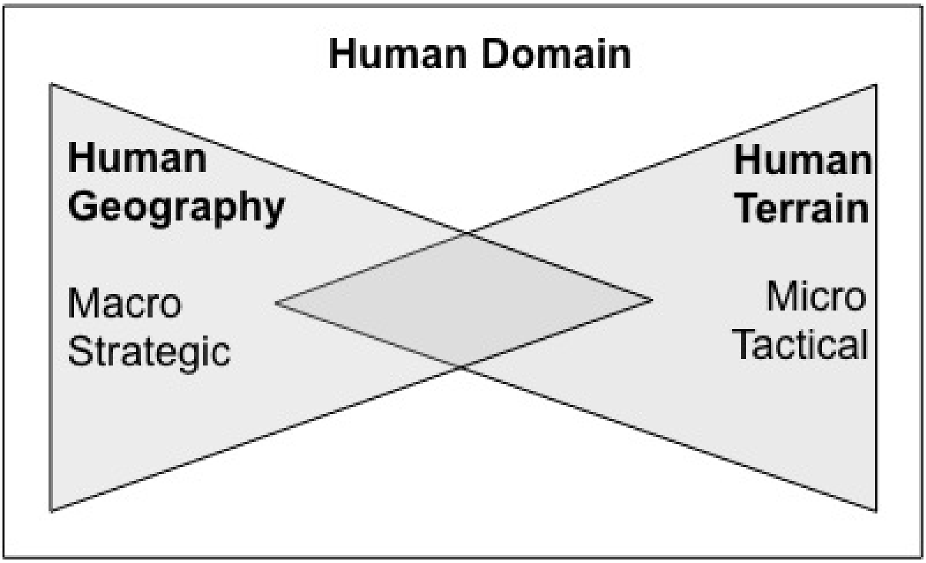


Fig. 1. Overlapping constructs of Human Geography and Human Terrain

3.1 Human Geography

Human Geography encapsulates the macro-level detail of a situation and attributes that are unlikely to change in the short term (e.g. physical characteristics of time and space such as buildings, runways, radio masts, etc) and also perhaps stable attributes of a population that may remain largely stable over time (e.g. social identity, ethnicity, etc). These provide a basis for strategic decision-making based on known data that have been gathered and assimilated over a period of time. That said, the period of time over which data are collected is important. Information changes over time in light of new intelligence about a situation. In due course it will be updated when the new data are processed and their credibility confirmed. For example if a bridge has been compromised, intelligence reports will be fed back at a strategic level so that maps and other decision support tools can then be updated. For social factors this may be more challenging. It may be possible to map aspects of general sentiment for a given population (e.g. attitudes are likely to remain static for periods of time) but perhaps more difficult to map sentiment for more dynamic factors (e.g. anger at perceived injustices). However, as societies evolve over time, so do the social relations that are:

- established through space (such as the organisation of resource-based activities and environments);
- constrained by space (such as the inertia imposed by the built environment or the limits imposed by natural hazards);
- mediated through space (including the development of ideology and beliefs within geographically confined regions).

In many ways this aspect of Human Geography overlaps with elements of Human Factors. Whilst Human Geography is defined by economies, cultural identities, politics and society, Human Factors focuses on the study of people and human interaction within specific contexts [7]. Many aspects shaping human performance and the way Human Factors considers complex systems are relevant to Human Geography. The temporal dynamics and location-based perspective to Human Geography translates across to fundamental issues in Human Factors such as human performance, workload and situation awareness, where the ability of people to conduct tasks under increased pressure or when dealing with uncertainty are both time and location dependent.

3.2 Human Terrain

Human Terrain is defined within Army Tactical Doctrine as the social, political, and economic environment, belief systems and forms of interaction of the people among whom soldiers operate and includes social activities, political frameworks, organisations, people and events [8]. In addition to the identification of such interactions it is critical that they are understood (or interpreted) from the perspective of the local population based on an understanding of what that population wants, needs, and fears in their daily lives [8].

An understanding of the Human Terrain can enable soldiers to more readily distinguish friend from foe [8]. This is critical in operations where soldiers are tasked with operating close to a local population and interacting directly with local forces. Military action, if grounded in an understanding of the Human Terrain, can help build relationships and trust with the local population. This can then be used to foster a sense of safety and security, both of which are foundations for future opportunities and development [8]. In addition, an increase in Human Terrain understanding can also be used to ascertain points of influence, leverages and vulnerabilities within the local population [8]. These can then be avoided when planning operational objectives, all of which should ensure a balance between the officially conveyed command intent and the perceived intent that the local population might derive from the behaviours they observe [8]. In this way, Human Terrain encompasses planning and decision-making in a dynamic environment (relevant to some of the other complex systems/environments in which Human Factors operates).

At a tactical level, Human Terrain reflects the dynamic aspects of time and space (and underlying issues of uncertainty and trust in changing datasets). Human Terrain reflects the detail of a situation experienced at an individual or group level as events unfold in real time. Information is processed based on assumptions about the data available (and perhaps referenced back to known Human Geography analysis). At a simple level, whilst the location of a population might remain stable over a period of time, it may well be transient within its boundaries (due to natural disasters, civil unrest, etc). Mass migration is a major issue in conflict when populations shift and need to be relocated/re-housed until they move back again and then resettle in often war-torn areas. At a tactical level, a radio mast may appear on the terrain when it previously was not there and therefore represents a dynamic attribute that may eventually become part of the Human Geography dataset if it remains in place and is authenticated. From a Human Terrain perspective not only is the radio mast present as a physical attribute in the landscape, but it also has a range of transmission (given the local topography and prevailing atmospheric and weather conditions). From the Human Domain perspective more subtle questions may be asked about what is being transmitted and what kind of influence does a particular message have over the local population, given the prevailing dynamics of a situation.

4 Re-constructing the Human Domain

From a military perspective, the Human Domain has evolved from aspects of Cultural Geography and involves the military examination of human components within an operational environment, including an integrated picture of the cultural landscape [1]. The result is an appraisal of a region's population, cultural groups, cultural institutions, settlement patterns, land use, economies, transportation and communication networks, and military capabilities [1].

The development of the Human Domain from Cultural Geography provides a basis for exploring aspects of Human Geography and Human Terrain. There is a clear emphasis on socio-cultural factors but linked to this are underlying economic and political factors. One example from a recent military workshop was that in building a new road to support military logistics, a lucrative contract awarded to a local company could have serious consequences if that contractor then develops local political power through being able to buy or bribe support from the local militia. A further example is that in situations where access to water is controlled through canals or sluice gates, different local actors can influence others and exercise power by their control of the waterways. Whilst these situations might be predictable if the right information is gathered and analysed, other situations might be more of a challenge. If a military unit ‘promises’ to support the local community (perhaps by refurbishing some premises, helping to repair a bridge, etc) but the verbal contract is lost between different rotations of military personnel, then the local community may be reluctant to trust or believe in new promises.

As the Human Domain incorporates the relationships between people and their natural and cultural environments, this relates well to fundamental aspects of Human Factors. There is also a focus on developing and preparing soldiers for specific tasks and this has similarities in training and cultural change within conventional organisations and understanding complex socio-technical systems from a Human Factors perspective. A key output is an enduring understanding (that evolves and matures) that is essential in briefing the rotations of military units through an area/region and providing immediate context for them when they arrive in an area.

5 Human Domain Methods and Interpretation

In a broad sense the Human Domain focuses on observed distributions and analytical explanations that explore social, economic, cultural, political, and demographic dimensions of human existence [9]. However, the Human Domain also includes temporal aspects and location-based perspectives to understand system dynamics.

Whilst there is nearly always data available, all too often the challenge becomes one of identifying the relevant data for a given situation [10]. As a consequence, there is often not enough 'good' data that is reliable and relevant, and what there is, is often incomplete, obsolete, inaccurate, and difficult to validate [10]. What emerges is a lack of detail that reflect spatio-temporal social, cultural, economic, political, religious, ethnic and other dynamic factors. Embedded within these are potential anomalies that could indicate areas of interest and so clearer methods are needed to capture, analyse, understand and ultimately visualise this information and link it to effects of scale and time, and map the motivations, intentions, reactions and interactions that will then provide both strategic and tactical insights for decision-making.

As a result, the collection and analysis of Human Domain data involves quantitative approaches (e.g. population numbers, demographic ratios, etc) and qualitative approaches (e.g. in-depth interviews, observations, ethnographic approaches, etc). Human Factors shares a number of these methods. From a methodological perspective the Human Factors community can help with eliciting information and interpreting the Human Domain. This is a vital contribution that Human Factors can offer and support the military who may not perhaps be used to questioning and collecting information from a civilian population. However, the manner in which such knowledge is produced, and what it means, is dependent on wider political, institutional, and cultural circumstances from where it came, and also on the perspective of the observer or interpreter who collected it [9]. In order to achieve this, it is useful to consider how information might be assimilated into information systems via data models and how that information might then be visualised in a way that is usable and fit for purpose.

5.1 Information and Data Models

In order to assimilate, classify and structure Human Domain data, the use of information and data models has been advocated. Building on political, economic and social constructs of geography (as mentioned earlier) these models can range from conceptual data management frameworks, such as the PMESII/ASCOPE frameworks to comprehensive military models such as the GEOINT Structure Implementation Profile (GSIP) designed to capture different levels of detail across the Human Domain.

PMESII sets out to develop a holistic understanding of Political, Military, Economic, Social, Information, and Infrastructure entities and is aligned to Human, Social, Cultural and Behavourial (HSCB) approaches [5]. ASCOPE builds on the PMESII framework by providing levels of detail based on Area, Structures, Capabilities, Organisations, People, and Events [4]. GSIP defines the methods and structure for specifying and encoding data, and includes specific sub-models that can be extended to cover aspects of Human Geography and Human Terrain data.

With reference to temporal dynamics and a location-based perspective, GSIP will form the basis of the NATO Geospatial Information Framework that aims to specify the content, structure, organisation and exploitation of geospatial information services, delivered over a NATO common geospatial information infrastructure. It will be built on a common content model that includes a feature content dictionary, feature catalogue and data model. Although these models have great capacity to structure data collection, and establish linkages and hierarchies of data, the use of these models also provides their own challenges.

One of the main challenges with the use of data models is that they are currently content-impoverished. This tends to be because Human Domain data does not yet have standard semantics or formats. The development of data standards through the creation of dictionaries, logical models, ontologies, application schemas, feature catalogs and encodings is therefore imperative work to support the population of these data models.

A further challenge is in the analysis of data at the right 'level of detail' for different users. This also means that the analysis must be done in a way that supports military collaboration with non-military agencies who may not be familiar with military concepts. Different levels of detail allow for a common framework of understanding within the Human Domain, from high-level strategic Human Geography detail (e.g. the cultural/social perspective) down to lower-level tactical Human Terrain detail (e.g. the individual perspective).

5.2 Visualisation

Once Human Domain data are collected and analysed, further challenges are faced in visualising the results. Traditional challenges are faced when dealing with Human Geography data in terms of representing the accuracy and precision of the data, generalising big data into products at smaller scales, dealing with data gaps, and displaying temporally dynamic data embedded within the Human Terrain. Further challenges are faced when attempting to visualise novel data (e.g. social media data) and representing uncertainty. Visualising sentiment and behaviour is likely to bring its own challenges, particularly if the data provenance is unknown. The solution is likely to be flexible, adaptable, potentially nested frameworks and methodologies that can deal with a range of data types in as consistent manner as possible. It is perhaps in this area that Human Factors can draw from established work in the discipline.

Within Human Factors, a number of topics are relevant to visualising Human Domain data: the interaction of people within systems, including methodologies for studying interactions; participatory approaches for user needs and knowledge elicitation; representing uncertainty within user interfaces and visualisations; trust (building trust and relationships) and transparency; planning and decision making in complex, dynamic systems (including systems based approaches); ethnographic and cultural assessment; and aspects of simulation and training.

Returning to the example of the radio mast, if such a feature is observed at the tactical level (e.g. a pilot spots it on a fly-by) but it is not then incorporated at a strategic level into databases or products (e.g. maps), levels of mistrust can result. This can undermine user confidence in existing products (e.g. reliability of the data) based on an assumption that other data may be left out if it has not been authenticated. The time it might take to verify the data also influences the temporal aspect of the Human Domain. If there is a delay in validating data sources portrayed on a map, it may be important to relate this to the user so they know which data may be less certain than others.

6 Distinctions in Human Geography and Human Terrain

Any localised military geographical analysis represents a partially processed snapshot of the operational environment at a specific point in time [1]. Earlier military geographical analyses must be revisited and continually updated in order to account for the changing and enduring attitudes and perceptions of a local population, any humanitarian efforts to help the country, and the continuing political, economic, and cultural influences from outside sources [1]. As experience has shown in countries such as Germany and Japan, and more recently in Bosnia and Kosovo, in the aftermath of war sustained peace can be achieved only through effective stability-and-support operations and long-term political and economic commitment [1]. This is a major issue as military objectives such as reconstruction often requires collaboration with Non Government Organisations (NGOs) and international bodies such as the United Nations and World Bank. In order for these activities to be effective and efficient, the data collected and the understanding gained should be sharable with other non-military players. A tactical understanding should be incorporated into all aspects of the planning process and intelligence collection process from: the identification of baseline and comparative data; pre-planning information gathering; to the development of the intelligence and potential effects; through to the understanding of COAs and decisions on control measures [8]. However, the appropriate level of detail is important. For example, an overlay of tribal locations for a particular area of interest may obscure particular dynamics of the environment that can impact on specific COAs and influence opportunities. Furthermore, from the Human Domain perspective, the boundaries of one tribal grouping may not meet the boundaries of another (e.g. tribes will fall across many political and administrative boundaries) making it difficult to collect all the information required to understand the tribe. Thus, integration should ideally be complementary to other forms of intelligence and not be treated as an additional component that can be dealt with in isolation [8].

7 Conclusion

Being aware of the Human Domain and its impact on operations is a fundamental requirement for planning and conducting military activities: the Human Domain is important to the military perspective. Furthermore, the Human Domain underpins a highly dynamic situation where different components may interact, often creating or at least adding to already difficult circumstances within which military units must function [1]. By creating a better understanding of Human Geography (e.g. the context and the arena in which activities take place) and Human Terrain (the importance of combining different user perspectives) there is a clear opportunity to draw from the knowledge-base and methodological toolbox of Human Factors.

Acknowledgement. This work has been supported and developed through the Advanced Geospatial information and Intelligence Services (AGIS) research project (MOD Contract No: DSTLX- 1000063699).

References

1. Palka, E.J., Galgano, F.A., Corson, M.W.: Operation Iraqi freedom: A military geographical perspective. Geographical Review 95(3), 373–399 (2005)
2. Campbell, D.G.: Egypt Unshackled: Using social media to @#:) the System. Cambria Books, Amherst (2011)
3. Pfeffer, J., Carley, K.M.: Social networks, social media, social change. In: Proceedings of 4th International Conference on Applied Human Factors and Ergonomics (AHFE 2012), San Francisco, July 21-25 (2012)
4. Hartley III, D.S., Lacy, L.: Creating the foundations for modeling irregular warfare. In: Proceedings of 4th International Conference on Applied Human Factors and Ergonomics (AHFE 2012), San Francisco, July 21-25 (2012)
5. McCloskey, M.J., Behymer, K.J.: Methods for capturing cultural lessons learned and training cross-cultural skills. In: Proceedings of 4th International Conference on Applied Human Factors and Ergonomics (AHFE 2012), San Francisco, July 21-25 (2012)
6. Dear, M.: The postmodern challenge: Reconstructing human geography. Transactions of the Institute of British Geographers, New Series 13(3), 262–274 (1988)
7. Hollnagel, E., Woods, D.D.: Joint Cognitive Systems: Foundations of Cognitive Systems Engineering. CRC Press, Boca Raton (2005)
8. de Vries, A.M.: Analysis for the future of Afghanistan. The human terrain of counterinsurgency operations: Developing the military mindset and social science support. DSTL on behalf of the Controller of HMSO (2010)
9. Gibson, C.: Human geography. In: Kitchin, R., Thrift, N. (eds.) International Encyclopaedia of Human Geography, vol. 5, pp. 218–231. Oxford, Elsevier (2009)
10. Moore, R.A., Stevens, C.M., Oonk, H.M., Averett, M.G.: Operational user requirements and example use cases for Human Social Culture Behavior (HSCB) technologies. In: Proceedings of 4th International Conference on Applied Human Factors and Ergonomics (AHFE 2012), San Francisco, July 21-25 (2012)

Part IV

Cognitive Issues in Health and Well-Being

Web-Based Architecture for At-Home Health Systems

Tiffany Chua and Mark Bachman

University of California, Irvine, Campus Drive, Irvine, CA 92697
{tchua,mbachman}@uci.edu

Abstract. Technologies are revolutionizing the way health care is managed and delivered, especially in the areas of telemedicine and home care. Many at-home e-Health products are being developed and brought to market, but one of the factors that prevent widespread adoption is the need for customized solutions that result in lack of economies of scale. In this paper, we describe our Web-ENabled Devices and Instruments (WENDI) platform that addresses this challenge, utilizing technologies that promote low-cost and rapid development of web-based applications.

Keywords: e-Health, telehealth, web-based architecture.

1 Introduction

Advances in information and communication technologies are revolutionizing the way health care is delivered and managed. One of the greatest areas of expansion in e-Health is in telemedicine and home care, which delivers personalized health systems and services to the home. A number of factors have contributed to its growth: the consumers' desire for greater involvement in their own health care, and their desire for more high tech products, diminishing levels of family care [1], the rise of chronic diseases [2], etc. In many developed countries, the increase in the aging population relative to the size of the workforce in the health sector is the major contributing factor.

While innovation in this field has continued and home e-Health products have proliferated, there are a few challenges that prevent widespread adoption and integration of at-home health systems. One is that the fragmentation of public demand and the need for customer-individualized solutions result in technological delay and lack of economies of scale [3].

In this paper, we describe a web-based architecture and platform for at-home health systems that partly addresses this challenge. Our solution utilizes web technologies that promote low-cost and rapid development of customized applications. We also present an example implementation of a customized system using this platform that provided heart rate variability (HRV) biofeedback for stress management.

2 Web-Based Architecture

Web architectures for e-Health systems generally follow a three-tier architecture [4], shown in Figure 1.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 315–322, 2013.
© Springer-Verlag Berlin Heidelberg 2013

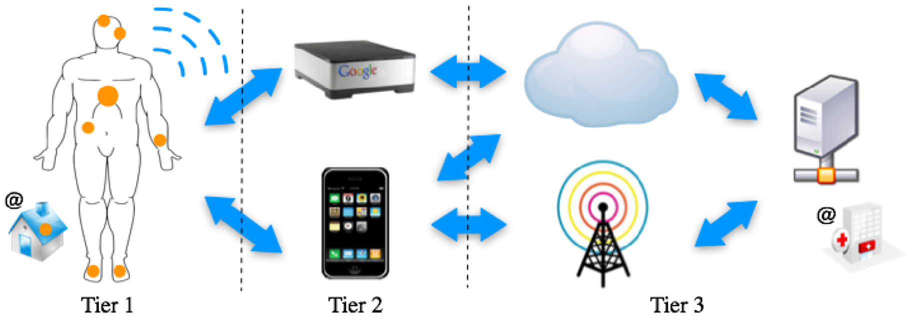


Fig. 1. General three-tier web-based architecture for e-Health systems: (1) data tier, (2) logic tier, (3) database and presentation tier

Tier 1 is the data tier, consisting of sensor networks that collect and pre-process physiological data, usually located on the body, but may also include ambient sensors around the home, such as temperature and humidity. It may also include actuators or other devices that provide feedback to the user. Tier 2 is the logic tier, which consists of a personal server or home gateway that manages the sensor networks and performs additional data processing or analysis if needed. It may also temporarily store data for later viewing or transmission to tier 3. Tier 3 is the database and presentation tier, which provides long-term data storage and presents the data through a web browser for web-based applications. For systems with telemedicine functions, tier 3 consists of a web portal or web application hosted by a server located at a hospital or health institution. For local systems, the web application and database may be hosted by the home gateway.

2.1 Tier 1: Sensor and Actuator Networks

Sensor and actuator networks are often heterogeneous, consisting of sensors used for different purposes such as ECG sensors, pulse sensors, blood pressure sensors, and temperature sensors. They may be wired or wireless, and may use different types of communication technologies -- general purpose technologies such as USB, Wireless USB, Wi-Fi, Bluetooth, and Zigbee, or proprietary technologies such as ANT, Sensium and Zarlink. To some degree, some technologies such as USB, Bluetooth and Zigbee provide integration mechanisms to promote interoperability. For example, Bluetooth has a Serial Port Profile, which sets up virtual serial ports between two devices using serial port emulation. To achieve higher levels of interoperability, a number of standards have been proposed, such as ISO/IEEE 11073 and Health Level 7. These standards may be complementary with each other [5].

2.2 Tier 2: Personal Server or Home Gateway

A personal server is usually a smartphone or mobile device that has constant connection to the body sensor network, supporting patient mobility. However, it has power and energy limitations, and higher cost of data transmission. On the other hand, a home gateway, although not always connected to the patient, can support more communication channels, can have wider network bandwidth and lower-cost data transmission to the next tier. Examples of home gateways include desktop computers, mini computers, and set-top boxes [6].

2.3 Tier 3: Database and Web Application Server

The server on tier 3 hosts a database server for storing data and a web server for hosting web applications, which a web client can access. The main advantage of using a web-based architecture is that the resulting client-server system is platform-independent. By using service-oriented architectural styles such as RESTful and SOAP- or RPC-style architectures, a loosely-coupled distributed system can be built. For example, data processing and analysis can be performed on the server-side while data visualization is implemented on the client-side. The client exchanges information with the server using well-defined messages – for RESTful services, using a uniform interface based on simple and universal HTTP operations such as PUT, POST, GET and DELETE, and for SOAP- or RPC-style services, using an application-dependent programming interface that supplant HTTP operations. In both types of services, clients connect to servers using techniques that are native to the web, but are limited to request/response type of communication. Other web technologies that support bi-directional data streaming, such as Comet or websockets, complement these services [7].

Using a web application or a web portal as a front-end brings several advantages. First, it ensures universal access, as every modern computer, laptop, netbook, tablet and smartphone has a browser. Second, installing and deploying the user interface is almost instantaneous. Third, managing, maintaining, and modifying applications is easier, since only the web application on the server-side needs to be updated [4]. Lastly, with more reusable components, developing web applications is a low-cost and rapid process compared to traditional client-server applications. With the advent of HTML5 technology, running media-rich interactive web applications on browsers is possible.

3 System Architecture and Design

For e-Health systems, the solution usually includes both hardware and software to provide a platform that collects, processes and presents data from sensors, and/or transmit data for further analysis or storage [8]. In this section, we describe our Web-ENabled Devices and Instruments (WENDI) platform, a hardware and web-based software solution for at-home systems.

Figure 2 shows the hardware components of an at-home system. It consists of a sensor and actuator network that is connected to a home gateway, a Linux-based mini-computer with standard interfaces such as USB and Bluetooth. The home gateway is connected to a router, which provides a local wireless network for the at-home health system. Client devices such as laptops, tablets and smartphones or any other device with a web browser can connect to the home gateway through the Wi-Fi network. For telemedicine or social networking functionality, the router can be connected to the internet.

Fig. 2. Components of an at-home health system using the WENDI platform

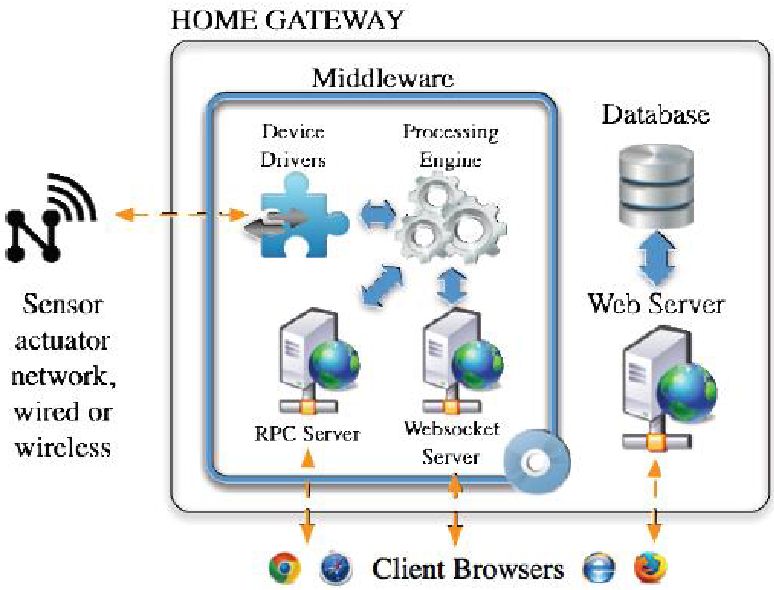


Fig. 3. Components of WENDI home gateway

The sensor and actuator network may include any device with serial port emulation, such as USB-serial and Bluetooth device with a serial port profile. Support for other types of devices may be added in the future.

The home gateway is an off-the-shelf mini-computer (Asus EEEBox 1007) running a Debian Linux-based operating system (OS). It contains middleware that collects, processes, and transmits data from sensors, as well as a database where data is stored, and a web server that hosts web applications. The components of the home gateway are depicted in figure 3.

The middleware on the home gateway was written in Python. The middleware hides the complexity and heterogeneity of underlying hardware, allowing programmers to focus on the retrieving data from sensors without knowing how to communicate with individual sensors, actuators, or other devices. It has a petrinet-based processing engine for analyzing data. It also includes an RPC server, which has basic commands to start and stop data acquisition, and a websocket server, which streams sensor data to client browsers. The middleware can be easily extended and customized to provide additional functions for applications.

A javascript client library was created to facilitate communication with the server. It has functions for creating a socket to communicate with the RPC and websocket servers, and functions to send commands / receive responses from the RPC server as well as process messages from the websocket server.

Figure 4 shows the deployment diagram for WENDI. On the home gateway, the platform relies on OS drivers and components for data acquisition and control of sensors, actuators or devices, such as the USB and Bluetooth device managers. A standard web server and relational database management system packaged for the OS are included for building customized web applications. The middleware consists of several modules: wendi_sources.py, which receives incoming data from sensors and devices, wendi_sinks.py, which sends outgoing data to actuators and devices, wendi_pn.py, which is the petrinet-based processing engine that takes the input data, processes the data, and produces output data, wendi_server_wamp.py and wendi_server_websocket.py, which handle RPC and communication over websockets, respectively. Libraries for RPC and communication over websockets are provided by the Autobahn Python package [9]. On the client device, a corresponding javascript library is provided to send requests and receive commands to the RPC server, and to send/receive streaming data over websockets. The wendi.js client library uses a library provided by Autobahn JS [10].

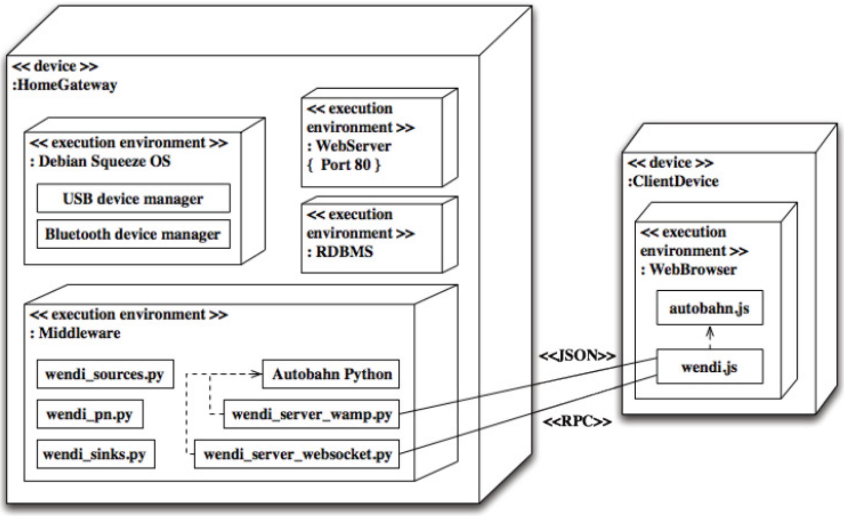


Fig. 4. Deployment diagram for WENDI

4 Example Application: HRV Biofeedback for Stress Management

In this section, we present an example application utilizing the WENDI platform. The objective in this application was to present heart rate variability (HRV) data from a pulse sensor to the user for biofeedback in stress management. A healthy heart is influenced by multiple factors that result in variations in heart beat intervals (HBIs), called HRV, at time scales ranging from less than a second to 24 hours. Ideally, the HRV waveform should contain strong sinusoidal components. However, when one is experiencing feelings of anger or frustration, the HRV waveform shows a disordered, jerky waveform [11]. The change in HRV is one of the body's secondary physiological responses to stress. The body also has natural tools for combatting stress, one of which is breathing at a slow rate in synchrony with your nervous system. Previous research has shown that by using the body's feedback loop, this breathing technique enhances the state of your nervous system, and induces feelings of calm and relaxation [12].

A web application was built to educate users in using breathing exercises as a stress management tool and provide HRV biofeedback for patients to monitor their progress. On the server side, pulse data was collected from a plethysmographic sensor located on a glove. A microcontroller on a hand glove acquired pulse data from the plethysmographic sensor positioned on the index finger, pre-processed and transmitted the data via Bluetooth to the home gateway, which performed further data processing and analysis. Processing involved detecting peaks in the pulse waveform,

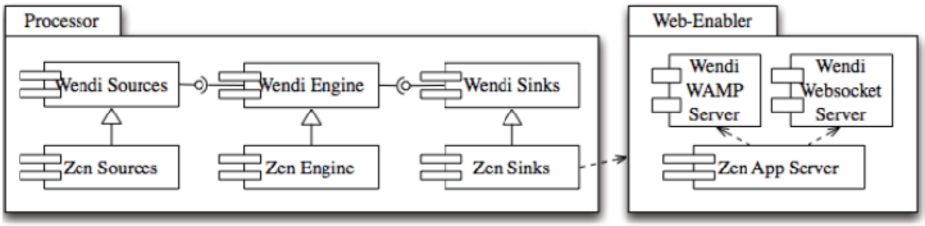


Fig. 5. Component diagram of extended middleware

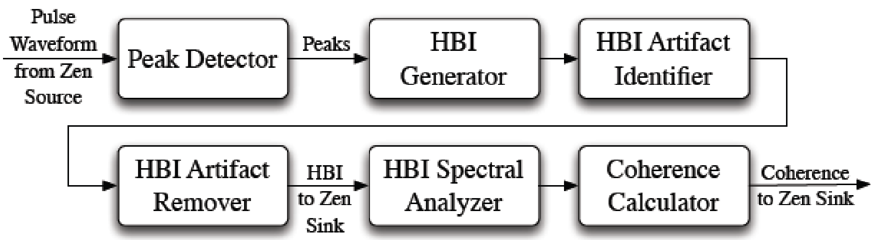


Fig. 6. Block diagram of Zen Engine for HRV biofeedback

extracting the HBIs, and calculating coherence. Coherence was a number between 0 to 100 that reflected the relaxation state of the user over a period of time; a higher figure indicated that the user was more relaxed. HBIs and coherence data were delivered to a client browser on a tablet, where data was visualized. Figure 5 shows a high-level component diagram that illustrates how middleware on WENDI was extended to provide other application-dependent functions. Figure 6 shows the block diagram of the Zen Engine.

On the client-side, a standard web application was built to educate and train patients on breathing techniques for stress management. The client communicated with the web server and database server as part of its standard web application, and used the javascript client library for WENDI platform to control data acquisition through RPC and obtain streamed data from the server through websockets.

5 Summary

In summary, we have presented a web-based platform for at-home health systems that utilizes web technologies that enable low-cost and rapid development of customized applications with rich user interfaces. The platform is modular and extensible, allowing programmers to extend the functionality of the platform with user-defined components such as sensor drivers and signal processing modules. The platform also provides an easy way to add the ability to acquire, process and stream data to existing web applications.

Currently, the platform is designed to be used by systems integrators, who have the ability to extend the functionality of the platform with plugins. More work is needed to transform the platform into one that can be used by domain experts, who may have basic IT skills, but need higher-level abstractions that simply the configuration of the system. Other areas for future work include utilizing existing standards such as IEEE 11073 and Health Level 7 for interoperability with other devices and systems, discovery of sensors, actuators or devices, management of devices and network resources, etc. Future work will address these issues.

References

1. Houser, A., Gibson, M.J., Redfoot, D.L.: Trends in Family Caregiving and Paid Home Care for Older People with Disabilities in the Community: Data from the National Long-Term Care Survey. AARP Public Policy Institute (2010)
2. Healthy Communities: Preventing Chronic Disease by Activating Grassroots Change. National Center for Chronic Disease Prevention and Health Promotion, Division of Adult and Community Health (2011)
3. Accelerating the Development of the e-Health Market in Europe. European Commission, Information Society and Media (2007)
4. Shopov, M., Spasov, G., Petrova, G.: Architectural models for realization of web-based personal health systems. Presented at the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing, New York, USA (2009)

5. Aragues, A., Martinez, I., Valle, P., Munoz, P., Escayola, J., Trigo, J.: Trends in entertainment, home automation and e-health: Toward cross-domain integration. IEEE Commun. Mag. 50, 160–167 (2012)
6. Spasov, G., Petrova, G.: Web-based Personal Health Systems – Models and Specifics. Presented at the International Scientific Conference Computer Science (2008)
7. Guinard, D., Trifa, V., Wilde, E.: A resource oriented architecture for the Web of Things. Presented at the 2010 Internet of Things (IOT) (2010)
8. Korhonen, I., Parkka, J., van Gils, A.M.: Health monitoring in the home of the future. IEEE Engineering in Medicine and Biology Magazine (2003)
9. Autobahn Python, http://autobahn.ws/python
10. Autobahn JS, http://autobahn.ws/js
11. Tiller, W.A., McCraty, R., Atkinson, M.: Cardiac coherence: A new, noninvasive measure of autonomic nervous system order. Altern. Ther. Health Med. 2, 52–65 (1996)
12. Lehrer, P.M., Vaschillo, E., Vaschillo, B.: Resonant Frequency Biofeedback Training to Increase Cardiac Variability: Rationale and Manual for Training. Appl. Psychophys. Biof. 25, 177–191 (2000)

Inclusive Design: Bridging Theory and Practice

Anita H.M. Cremers, Mark A. Neerincx, and Jacomien G.M. de Jong

TNO, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands
{anita.cremers,mark.neerincx,jacomien.dejong}@tno.nl

Abstract. Large groups in society lack the necessary skills to be sufficiently self-reliant and are in need of personal assistance. These groups could be supported by information and information technology (ICT), but only if this technology is designed to fit their (cognitive) abilities. Inclusive design theory and methods have already been developed in research contexts, but there is still a gap between theory and practice. There is a need for a practical aid, that helps to create awareness of inclusive design among ICT developers, and offers easy-to-use information and tools to actually apply the methods for diverse target groups. This paper describes the first steps taken towards an inclusive design toolbox for developing ICT applications that offer cognitive support for self-reliance. Dutch ICT companies were interviewed and participated in a co-design workshop, leading to a number of initial needs, user requirements, and an on-line community, that form input for further development of the toolbox.

Keywords: Cognitive abilities, toolbox, design patterns, personas, inclusive design methods, ICT, self-reliance, SME, co-design.

1 Introduction

Information and communication technology (ICT) has a large impact on personal and social lives of people. More and more, in order to access, request or provide information (e.g. to make personal choices in health care), and to participate actively in society (e.g. to use social media to maintain contacts with peers), people need to be able to make use of this technology. Also, the government expects citizens to become more self-reliant. Large groups in society, however, lack the necessary skills to be sufficiently self-reliant and are in need of personal assistance. These groups could be supported by information and information technology, but only if this technology is designed to fit their abilities. In that way, technology is not another barrier, but would instead serve as a means to achieve self-reliance.

Diverse groups exhibiting limited self-reliance include people with specific physical and cognitive limitations, ageing people and people with a low education and/or a low socioeconomic status, all of whom adhere to specific values in life. Self-reliance is relevant in many areas of society, but in particular in social security (absence of threats as a result of criminal acts, offenses, serious nuisances of other citizens), health care and well-being (make personal choices, life style, adherence to therapy), and participation (education, work, social engagement). An important determinant for

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 323–332, 2013.
© Springer-Verlag Berlin Heidelberg 2013

self-reliance is self-efficacy: the ability and belief to act adequately and efficiently in a given situation (Bandura, 1997), which should also be present for ICT use. Problems with using ICT to support self-reliance mainly apply to people with suboptimal cognitive abilities, such as elderly persons, people of low literacy and non-natives, but also children.

In order to make ICT accessible to a large diversity of user groups with specific abilities and values, inclusive design methods should be applied. Inclusive design is defined as the design of mainstream products and/or services that are accessible to, and usable by, as many people as reasonably possible, without the need for special adaptation or specialized design (British Standards Institute, 2005; Langdon & Thimbleby, 2010). However, developers of ICT products and services are generally not aware of the existence of inclusive design theory and methods. Also, theory and methods have been developed in research contexts and are often hard to apply in real life. In short, there is a gap between theory and practice of inclusive design. There is a need for a practical aid, that helps to create awareness of inclusive design among ICT developers, and offers easy-to-use information and tools to actually apply the methods for diverse target groups. Such a toolbox should reflect state of the art knowledge on inclusive design and should easily be connectable to already existing tools.

This paper describes the first steps taken towards an inclusive design toolbox for developing ICT applications that offer cognitive support for self-reliance. First, a brief state of the art of inclusive design standards, guidelines, design patterns and methods is provided. Then, current practices of inclusive design are presented, in the form of existing toolboxes and an inventory of the use of inclusive design methods in Dutch ICT small and medium enterprises (SMEs). Finally, a first step towards co-design of the toolbox together with SMEs is described, resulting in initial use requirements of the toolbox. The paper ends with initial conclusions and directions for further steps in the co-design process.

2 Inclusive Design Theory and Methods

2.1 Standards and Guidelines

The term 'inclusive design' stands in the tradition of the terms 'design for all', 'universal design' and '(universal) accessibility'. Up till now a number of sets of 'Universal Accessibility' guidelines have been developed for people with a variety of limitations. These guidelines are an important source of information for inclusive user interface design and evaluation. Examples of guidelines that have been issued by official bodies are the "Web Content Accessibility Guidelines" and the "User Agent Accessibility Guidelines" of the World Wide Web Consortium (W3C) and the "Guidelines for ICT products and services; ''Design for All'' of the European Telecommunications Standards Institute (ETSI). W3C aims specifically at people with visual disorders who want to use the internet (World Wide Web Consortium). ETSI has written guidelines for various disorders, but focuses more on products than on user interfaces (ETSI, 2009).

For other target groups and applications no official guidelines or standards exist. Although a lot of research has been carried out into various target groups and applications, which has often resulted in lists of design recommendations or guidelines. There are design principles for elderly people (Fisk et al., 2009), for children (Hourcade, 2008), and design 'considerations' for persons with a cognitive disability (WebAIM, Van der Pijl et al., 2005) and for people of low literacy (Cremers et al., 2012).

2.2 Human Values into Design Patterns

Value Sensitive Design (VSD) is a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process (Friedman et al., 2006). Values concern "principles or standards of a person or society, and personal or societal judgments of what is valuable and important in life" (Oxford English Dictionary) on personal, cultural or ethical issues (Cheng and Fleischmann, 2010). The values that should be addressed in inclusive design practices, such as access to information and services, can be investigated via three complementary approaches (cf. Friedman et al., 2006). First, a conceptual investigation starts with an analysis of the direct (e.g. frail elderly) and the indirect stakeholders (e.g. caregivers). Such an analysis conveys theoretical-founded values like dignity, autonomy, independence, safety, trust, and privacy. Second, an empirical investigation, encompassing different techniques like focus groups, observations, interviews and surveys, can provide additional or elaborated "situated values" like freedom from discriminatory bias. Third, a technical investigation (e.g., domotics) acquires the values that relate to technical constraints and opportunities, such as comfort and affordability.

Friedman et al. (2008) used VSD to enhance the public participation and value advocacy in a simulation-supported city design environment, aiming at mutual understandings (without manipulative or strategic actions) and freedom from bias (the absence of systematic and unfair discrimination). To achieve these value-driven goals, a design pattern was formulated that clearly demarcates a more factual presentation of information from opinion (in order to avoid misperceptions). This example shows that interaction design patterns provide a practical and sound method to establish best practices of inclusive design, incorporating the relevant human values. Interaction design patterns are structured descriptions of an invariant solution to a recurrent problem within a context (Dearden and Finlay, 2006). They are used both to record and communicate design knowledge and to support the design process. We aim at an incremental development of an inclusive design pattern library (cf. http://www.welie.com/patterns).

2.3 Situated Inclusive Design

Both the standards and guidelines (section 2.1) and the Value Sensitive Design and Interaction Design Pattern (section 2.2) approaches, have to be integrated into a coherent user-centered design rationale to establish an effective and efficient engineering process. The design rationale 'situated Cognitive Engineering' (sCE) has been developed to channel this human-centered, iterative process of deriving, refining, shaping and validating user requirements (Neerincx & Lindenberg, 2008). Values, standards and guidelines explicitly feed into the requirements, combined with the identification of specific accessibility-related user characteristics or technological preconditions (Neerincx et al., 2009; Lindenberg & Neerincx, 2001). Use case analyses drive this specification and refinement process, integrated with claims analyses that provide the justification (i.e. the expected outcome of the interaction). The use cases and user requirements with an appropriate justification are shaped into interaction design patterns. However, if appropriate design patterns are already available, these practices can be selected and re-used. The set of patterns can be implemented in a prototype for evaluation.

For inclusive design, it is essential to involve all relevant user groups in the process. Evaluation should include aspects that can be perceived objectively (performance) as well as subjective factors (affect, privacy, trust) and be executed in a realistic use context. Examples of this 'situated inclusive design process' are applications for people of low-literacy (Cremers et al., 2008), cognitively disabled (Pijl et al, 2005) and elderly (Blanson Henkemans et al., 2008; Bojic et al., 2009). Such examples show a large variety of methods and solutions with specifications on different levels of abstraction. Currently, we are developing a situated Cognitive Engineering Tool (sCET) that supports both (1) the analytical and empirical activities to acquire and assess information and (2) the recording and sharing of this information in a concise and coherent format (cf., Neerincx, 2011, and see http://www.scetool.nl).

3 Practices of Inclusive Design

3.1 Existing Design Toolboxes

A large collection of design methods is currently available, applicable in various phases of the design process, for both specification and evaluation. A number of practical on-line design toolboxes already exist that try to bridge the gap between theory and practice. For each toolbox, the aimed users, the target groups, the method description, method organization and selection, and the presentation/visualization are described.

- **Inclusive design toolkit.** The toolkit contains an introduction on inclusive design and the need for doing it. Aimed users of the toolbox are not specified; it focusses on both designers and businesses. The focus lies on descriptions of all possible target groups and their capabilities, including a model of interaction between the different user capabilities and design guidance for each capability. The toolkit contains a limited number of method descriptions: design process checklist, integrated design log, business case materials, exclusion calculator, Cambridge simulation gloves, Cambridge simulation glasses, impairment simulator software, example set of personas. The descriptions consist of definitions, case studies, and guidelines.

The information presentation employs short text sections, and lots of icons, graphics, pictures and charts (http://www.inclusivedesigntoolkit.com/).

- **55plus toolbox.** The toolbox (in Dutch) focuses on topics that change the innovation process as a result of the choice of a target group (in this case: of 55plus people). Aimed users of the toolbox are entrepreneurs, focusing on both product development and marketing and sales. There is a phasing for product design consisting of: exploration, product development, production and marketing. In each phase the user can choose from a number of guiding questions to obtain information on the target group, useful tools and cases. Suitable tools for the particular phase and target group are suggested and illustrated in factsheets containing step by step guidance, visualizations, relevant links and references (http://www.55plustoolbox.nl).
- **UCD toolbox.** This toolbox (only a private beta-version) presents some benefits of applying user-centered design. Aimed users of the toolbox are not specified. It contains an overview of 35 design methods, which can be pre-selected by criteria: type of product, design goal, resources, participants and method characteristics. Also, a pre-selection of methods can be made for various target groups: elderly, children, physically challenged, visual/hearing impaired or cognitively challenged. However, no background information on specific target groups is offered and it does not become clear why the methods are suitable for the target groups. Method descriptions contain: overview (visuals, possible outcomes, benefits, limitations, written by and reviewed by), description, tweaks (optimization), instructions (preparation, execution, analysis) and literature (http://www.ucdtoolbox.com).
- **HCD toolkit.** The toolkit shows the theory on HCD with visualizations and models. Aimed users of the toolbox are people, nonprofits, and social enterprises that work with low-income communities throughout the world (target group). The HCD Toolkit walks users through the human-centered design process and supports them in activities such as building observation and empathy skills, prototyping, leading workshops, and implementing ideas. This HCD process identifies 3 phases: Hear, Create, Deliver. Per phase a number of methods are presented. After selecting a method, you see detailed information on the method containing instructions and tips and indications of time, difficulty, materials and participants. Each method ends with one or more related stories (cases) which are submitted by users of the toolbox, creating a large involvement and experience sharing of these users (http://www.hcdconnect.org/methods).

The collection of current toolboxes already contains extensive information on target groups and (the added value of using) inclusive design methods. However, there is no toolbox yet that makes an explicit connection between (cognitive) characteristics of target groups and suitable methods. Also, current toolboxes do not contain many design guidelines yet; best practices are offered but not as design patterns. Finally, toolboxes could benefit from a better description of the commercial interest of using inclusive design for (ICT) companies.

3.2 Actual Use of Methods in the Netherlands

Interviews. A selection of Dutch SMEs who are involved in the development of products or services for end-users was made. All companies were approached by telephone to make an inventory of end-user activities they were already employing. The following questions were asked:

- What kinds of products or services do you develop?
- Who are your end-users?
- Do you involve end-users in your product or service development processes?
- Which methods do you use when involving end-users?
- What questions and needs do you have with respect to involving end-users?

Results. An inventory of a total of 56 Dutch companies was made. Of all companies, 13 (23%) were finally interviewed. The remaining 43 companies did not participate for various reasons: they did not react to requests via email or voice mail, it was hard to find the right point of contact, they turned out not to produce products or services for end-users, or they were not interested in being interviewed.

These companies produced a variety of products and services: applications of agent technology, help artists from a concept to a concrete product, virtual environments, training simulation, television apps, mobile services, health care robots, sensor technology, document management, web sites, and mobile apps. End user groups are diverse: children, people with multiple disabilities, general public, police, fire brigade, military, consumers, elderly, chronically ill, professional users.

Of the 13 companies that were interviewed, 4 indicated they never consulted end-users during their development process. One indicated they had not selected their own methods yet, but sometimes hired students industrial design to do end-user research. Another said that they sometimes used 'AB'-testing: two versions of a design, e.g. a banner, are used and the number of clicks on the banners are counted to see which one is preferred. Reasons why companies did not involve end-users is that creative persons want to pursue their own ideas, speed is very important there is no time to wait for a report, client keeps in touch with end-user, sometimes aware of adjustment of method to specific target group. Nine companies indicated they sometimes involved end-users. Various methods were mentioned: set up user evaluation in cooperation with the client, place a camera in a test setup, recruit employees from a certain application domain who bring in background knowledge, own engineers act as end-users, observations of use in context, test sessions of use in context, scenarios of future use, interviews, round table sessions, get a feeling with the market (clients) and new technology (conferences, exhibitions), review off-line (form filling), play scenarios in the lab, interview stakeholders, acquaintances of end-users, 'undercover' observation, contextual design, observation in lab, observe clients making use of products, workshops with stakeholders, customer journeys, service blue prints, concept evaluation, visit/talk to clients. However, the number of methods mentioned varied a lot between companies: larger companies and design companies employ far more methods than ICT companies.

Companies indicated some questions and needs, which a toolbox could possibly address:

- How do you know whether you have done enough research?
- How do you know you have sold the real problem with your product?
- Who do you recruit to contact the end-users?
- Who do you select as test persons if the target group is large and varied?
- How do you behave towards the end-user? What is the right attitude?
- How do you avoid politically correct answers from users?
- How can you test with end-users if the product concept is still confidential?
- What is a structured way to handle requirements?
- The need for a platform to find other companies, share information and tips & tricks.

4 Co-design of an Inclusive Design Toolbox

4.1 Toolbox Considerations

Descriptions of relevant methods, target groups and aspects of self-reliance, examples of applications of these methods and lessons learned will be collected in an inclusive design toolbox. The toolbox will be made available to SMEs who develop ICT products and services for end-users with suboptimal cognitive abilities, and who want to involve these end-users in their development processes. In order to develop a toolbox, the following aspects need to be considered:

- Who are the users of the toolbox?
    - o Expert vs. non-expert users (with respect to inclusive design)
    - o Designing for all vs. designing for special target groups
    - o Knowledge of the target group vs. unknown target group
- What are the goals of the toolbox?
  - Offering help with choosing a method in general
    - o Choice criteria (design phase, budget, time, etc.)
    - o Short description and visualization of the method
  - Offering help with executing the method in general
    - o (Extensive) description of the method and procedure
    - o Guidelines, design patterns, best practices
    - o Tips & tricks, do’s & don’ts
  - Offering help with executing methods for specific target groups
    - o Which methods are suited for specific target groups?
    - o How to adjust methods to a specific target group?
    - o Guidelines, design patterns, best practices per target group
    - o Tips & tricks, do’s & don’ts
  - Offering help with designing for specific target groups in general
    - o Descriptions of target group characteristics
    - o Personas, scenarios, storyboards of target groups

- Offering help with designing for self-reliance
  - Definition aspects of self-reliance (social security, social engagement, self-efficacy)
  - Tips & tricks, do's & don'ts

4.2 Workshop with SMEs

Method. In order to be able to develop an inclusive design toolbox that is useful for developers of products and services, requirements for functionality and design of the toolbox should be collected from the prospective users of the toolbox. In order to get input from these users, a workshop was organized with a selection of the companies, to delve deeper into requirements for the toolbox and create a Dutch inclusive design ('design for diversity') community. The goal of the workshop was to:

- Gather knowledge on special target groups such as elderly, people of low-literacy, immigrants and children
- Gather knowledge on different methods and techniques for developing and evaluating products and services for the target groups
- Inspiring best practices of peer companies of involving end users
- Exchange experiences and build up a network with companies and knowledge institutions with respect to designing for a diversity of target groups.
- Gather first requirements for the toolbox.

Participants were 14 representatives of ICT and design companies (SMEs). The workshop started with introductions on the commercial interest of designing with end-users for SMEs and the societal relevance of the inclusive design toolbox. Then, three presentations of best practices of inclusive design methods (context mapping, focus group, questionnaire) with, respectively, children, low-literate and elderly persons were given by three companies. In interactive sessions, three groups discussed how the best practices could be applied in their own companies and how a toolbox could help. Finally, a 'design for diversity' social media group was launched.

Results. Some requirements for the toolbox came up during the workshop:

- Availability of personas, to create an image of standard end-users
- Examples (best practices) of products and processes, success/failure factors
- Preconditions for the use of methods and solutions
- SME-proof: simple, not scientific (but rooted in scientific knowledge)
- Available methods within a certain time period and budget
- First present solutions (best practices), then method on how to get there
- Possibility to shop and snitch from toolbox
- Possibility to share information (open source)

5 Discussion, Conclusions and Future Work

This study confirms our assumption that there is a substantial gap between theory and practice of inclusive design.

From a theoretical point of view, there is still a lot of work to be done in the selection of appropriate inclusive design methods, standards, guidelines and design patterns for diverse target groups, as well as the validation, description and disclosure of this knowledge base. Moreover, inclusive design theory and methods should form input to the situated cognitive engineering process, to become 'situated inclusive design'.

From a practical perspective, it can be concluded that awareness of inclusive design is still lacking. It has proven to be hard to find Dutch ICT companies that produce end-user products and are willing or able to discuss methods for end-user involvement. Unfamiliarity with inclusive design seems to be a barrier for participation in interviews or a workshop. However, we found that some companies already involve end-users in their development processes, in less or more structured ways. Also, we have been able to derive some initial needs and requirements for an inclusive design toolbox, in particular the need for concrete personas (target groups) and design patterns (best practices). Finally, the creation of a social media group 'design for diversity' should ensure more awareness and involvement of companies.

In the future, next steps in the co-design process of the inclusive design toolbox should be taken, with regular involvement of ICT companies. The inclusive design research community has a strong responsibility to guarantee the quality of this toolbox. Ultimately, the toolbox should lead to ICT solutions that match cognitive abilities and reflect values of diverse target groups, in order to empower these citizens to become self-reliant in society.

Acknowledgements. The current work was carried out as part of COMMIT, a Dutch national public-private research program on information and communication technology, project P02: Interaction for universal access (http://www.commit-nl.nl). We thank Judith Kessens for reviewing the paper.

References

1. Bandura, A.: Self-efficacy: the exercise of control. Freeman and Company, New York (1997)
2. Blanson Henkemans, O.A., Rogers, W.A., Fisk, A.D., Neerincx, M.A., Lindenberg, J., van der Mast, C.: Usability of an adaptive computer assistant that improves self-care tasks and health literacy of older adults. Methods of Information Medicine 47(1), 82–88 (2008)
3. Bojic, M., Blanson Henkemans, O.A., Neerincx, M.A., Van der Mast, C.A.P.G., Lindenberg, J.: Effects of Multimodal Feedback on the Usability of Mobile Diet Diary for Older Adults. In: Stephanidis, C. (ed.) UAHCI 2009, Part III. LNCS, vol. 5616, pp. 293–302. Springer, Heidelberg (2009)
4. British Standards Institute. British Standard 7000-6:2005. Design management systems - Managing inclusive design – Guide (2005)

5. Cheng, A.-S., Fleischmann, K.R.: Developing a meta-inventory of human values. Proc. Am. Soc. Info. Sci. Tech. 47, 1–10 (2010)
6. Cremers, A.H.M., Kranenborg, K., Kessens, J.M.: Guidelines for user interfaces for illiterates. Draft TNO Report (2012)
7. Cremers, A.H.M., de Jong, J.G.M., van Balken, J.S.: User-centered design with illiterate persons: The case of the ATM user interface. In: Miesenberger, K., Klaus, J., Zagler, W.L., Karshmer, A.I. (eds.) ICCHP 2008. LNCS, vol. 5105, pp. 713–720. Springer, Heidelberg (2008)
8. Dearden, A., Finlay, J.: Pattern Languages in HCI: A Critical Review. Human–Computer Interaction 21(1), 49–102 (2006)
9. ETSI. Human Factors (HF); Guidelines for ICT products and services; "Design for All". Sophia Antipolis Cedex, ETSI Guide ETSI EG 202 116 V 1.2.2 ETSI (2009-3) (2009)
10. Fisk, A.D., Rogers, W.A., Charness, N., Czaja, S.J., Sharit, J.: Designing for older adults: principles and creative human factors approaches, 2nd edn. Human Factors & Aging Series. CRC Press Taylor & Francis Group, Boca Raton (2009)
11. Friedman, B., Kahn Jr., P.H., Borning, A.: Value Sensitive Design and information systems. In: Zhang, P., Galletta, D. (eds.) Human-Computer Interaction in Management Information Systems: Foundations, pp. 348–372. M.E. Sharpe, Armonk (2006)
12. Friedman, B., Borning, A., Davis, J.L., Gill, B.T., Kahn Jr., P., Kriplean, T., Lin, P.: Laying the foundations for public participation and value advocacy: Interaction design for a large scale urban simulation. In: Proc. 9th Annual Int. Conference on Digital Government Research. Digital Government Society of North America, Montreal (2008)
13. Langdon, P., Thimbleby, H.: Editorial: Inclusion and interaction: designing interaction for inclusive populations. Interacting with Computers 22, 439–448 (2010)
14. Lindenberg, J., Neerincx, M.A.: The need for a "universal accessibility" tool. ACM SIGCAPH Computers and the Physically Handicapped 69, 14–17 (2001)
15. Neerincx, M.A.: Situated Cognitive Engineering for Crew Support in Space. Personal and Ubiquitous Computing 15(5), 445–456 (2011)
16. Neerincx, M.A., Lindenberg, J.: Situated cognitive engineering for complex task environments. In: Schraagen, J.M.C., Militello, L., Ormerod, T., Lipshitz, R. (eds.) Naturalistic Decision Making and Macrocognition, pp. 373–390. Ashgate, Aldershot (2008)
17. van der Pijl, D.J., Cremers, A.H.M., Soede, M.: Personalized PDA accessibility for intellectually disabled persons: concept guidelines based on the development of an electronic travel companion. In: HCI International, Las Vegas, July 22-27 (2005)
18. WebAIM. Cognitive disabilities Part 1 (We Still Know Too Little, and We Do Even Less) and Part 2 (Conceptualizing Design Considerations). Web Accessibility In Mind, `http://webaim.org/articles/cognitive`
19. World Wide Web Consortium. Web Content Accessibility Guidelines 1.0; User Agent Accessibility Guidelines 1.0., `http://www.w3.org`

Online Single EEG Channel Based Automatic Sleep Staging

Gary Garcia-Molina[1,2], Michele Bellesi[2], Sander Pastoor[3], Stefan Pfundtner[3], Brady Riedner[2], and Giulio Tononi[2]

[1] Philips Research North-America, Briarcliff, NY, USA
[2] Department of Psychiatry, University of Wisconsin, Madison, WI, USA
[3] Philips Research Europe, Eindhoven, The Netherlands
gary.garcia@philips.com, bellesi@wisc.edu

Abstract. Recent evidence supports the positive effects of external intervention during specific sleep stages (e.g. enhanced memory consolidation and depression relief). To enable timely intervention, online automated sleep staging is required and preferably with short latency. In this paper, we propose an approach to achieve this based on the analysis of spectral features of a single electroencephalogram (EEG) channel and the use of Gaussian Mixture Models. We compare among several choices for the EEG signal location, the type of spectral features, and the duration of the signal segment (epoch) that is required to automatically identify the sleep stage. The performance metric used for comparison purposes is the kappa statistic, which measures the agreement between the automatic and manual sleep staging. The performance is higher when central EEG locations (C3, C4), longer epochs, and the power in five frequency bands are used. However, good results (kappa=0.6) can also be obtained for an epoch duration of 12 seconds.

Keywords: automatic sleep staging, online, single EEG signal, spectral features, GMM.

1 Introduction

Sleep is a state of reversible disconnection from the environment characterized by quiescence and reduced vigilance. Although the precise function of sleep remains to be elucidated, it appears that sleep primarily benefits the brain. For example according to the synaptic homeostasis hypothesis (SHY) [1], plastic processes occurring during wakefulness result in a net increase in synaptic strength in many cortical circuits, and sleep is needed to renormalize synaptic strength in a way that is beneficial to nerve cells and to memory. SHY also emphasizes that when cortical neurons begin oscillating at low frequencies during non-rapid eye movement sleep, these oscillations become synchronized. This is why the electroencephalogram (EEG) exhibits high power in the delta band (0.5-4 Hz) particularly at the beginning of a sleep episode. SHY also claims that slow waves in sleep do not merely reflect synaptic strength, but

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 333–342, 2013.
© Springer-Verlag Berlin Heidelberg 2013

also play a functional role in "renormalizing" synaptic strength to a baseline level that is energetically sustainable and beneficial for performance [1].

Recent research evidence indicates that modulating sleep activity patterns, specifically slow waves via sensory, magnetic, or electric stimuli at specific sleep stages can be beneficial in a wide range of contexts including memory acquisition and consolidation [2][3] and relief from depression [4][5]. To verify the validity of such interventions in practice requires conducting research in a larger population using automated means for online sleep staging with low latency to allow timely intervention.

Conventional sleep staging relies on various bio-signals (polysomnography) for human experts to decide on sleep stages typically on the basis of 30-second long segments. Real-time sleep staging is proposed in [6] using EEG, electro-oculogram (EOG), and electro-myogram signals (EMG). We consider here the option of achieving online automatic sleep staging on the basis of a single channel (or signal). EOG and EEG electrodes are considered for this purpose. Using a single signal permits to simplify the research setup and increases the subject comfort.

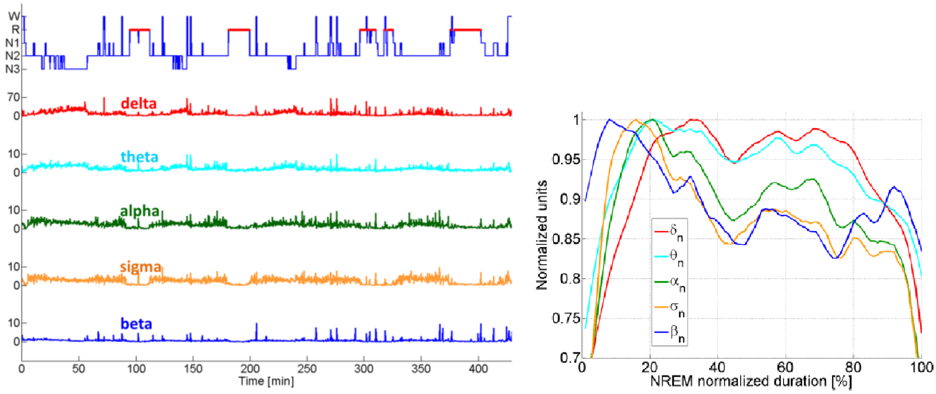
This paper is organized as follows. Section 2 presents an overview of the time course of the EEG signal during sleep. Section 3 describes the methods used for automatic sleep staging. The results in terms of performance are presented in Section 4. Section 5 concludes the paper.

2 EEG during Sleep

Two distinct types of sleep occur in humans: rapid eye movement (REM) sleep, and non–REM sleep. Compared to the low voltage, high frequency patterns appearing in the awake EEG, non–REM (NREM) sleep is associated with a synchronized EEG pattern. NREM is subdivided into stages N1, N2, and N3. During REM, the EEG exhibits a pattern similar to that observed during wakefulness [7].

Global trends for the EEG power during sleep in the classical frequency bands: delta (0.5–4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (11-15 Hz), and beta (15-30 Hz) are analyzed in detail in [8]. Fig. 1a summarizes the typical time courses of the power in those bands depending on the sequence of sleep stages as indicated by the hypnogram (top graph in Fig. 1a). The power in Fig. 1a is reported in RMS units which are obtained by calculating the square root from the average energy (in a window of a given duration) of the signal band-pas filtered in the frequency of interest. As sleep deepens the power in the delta and theta bands increase whereas the power in the alpha, sigma, and beta bands follow a quasi-opposite trend.

A more detailed view of the power trends during non-REM (NREM) can be observed in Fig. 1b where the average power in all frequency bands over 10 subjects is portrayed. The time was normalized to percent units to account for the difference in NREM duration for different subjects. The power was also normalized (by scaling w.r.t. the highest value) to facilitate the comparison. The power in delta and theta increase at the beginning of the NREM episode, remains high for most of the NREM duration, and decreases at the end of NREM (i.e. following an inverted-U trend). The power in alpha, sigma, and beta rapidly increases at the beginning of the NREM



(a) (b)

Fig. 1. a) Top: hypnogram of a representative night (red segments correspond to REM sleep). Time courses of the power, in RMS units, in the delta, theta, alpha, sigma, and beta frequency bands. b) (Average) normalized time course of the normalized power in delta, theta, alpha, sigma, and beta throughout a NREM episode.

episode, reaches its maximum, and decreases afterwards. The power in beta slightly increases by the end of the NREM episode. These observations are used to guide the spectral feature extraction as described in Section 3.2.

3 Methods

3.1 Dataset

The data from ten healthy, medication-free, right-handed subjects (five female; age 21.9 ± 0.5 years; mean ± std.) who participated in a previous study run at the University of Wisconsin [9] was used in this paper. The data includes right/left electro-oculogram signals (EOGR and EOGL), six bipolar EEG signals (F3-A2, F4-A1, C3-A2, C4-A1, O1-A2, and O1-A2), other polysomnography signals (details in [9]), and the manually annotated hypnogram done by an expert according to the standard rules in [7] and on the basis of 30-second long segments. To permit the analysis of segment durations as described in section 3.2, sleep stages were assigned to each 6-second long segment by simple extrapolation from the 30-second long based staging. Five sleep stages were considered: 1) Wake (W), 2) REM (R), 3) N1, 4) N2, and 5) N3.

3.2 Automated Sleep Staging

The procedure used to automatically score a hypnogram from the data of a single signal, is designed in such a way so as to accommodate online operation. A signal segment (*epoch*) of a given duration is presented to a *classifier* which decides the sleep stage to which this epoch belongs independently from previous decisions. In this paper, we consider epochs with durations: 6, 12, 18, 24, and 30 seconds. The classifier is *trained* with ground truth data, i.e. epochs for which the sleep stage is known (annotated by the human expert).

Spectral Features per Epoch

Spectral features from the epochs are first extracted by calculating the log of the estimated power (in RMS units) in the delta, theta, alpha, sigma, and beta frequency bands (δ, θ, α, σ, and β are used to refer to the log of the power in RMS units). The logarithm is extracted to equalize the feature range across frequency bands (especially relevant for delta because the power in this band is higher than in other bands, see Fig. 1a) as this prevents the estimation of singular covariance matrices [10]. A combination of spectral features (e.g. δ and β) defines a *feature vector* associated with each epoch. For a sleep night recording, the number of feature vectors corresponds to the number of epochs in the recording.

The spectral combinations considered are:

1) Two-band spectral combinations between i) the bands where the power follows an inverted-U trend during NREM (i.e. δ and θ according to Fig. 1b) and ii) the bands where the power trend quasi-mirrors that of δ and θ during NREM (i.e. α, σ, and β according to Fig. 1b). This results in 6 two-band spectral combinations where the feature vectors have two components.
2) All possible four-band spectral combinations. This amounts to five combinations: δθασ, δθαβ, δθσβ, δασβ, and θασβ where the feature vectors have four components.
3) The combination of all five bands which results in feature vectors with five components.

Classifier

The classifier decides on the sleep stage for a given epoch on the basis of that epoch's feature vector (referred to as x, d is the feature vector dimension). The classification approach in this paper is based on the estimation of the probability that a given feature vector belongs to each sleep stage, referred to as $p(x|stage)$, and select the sleep stage (s_x) for which the likelihood is the largest (see Eq. 1).

$$s_x = \underset{stage=\{W,R,N1,N2,N3\}}{\arg\max} \left\{ p(x|stage) \right\} \tag{1}$$

The probability density function $p(x|stage)$ is estimated using a Gaussian Mixture Model (GMM) [11] and can be written as:

$$p(x|stage) = \sum_{m=1}^{M} a_m p\left(x|stage,m\right), \tag{2}$$

where a_m is the m-th mixture coefficient, $\sum_{m=1}^{M} a_m = 1$, and $p(x|stage, m)$ is the m-th multimodal Gaussian mixture which can be written as in Eq. 3 (the "stage" sub-indices were removed in this equation for convenience of notation).

$$p(x|m) = \frac{1}{(2\pi)^{d/2}|\Sigma_m|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_m)^T \Sigma_m^{-1} (x-\mu_m)\right\}, \tag{3}$$

where Σ_m is the $d \times d$ covariance matrix, and μ_m is the mean vector having d elements. The mixture model can approximate any probability density function arbitrarily closely provided that it contains enough components[12].

The GMM parameters are estimated using data from a *training set*. Given the order of the model *M*, the mixture coefficients, covariance matrices, and the mean vectors are estimated using the expectation maximization algorithm. This algorithm is described in detail in[10]. The order of the model is chosen in such a way so as to balance the model complexity (the higher *M* is, the higher the complexity is) and the fit with the data. For the data considered in this paper, we use *M=2*.

3.3 Performance Evaluation

Evaluation of the automated staging performance was made following a leave-one-subject-out procedure where the data from all the subjects but one in the dataset is used to estimate the GMMs for each sleep stage. The data from the remaining subject (e.g. subject *k*) is used to test the models and estimate a 5×5 *confusion matrix* C_k where the *i*, *j* element represents the number of times an epoch manually scored as belonging to sleep stage *i,* was classified as belonging to sleep stage *j*. The correspondence between numeric indices and sleep stages is as follows: 1) W, 2) R, 3) N1, 4) N2, and 5) N3. A visualization of the performance evaluation is illustrated in Fig. 2, where *N* is the number of subjects (10 in this paper), and $F_{k,1}, F_{k,2}$ are the feature vectors of the epochs of subject *k* corresponding to the first and second nights respectively. The training data consists of the epochs for both nights of all other (*N*-1) subjects in the dataset (i.e. $k_1, \ldots, k_{N-1} \neq k$). The training data is organized per sleep stage and the corresponding GMMs are estimated.

Although the confusion matrix provides sufficient information to assess the automatic staging performance, comparing the performance for different feature selections and electrode locations is facilitated by extracting a single quantity that reflects the information contained in the confusion matrix. This can be accomplished by estimating the so-called *kappa statistic* [13] from the global confusion matrix $C = \sum_k C_k$.

Kappa measures the agreement between the manual and automatic sleep staging and is defined as: $\kappa = (P_a - P_e)/(1 - P_e)$, where P_a is the relative observed agreement, and P_e is the hypothetical probability of chance agreement. P_a and P_e can be obtained from the confusion matrix as: $P_a = \sum_{i=1}^{5} C_{i,i} / \sum_{i=1}^{5}\sum_{j=1}^{5} C_{i,j}$ and $P_e = \sum_{i=1}^{5} S_{:,i} S_{i,:} / \left(\sum_{i=1}^{5}\sum_{j=1}^{5} C_{i,j} \right)^2$, where $S_{:,i}$ and $S_{:,i}$ are respectively the sums across columns and rows of the confusion matrix C.

As discussed in [14], kappa values greater than 0.80 represent almost perfect, between 0.61 and 0.80 substantial, between 0.41 and 0.60 moderate, between 0.21 and 0.40 fair and between 0 and 0.21 slight agreement. The average kappa value characterizing the agreement between two human experts is 0.87 [15]. For illustration, the automatically generated hypnograms when κ=0.6 and κ=0.4 are represented in Fig. 5.

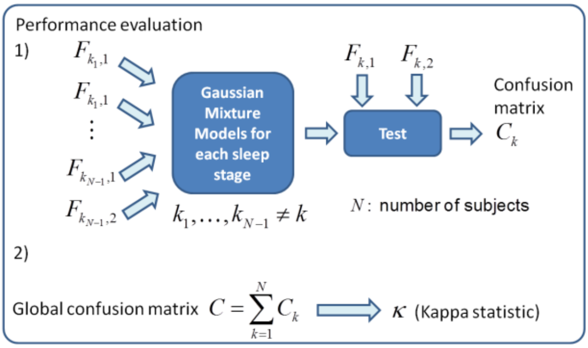


Fig. 2. Visualization of the performance evaluation method. A leave-one-subject-out procedure is used and the corresponding κ is estimated.

4 Results and Discussion

4.1 Two-Band Spectral Combinations

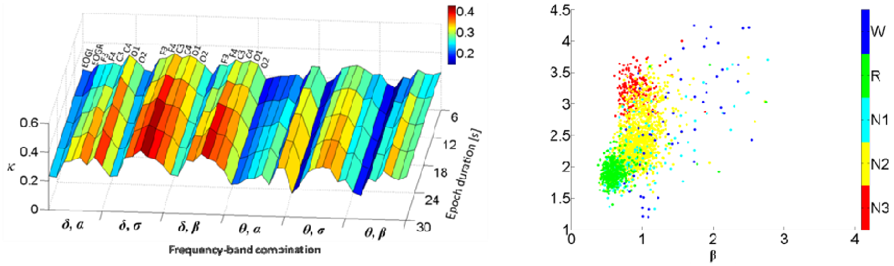
The results in terms of kappa for the six two-band spectral combinations (δα, δσ, δβ, θα, θσ, θβ) are represented in Fig. 3a. The electrode location and the epoch duration are also considered as independent variables.

As expected, kappa increases when the epoch duration increases. As for the electrode locations, the highest kappas are reached on central locations (C3 and C4), the lowest kappas are for EOG electrodes, and intermediate kappa values are for frontal and occipital locations. The maximum kappa is 0.42 (moderate agreement) for location C3, epoch duration 24 seconds, and the combination δσ.

If θ is used instead of δ, kappa decreases which suggests that δ plays a key role in defining the sleep stages. An illustration of the mapping of sleep stages in the δ-β plane is depicted in Fig. 3b.

4.2 Four-Band and Five-Band Spectral Combinations

The kappa results for all possible four-band spectral combinations and the five-band spectral combination are represented in Fig. 4. The electrode location and the epoch



(a) (b)

Fig. 3. a) Values of κ for: the 2-band spectral combinations for all considered electrode locations and epoch durations. b) Mapping of sleep stages in the δ-β plan for subject 1.

duration are also considered as independent variables. As in the case of 2-band spectral combinations, the value of kappa increases with the epoch duration. Kappa is larger for frontal and central electrode locations, low for EOG signals, and intermediate for occipital locations.

Using all the five bands improves kappa as compared to the four-band combinations. The maximum kappa is 0.63 (substantial agreement with the manual sleep staging) for location C4, epoch duration 24 seconds, and using the five frequency bands. When using only four frequency bands, the maximum kappa is 0.60 (moderate agreement) for location C4, epoch duration 24 seconds, and the combination δθαβ. For the four-band combinations, the one that is associated with lower kappas is: θαβσ. This confirms the results in Section 4.1 where δ was found to be essential for the automatic staging.

Table 1. Kappa for short epoch durations, frontal & central locations, and using δθασβ

	Location			
Epoch Duration	**F3**	**F4**	**C3**	**C4**
6 seconds	0.53	0.55	0.54	0.56
12 seconds	0.57	0.60	0.58	0.60

For the sake of low latency online sleep staging, it is useful to consider the performance for shorter epoch durations (i.e. 6 seconds and 12 seconds). This is reported in Table 1 where it can be observed that a moderate agreement with manual sleep scoring can be achieved for 12-second long epochs. An illustration of the automatically scored hypnogram using 12-second long epochs is shown in Fig. 5.

On the basis of 30-second long epochs using a single EEG signal, with a smoothing step (deciding on the stage by considering the results on adjacent epochs), [16] reports a kappa of 0.72 for 5 sleep stages. The smoothing step is of course not suitable for online operation. Using 30-second long epochs a kappa value of 0.79 is reported in [15] using a single EEG signal. The advantage of our approach relies on the fact that shorter epochs can be used, maintaining a moderate level of agreement, which ensures low latency, favoring real-time operation.

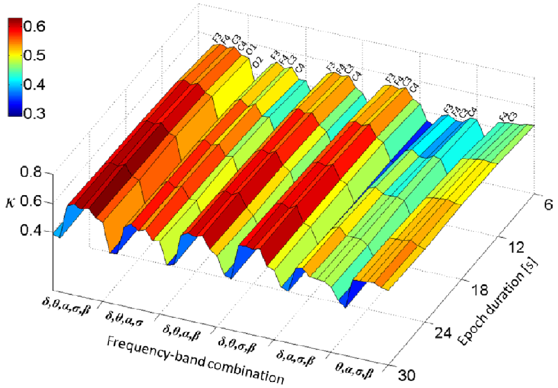
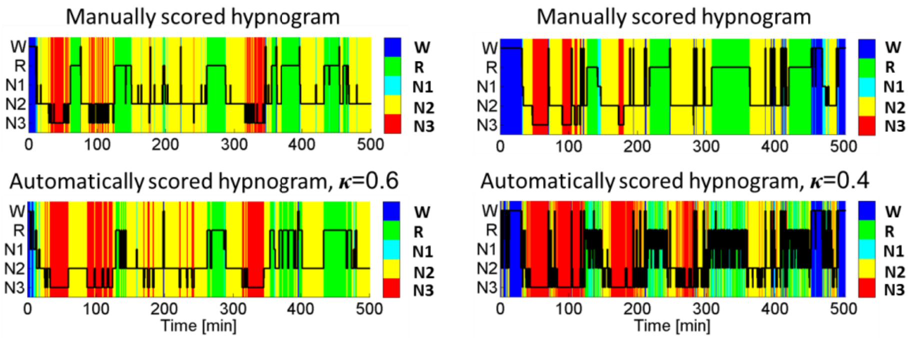


Fig. 4. Values of κ for: all 4-band spectral combinations and the 5-band spectral combination versus all considered epoch durations



(a) (b)

Fig. 5. Examples of automatically generated hypnograms; κ=0.6 (left) and κ=0.4 (right). (a) Top: manually scored hypnogram (reference). Bottom: automatically scored hypnogram (κ=0.6) using spectral features δ, θ, α, σ, and β extracted from the signal at electrode C4 and on the basis of 12-second long epochs. (b) Top: manually scored hypnogram. Bottom: automatically scored hypnogram (κ=0.4) using spectral features δ, and β extracted from the signal at electrode C3 and on the basis of 12-second long epochs.

5 Conclusions

In this paper, we propose an approach to automatic sleep staging on the basis of a single channel using spectral features and estimating the probability density function of feature vectors per sleep stage through GMMs. The performance was evaluated using the kappa statistic which can be calculated from the confusion matrix.

The performance was compared across different selections for the spectral features, epoch durations, and electrode location. Higher performance can be reached if central electrodes (C3, or C4), longer epochs, and all frequency bands are used. If only two frequency bands were to be used (e.g. delta and sigma), the performance is moderate (kappa=0.4).

Interestingly, shorter epochs (6 and 12 seconds) can be used at an acceptable level of performance, kappa=0.56 (6 seconds, C4) and kappa=0.6 (12 seconds, C4). The sleep stage can then be decided for each 12-second long epoch independently from the stage of the previous epochs. This permits almost real-time operation and timely intervention during sleep.

Since the decision on the sleep stage to which an epoch belongs is taken independently from previous epochs, an obvious performance improvement step would consist in smoothing the results by considering the sleep stages from previous epochs. This step however, could increase the latency which may adversely impact online operation.

The use of standard manually scored 30-second long epochs as reference for the staging of shorter epochs may not be the best choice because this is a standard for offline analyses that was established in the clinical framework.

References

1. Tononi, G., Cirelli, C.: Sleep function and synaptic homeostasis. Sleep Medicine Reviews 10(1), 49–62 (2006)
2. Marshall, L., Helgadóttir, H., Mölle, M., Born, J.: Boosting slow oscillations during sleep potentiates memory. Nature 444(7119), 610–613 (2006)
3. Marshall, L., Mölle, M., Hallschmid, M., Born, J.: Transcranial direct current stimulation during sleep improves declarative memory. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 24(44), 9985–9992 (2004)
4. Landsness, E.C., Goldstein, M.R., Peterson, M.J., Tononi, G., Benca, R.M.: Antidepressant effects of selective slow wave sleep deprivation in major depression: a high-density EEG investigation. Journal of Psychiatric Research 45(8), 1019–1026 (2011)
5. Vogel, G.W., Vogel, F., McAbee, R.S., Thurmond, A.J.: Improvement of depression by REM sleep deprivation. New findings and a theory. Archives of General Psychiatry 37(3), 247–253 (1980)
6. Kuwahara, H., Higashi, H., Mizuki, Y., Matsunari, S., Tanaka, M., Inanaga, K.: Automatic real-time analysis of human sleep stages by an interval histogram method. Electroencephalography and Clinical Neurophysiology 70(3), 220–229 (1988)
7. Iber, C., Ancoli-Israel, S., Chesson, A.L., Quan, S.F.: The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications, First. American Academy of Sleep Medicine (2007)
8. Merica, H., Fortune, R.D.: State transitions between wake and sleep, and within the ultradian cycle, with focus on the link to neuronal activity. Sleep Medicine Reviews 8(6), 473–485 (2004)
9. Hulse, B.K., Landsness, E.C., Sarasso, S., Ferrarelli, F., Guokas, J.J., Wanger, T., Tononi, G.: A postsleep decline in auditory evoked potential amplitude reflects sleep homeostasis. Clinical Neurophysiology 122(8), 1549–1555 (2011)
10. Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press (1995)
11. Reynolds, D.A., Rose, R.C.: Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing 3(1), 72–83 (1995)

12. McLachlan, G.J., Basford, K.E.: Mixture models: Inference and applications to clustering. CRC Press (1987)
13. Cohen, J.: A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 20(1), 37–46 (1960)
14. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics 33(1), 159–174 (1977)
15. Jensen, P.S., Sorensen, H.B.D., Leonthin, H.L., Jennum, P.: Automatic sleep scoring in normals and in individuals with neurodegenerative disorders according to new international sleep scoring criteria. Journal of Clinical Neurophysiology 27(4), 296–302 (2010)
16. Berthomier, C., Drouot, X., Herman-Stoïca, M., Berthomier, P., Prado, J., Bokar-Thire, D., Benoit, O., Mattout, J., D'Ortho, M.-P.: Automatic analysis of single-channel sleep EEG: validation in healthy individuals. Sleep 30(11), 1587–1595 (2007)

CogWatch – Automated Assistance and Rehabilitation of Stroke-Induced Action Disorders in the Home Environment

Joachim Hermsdörfer[1], Marta Bienkiewicz[1], José M. Cogollor[2], Martin Russel[3], Emilie Jean-Baptiste[3], Manish Parekh[3], Alan M. Wing[4], Manuel Ferre[2], and Charmayne Hughes[1]

[1] Institute of Human Movement Science, Department of Sport and Health Science, Technische Universität München, 80992 Munich, Germany
{joachim.hermsdoerfer,marta.bienkiewicz,charmayne.hughes}@tum.de
[2] Centre for Automation and Robotics CAR (UPM-CSIC), Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain
{m.ferre,jm.cogollor}@upm.es
[3] Electrical, Electronic and Computing Engineering, The University of Birmingham, Birmingham, B15 2TT, U.K.
{M.J.RUSSELL,EMJ198,MXP534}@bham.ac.uk
[4] School of Psychology, The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK.
a.m.wing@bham.ac.uk

Abstract. Stroke frequently causes apraxia, particularly if it affects the left-hemisphere. A major symptom of apraxia is the presence of deficits during the execution and organization of activities of daily living (ADL). These deficits may substantially limit the capacity of stroke patients to live independently in their home environment. Traditional rehabilitative techniques to improve ADL function revolve around physical and occupational therapy. This approach is labor intensive and constraints therapy to clinical environments. The CogWatch system provides an supplementary means of rehabilitation that is based on instrumented objects and ambient devices that are part of patients' everyday environment and can be used to monitor behavior and progress as well as re-train them to carry out ADL through persistent multimodal feedback.

Keywords: Apraxia, activities of daily living, rehabilitation, stroke, assistive technology.

1 Introduction

Following a cerebrovascular accident or stroke, a significant proportion of patients can suffer from apraxia, particularly if it affects the left-hemisphere. Major symptoms of apraxia are deficits of tool use and errors in executing and organizing action sequences of activities of daily living (ADL) [Goldenberg and Hagmann1998b;Schwartz and Buxbaum1996;Foundas et al.1995]. These deficits may substantially limit the

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 343–350, 2013.
© Springer-Verlag Berlin Heidelberg 2013

capacity of stroke patients to live independently in their home environment. Effective rehabilitation is therefore a high priority, however, traditional methods are labor intensive and opportunities for patients to practice recovered skills are very limited [Smania et al. 2000;Goldenberg and Hagmann 1998a;Bowen et al. 2009].

Automated computer-based systems are therefore seen as a viable alternative. To date, the majority of computer-based rehabilitation systems is focused on treating physiological aspects of stroke, such as limb movement, and is based on robot and/or virtual environment platforms. Moreover, they require the patient to function within the system workspace rather than being adapted to the patient's normal environment. As an alternative, partners in the CogWatch Project are developing a Personal Healthcare System (PHS) that provides customized, continuous and long-term cognitive rehabilitation in the patient's familiar environment (e.g., kitchen), and allows remote monitoring of the progress of the patient by relevant clinicians.

In this study we detail the core components of the CogWatch system. The development involved the following steps: The ADL tasks relevant for patients and also suitable for automated action recognition were determined. Subsequently, clinical studies with apraxic patients were conducted in order to define typical error patterns during ADL performance. Sensors and algorithms that could provide sufficient information to recognize actions were then developed. Lastly, feedback and cueing procedures were designed to alert patients and guide error correction during task performance. Fig. 1 provides an overview of the components of the system.

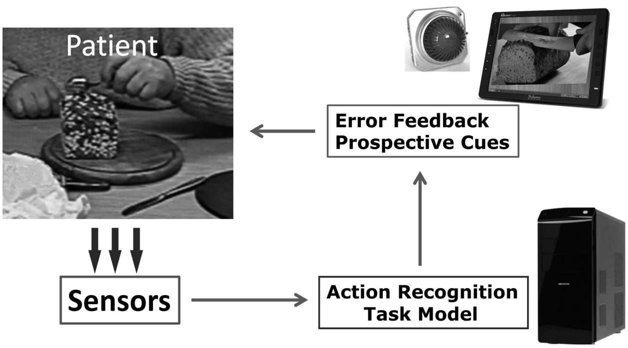


Fig. 1. Simplified schema of an automated system (CogWatch) for rehabilitation of ADL impairments following stroke

2 ADL Task Selection

In the first stages of CogWatch development ADL's with a clear relevance for home scenarios (from activities such as drink and meal preparation and consumption, dressing, and personal grooming) were considered for inclusion in the development. After careful consideration, the selected ADL scenario for the first prototype was a tea-making task. This task was chosen because it is highly relevant to everyday life, should be familiar to the majority of participants, is sufficiently complex to ensure the inclusion of a substantial number of apraxic patients, and also enables analysis about the selectivity of the effects of apraxia. Tea making has also been thoroughly studied in the literature, and thus provides a basis against which the patient results could be compared.

3 Clinical Studies to Determine Typical Error Patterns

We conducted experiments in a clinical setting to ascertain typical error patterns in a tea making task. To this end, we asked twenty-one apraxia patients (age = 58.10 y, SD = 13.81, 12 men, 9 women) with lesions following a single cerebrovascular accident (CVA) to prepare two cups of tea with different ingredients ("2 cup tea-making task").

Analysis indicated that patients committed errors in 48.6% of trials, with a total of 68 errors recorded. The number of errors per trial ranged from 0 - 10 (mean = 3.14, SD = 2.46). Consistent with previous research [Buxbaum 1998;Schwartz et al. 1998] the most frequently occurring error was that of ingredient omission (35% of errors), with patients typically failing to put tea bags into one or both cups, or adding one or both sweetener tablets to cup2. There were also a number of trials in which apraxia patients omitted a step in the action sequence (sequence omission = 18%), for example, turning on the empty kettle. Patients also added an unnecessary ingredient (ingredient addition = 9%), or substituted an unnecessary ingredient for a necessary one (ingredient substitution = 9%), carried out an action in an inappropriate way (quality = 7%), and performed an action earlier than required (sequence anticipation = 6%). Lastly, there was a small number of toying (4%), misestimation (3%), mislocation (3%), perseveration (3%), ingredient retraction (2%), and perplexity errors (2%). The data obtained from the clinical studies was used to develop the Task Models (TMs) for the first prototype (see section 5).

4 Monitoring and Feedback Devices

The CogWatch system is designed to be personalized to suit the needs of individual patients at the same time as being practical and affordable for home installation so that rehabilitation takes place in familiar environments performing familiar tasks. As such, it is imperative that the home-based action recognition system is affordable and easy to implement. Fortunately, emerging technologies have led to the rapid development of low-cost and easy to use monitoring and feedback devices. These devices are capable of monitoring patients' goal-oriented behavior and delivering multimodal action-guidance, and feedback to correct errors that have occurred and risk alerts to prevent accidents.

4.1 Monitoring Devices

Microsoft Kinect™ Sensor. The current version of the CogWatch system utilizes Kinect™ technology to capture and record 3D hand positions and color images from the scene. The Kinect™ sensor is connected to a computer that interprets all signals and also allows the clinician or patient to interact with the software via a user-friendly interface (Fig. 2). This interface provides real-time visualization of the patient and image scene, as well as the signals from the sensorized objects. Furthermore, the acquired raw signals will be saved in a repository/database, and can be used for offline evaluation of patient behavior by relevant clinical personal.

The usability of the Kinect™ sensor in the CogWatch system was recently evaluated in both healthy and apraxic populations [Cogollor et al. 2012]. In that study, hand movements were recorded during a tea making task by a Kinect sensor™ and compared to data obtained by a commercially available marker-based motion capture system (CMS Zebris, Isny, Germany). The results indicated a moderate to strong correlation between signals (regardless of the neurological status of the individual), suggesting that the Kinect™ device is a reliable motion capture system in a cognitive rehabilitative context.

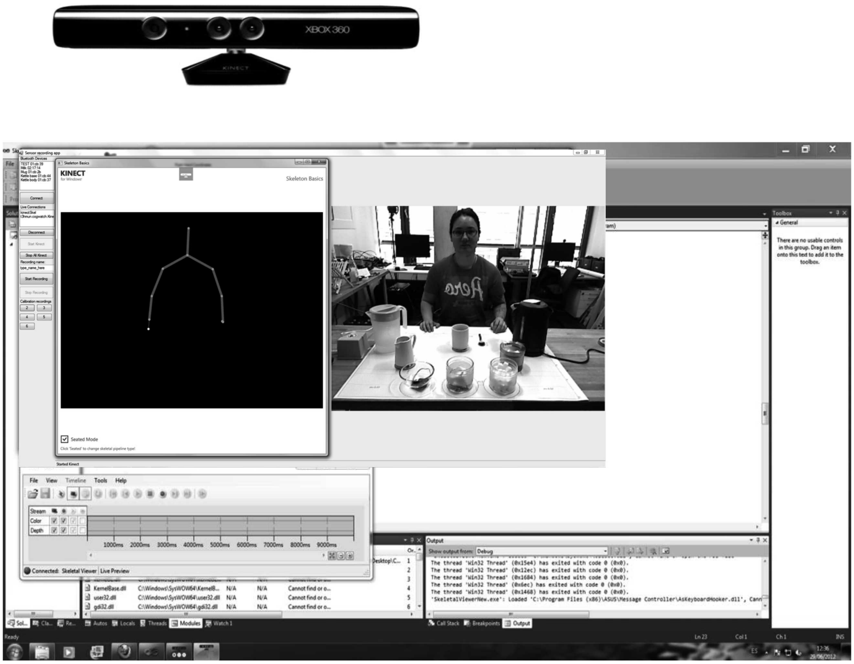


Fig. 2. Kinect™ user interface with skeleton view of arms and hands

Instrumented Objects. To detect actions at the sub-goal level the CogWatch system uses direct instrumentation of some of the objects used. A device called the CogWatch instrumented coaster (CIC) has been created to hold a 3-axis accelerometer and 3 force sensitive resistors (FSRs) in a small package that fits to the base of objects such as mugs, jugs and kettles. It is designed to be small and unobtrusive so that it does not affect the usage of the objects and has very little effect on their appearance. The CIC streams the sensor data to a host computer via Bluetooth, sampling each sensor at a rate of 200Hz with a 12bit analog to digital converter. As with the Kinect™, 3-D accelerometer values and the outputs of three FSRs are saved in the CogWatch system repository/database.

The accelerometer provides data that can be used to monitor changes in movement and orientation of the object that the CIC is attached to. The FSRs are used to monitor changes in weight of the object. Initial data collection has shown that these sensors are able to create very strong characterizations of actions such as pouring a kettle of water into a mug (see Fig. 3). The FSR sensors further allow the system to detect if the water is actually being poured into a mug (an example error is failure to pour into the correct container).

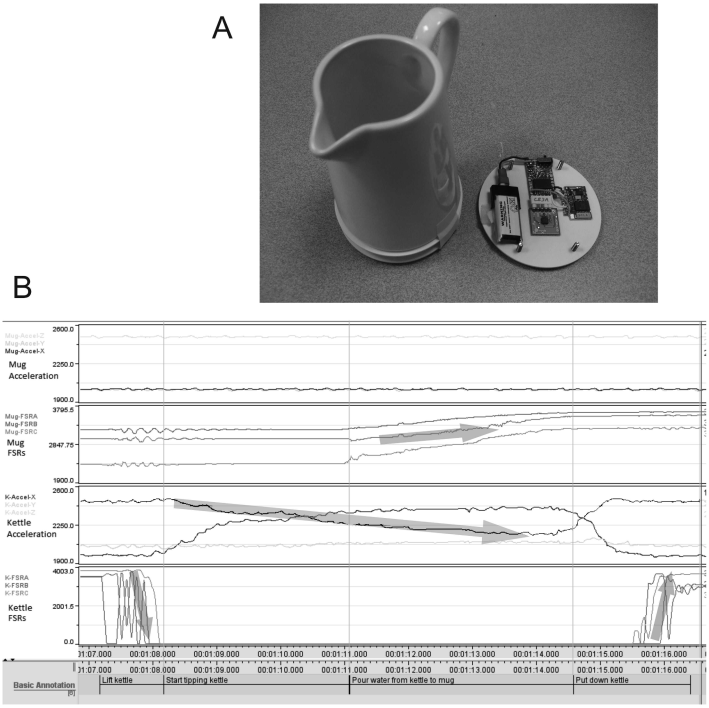


Fig. 3. CogWatch Instrumented Coasters (CIC) (A) Milk jug with fitting CIC (B) Example signals during a tea making scenario. FSRs of the kettle-CIC register lifting the kettle from the table and putting it back (see arrows 4th panel), acceleration sensor of the kettle-CIC register kettle movement and tilt (3rd panel) and FSRs of the mug-CIC register filling of the mug (2nd panel).

4.2 Feedback Devices

Wearable Watch. A multi-functional smart watch (Meta Watch, Ltd. Dallas, Texas; see Fig. 4) is paired with the host computer via Bluetooth 4.0 technology, and is used to produce vibration signals in the event that action resulting in a safety issue has occurred.

VTE GUI monitor. A monitor is in charge of providing Virtual Task Execution (VTE) and to provide feedback and cues to the patient during task execution (see 6 and Fig. 4). A tactile user interface (GUI) allows starting/stopping the application/simulation and selection of the task. Finally the computer is responsible for collecting all the corresponding data in the processor.

5 Action Recognition Techniques

- The architecture of prototype's Action Recognition Algorithm (AAR) system is defined as a set of parallel, Hidden-Markov-Model-based (HMM-based) sub-goal detectors interacting with an interpretative task model (see Fig. 4).
- The AAR interprets the sequences of measurements from the sensors (attached to the objects and tools involved in the task and to the participant's body) in terms of the actions and sub-tasks that the participant is performing. Hidden Markov Models (HMMs) were chosen for this task, because they are an appropriate technology for processing time-varying sequences of data and because many of the issues that arise in the context of the CogWatch application have already been addressed for HMMs in the context of automatic speech recognition (ASR). However, the chosen architecture of the HMM system, namely a set of parallel classifiers each modeling its own separate sub-task (defined by hierarchical task analysis) and using sub-task level HMMs, is novel from the perspective of ASR.
- A Task Model based on a Markov Decision Process (MDP) has been implemented. The states of the task model comprise sub-goals in tea making such as fill kettle, boil water, add teabag to cup, add water, stir, remove teabag, add milk, etc. The model has been tested and verified using synthetic AAR data. The notions of optimal strategy and optimal plan, which are used to define cost functions, are introduced. A shortcoming of the MDP-based approach is that it is not well-suited to dealing with ambiguity. Ambiguity arises because the AAR is imperfect and makes classification errors. A potential solution is described where the MDP is replaced by a Partially Observable MDP.

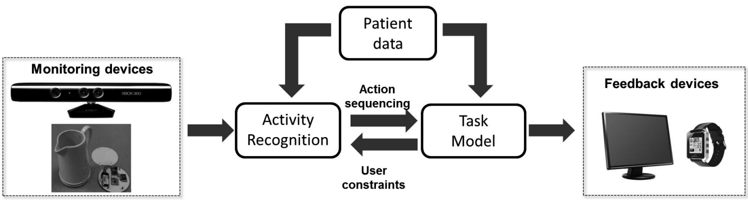


Fig. 4. Integration of Activity Recognition (AAR) and Task Models (TM) into the system

6 Feedback and Cueing Procedures

- Effective multimodal cues and feedback have to be investigated to determine their potential to (a) prospectively guide patients through actions to prevent the occurrence of apraxic errors, to (b) make patients aware of the errors when one is committed in an online fashion and to (c) make patients aware of the appropriate action. The multimodal feedback can include visual displays and markers, auditory signals, speech and gestures, vibrotactile stimulation as well as naturalistic multi-modal distractors which may occur in the patients' environment (see Fig. 4 and 5).

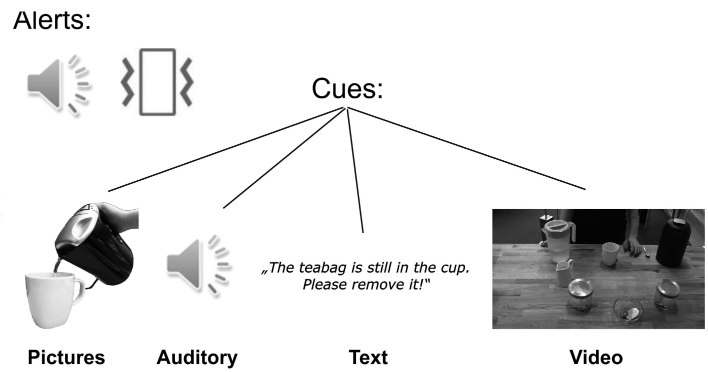


Fig. 5. Examples for feedback and type of cues available for guidance, warning and error correction

When the TM has detected an error, feedback is provided to the patient via verbal, auditory and visual (or a combination of) modalities in order to reduce the consequences of the error, prevent further errors from occurring at later stages of the task, and to preserve goal attainment if possible. Feedback information is provided in a cascading fashion, starting with simple alerts (e.g., auditory signal from a computer, vibratory signal on the wristwatch). If the patient fails to correct his or her actions, more detailed prospective cues are provided (e.g., auditory and text message, still images or videos displayed on the VTE), Feedback information can be tailored to the individual capacities of the patient who may exhibit stroke-related cognitive or language deficits. Initial evaluation of feedback alerts and information yielded good functionality, although more investigation is needed in order to clarify the most effective means of feedback for patients with different apraxic characteristics.

7 Conclusions

In conclusion, initial tests of an automated system to support action execution in daily life situation for apraxic patients have successfully demonstrated the feasibility of the CogWatch concept and the possibility of technical realization. The next steps include the extension of the CogWatch system to other ADL tasks, the tailoring of the tasks and system to the individual needs and abilities of the patient, and the application in home environments with affordable and failure-resistant equipment.

Acknowledgments. The authors wish to thank all CogWatch cooperation partners for their contributions. This work has been supported by the European Commission under the grant FP7-ICT-2011-288912.

References

1. Bowen, A., West, C., Hesketh, A., Vail, A.: Rehabilitation for Apraxia. Evidence for Short-Term Improvements in Activities of Daily Living. Stroke 40, e397 (2009)
2. Buxbaum, L.J.: Ideational apraxia and naturalistic action. Cogn. Neuropsychol. 15, 617–643 (1998)
3. Cogollor, J.M., Hughes, C.M.L., Ferre, M., et al.: Handmade task tracking applied to cognitive rehabilitation. Sensors 12, 14214–14231 (2012)
4. Foundas, A.L., Macauley, B.L., Raymer, A.M., Maher, L.M., Heilman, K.M., Rothi, L.J.G.: Ecological implications of limb apraxia: evidence from mealtime behavior. J. Int. Neuropsychol. Soc. 1, 62–66 (1995)
5. Goldenberg, G., Hagmann, S.: Therapy of activities of daily living in patients with apraxia. Neuropsychological Rehabilitation 8, 123–141 (1998a)
6. Goldenberg, G., Hagmann, S.: Tool use and mechanical problem solving in apraxia. Neuropsychologia 36, 581–589 (1998b)
7. Schwartz, M.F., Buxbaum, L.J.: Naturalistic action. In: Rothi, L.J.G., Heilman, K.M. (eds.) Apraxia: The neuropsychology of Action. Lawrence Erlbaum (1996)
8. Schwartz, M.F., Montgomery, M.W., Buxbaum, L.J., et al.: Naturalistic action impairment in closed head injury. Neuropsychology 12, 13–28 (1998)
9. Smania, N., Girardi, F., Domenicali, C., Lora, E., Aglioti, S.: The rehabilitation of limb apraxia: a study in left-brain-damaged patients. Arch. Phys. Med. Rehabil. 81, 379–388 (2000)

Using Light Guiding to Structure Everyday Life

Guido Kempter[1], Walter Ritter[1], and Markus Canazei[2]

[1]University of Applied Sciences Vorarlberg, Dornbirn, Austria
{guido.kempter,walter.ritter}@fhv.at
[2]Bartenbach Light Laboratory GmbH, Aldrans, Austria
markus.canazei@bartenbach.com

Abstract. We present an approach using room lighting for strengthening individual daily structure or changing structure of daily routines if required. This new healing environment concept includes a monitoring system based on standard passive infrared presence sensors as well as a zonal and ambient room lighting system using direct and indirect lighting with variable light intensities and light colors.

Keywords: Ambient Assisted Living, Mobility, Lighting, Motion Detection.

1 Introduction

The "Harmonised European Time Use Survey" (Eurostat, 2009) shows that people normally have a well structured day. There exist, however, differences between nations, age, as well as gender and there might be also individual differences in daily structure but we can identify daily routines in every sub-sample. The characteristics in a structure range from day/night-rhythm to the regularity in doing specific things at certain times. Impairment in vitality, mobility and cognitive performance also leads to a loss of daily structure as a result of which people mostly have to change to assisted living residences (Sinoo et al., 2011). Consequently, day-structuring measures will become relevant interventions for those people (e.g. for older persons). In this paper we show an approach for healing environments (Huismann et al., 2012) using ambient and zonal room lighting as well as a solution for monitoring the impact of this lighting design on individual daily structure.

2 Lighting Design

There are many well documented psychophysiological effects of light on human (see Veitch & Galasiu, 2012). Light has a direct biological impact on sleep-wake regulation, cognitive performance and physical activity, improves visual performance (e.g. contrast sensitivity, visual acuity, color perception), spatial navigation, and indirectly influences emotions by means of brighten up our mood, creates a room atmosphere for better activation or relaxation and positively influences spatial orientation. Taken these lighting impacts together, light can lead to general wellbeing and health (Fig.1).

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 351–357, 2013.
© Springer-Verlag Berlin Heidelberg 2013

There is some evidence, that spatial appearance manipulated by light also influences mobility as well as assumed temporal and spatial orientation. Figueiro et al. (2011) showed that ambient lighting and lighting cues can have positive effect on gait measures of older adults. Bieske & Dierbach (2006) give evidence that room lighting influences temporal and spatial orientation such as specifying the time of day and personal location within senior residence. People are rating the space of corridors more navigable under certain illumination (Hidayetoglu et al., 2012), and Wardano et al. (2012) show that certain lighting conditions increase duration of stay in well defined social situations.

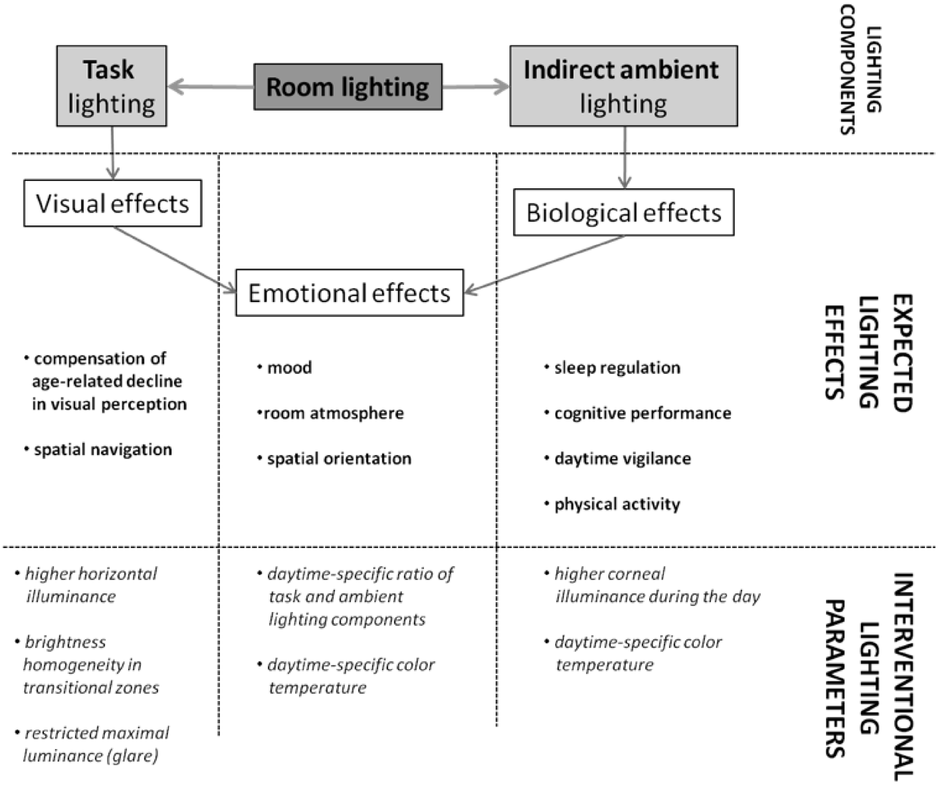


Fig. 1. Psychophysiological model of lighting effects on human

In order to design an environment with illuminated space in a way that supports structuring daily living we defined a lighting scenario which should influence circadian regulation, visual performance, subjective wellbeing as well as temporal and spatial orientation. This scenario consists of task and indirect ambient lighting components and aims at varying ambient and zonal light intensities and light colors over time within private residences. Ambient room lighting provides a higher illuminance level at eye level by means of glare-free indirect lighting components during day and warm-white light color during night and is primarily used as a chronobiological stimulus to regulate sleep-wake cycles, since there is strong evidence that circadian regulation can be influenced by the characteristics of bright light during the day and biological inactive light (e.g. warm-white light) in the evening (Veitch & Galasiu, 2012).

Zonal room lighting with glare-free spotlights will direct attention of residents to certain room zones by means of a higher horizontal illuminance level in the working area to maintain visual tasks and encourage time-of-day related activities, since there is evidence that rooms with distinct purposes and clearly legible meaning provides a space of perception that can better be memorized and may support individuals in their activities of daily living (Wardono et al., 2012). Furthermore, our lighting design is based on evenly lit transitional zones which will guarantee safer navigation and spatial orientation.

Finally our lighting concept aims at varying the color temperature and the amount of light provided either by the task lighting components or the ambient lighting components. Thus we are able to create recreational or stimulating room atmospheres according to daytime specific activities. This will be implemented by automatically triggering lighting by a pre-programmed lighting control system, which has been defined in advance on the basis of individual daily routines (see chapter 3). This includes 24-hours characteristic curves of light intensity and light color for each luminaire which will be turned on by the system, if there is not enough daylight available.

On the basis of such a prototypical activity rhythm, a person with impairments in spatial and temporal orientation could be softly reminded to start a certain activity at a certain time of the day prior to the actual start of the activity. There is, however, an option for manual control of lightings but after predefined delay automatic control will be activated again. Inhabitants are able to provide lighting conditions for optimal visual performance at any time even with totally unexpected events. Zonal lighting will induce unobtrusive stimuli for time-of-day related activities, e.g. smooth transition and contrasts. One example of such a smooth transition is a light alarm clock which starts around the usual wake-up time. The system will turn off all luminaires, if there is no person within the apartment in order to assure an energy-efficient lighting solution.

3 Room Activity Monitoring

Knowing about the individual daily structure of an inhabitant is a key requirement for providing such a lighting system. Besides interviews as basic information source for identifying such a daily structure, an automatic way of discovering changes to this structure is needed to allow for automatic or manual adjustments as well as ongoing evaluation purposes for the lighting system.

Actigraphy is used in a wide array of applications which include fitness or sleep-monitoring (see Ancoli-Israel et al., 2003), and could therefore also be well fitted for the purpose of identifying a daily structure. However, traditional accelerometer-based activity sensor devices have two major practical drawbacks: First, they have to be worn by the persons, mostly on a specific location like wrist, a belt, a foot-strap or in/on a shoe. Unfortunately, often persons don't like to wear such sensors or simply forget to put them on. Also these devices need care on the user's part, like charging them regularly. Second, while such sensors have very good resolution (typically sampling rate up to 20 Hz and acceleration in 3 axis) and a wide array of research has already been done on extracting information out of such signals (e.g, see Bosch et al., 2012), they don't tell in what location and context an activity was happening. One solution to the latter issue would be to add a separate tracking system that records where a person has been at specific times inside the apartment. By combining both information streams, some form of a daily structure could then be retrieved. However, the compliance issue mentioned in first place would still exist.

In an attempt to retrieve activity information in a way that doesn't require a participant to wear sensors and at the same time yields approximate information about the participant's location inside the apartment, we investigated the fitness of standard passive infrared (PIR) presence sensors for such a purpose. PIR-sensors typically yield on/off signals depending on the presence of humans (or, more precisely, depending on location changes of heat sources similar to human bodies) and are slow in reporting their status compared to accelerometers. However, by combining multiple sensors to cover relevant sectors in an apartment and by recording each sensors actuations it is possible to deduce some form of daily activity in an apartment (compare to Lymberopoulos, 2008). Even though the information delivered by the sensors itself is basic, by aggregating and relating sensor-actuations, one can get activity measures within a sector for a specific time period (e.g. number of actuations, duration of the on-phase), or sequences of actions among multiple sectors (e.g. person moves from the bed-room to the toilet an back).

In a small field test we investigated the basic applicability of this idea in a real world context. For this we equipped two apartments of two single elderlies with PIR sensors and a data logging facility. Figure 2 shows the room setup with a total of 12 PIR sectors. Each room was equipped with oneoffering four sections of detection and a minimum detection-interval of one second between actuations.

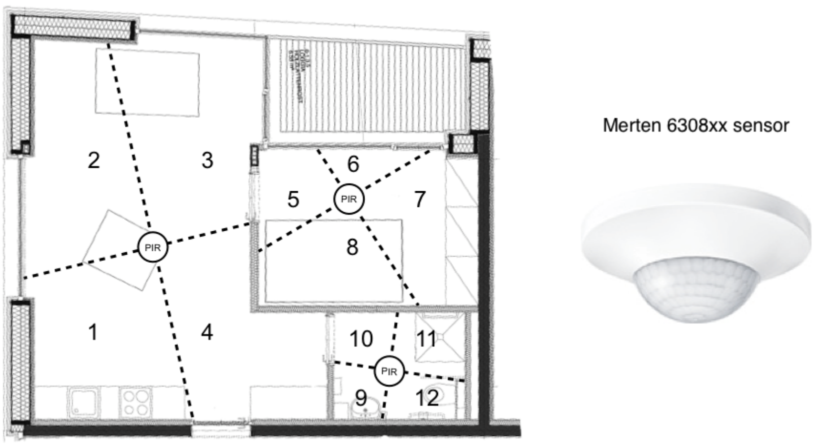


Fig. 2. Plan of test apartment with numbered PIR-sensor sections (total 12 sections)

To see if we could deduce daily structure parameters from the logged data, we evaluated data collected over a period of 6 month. Figure 3 shows a daily activity profile of four PIR-sectors in the living room on sundays averaged over 6 month. The x-axis shows the time whereas the y-axis the number of individual activations in a sector per 15 minutes time slot. One can clearly see the lack of activity from 11:45 to 12:15 hinting at a lunch break.

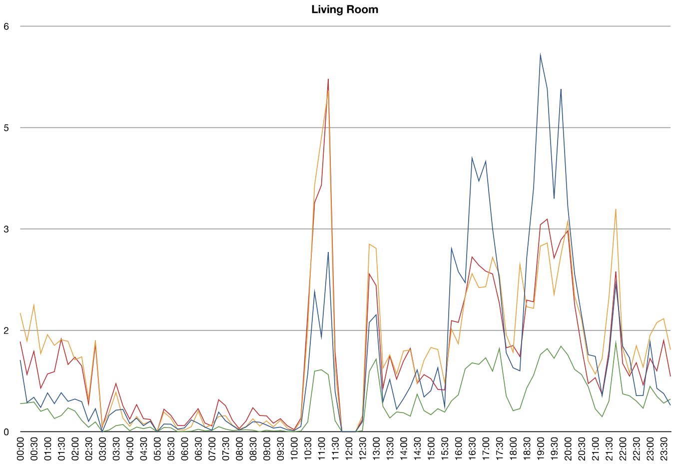


Fig. 3. Activity profile of the 4 PIR-sectors in living room for sundays averaged over 6 month

Figure 4 shows a daily activity profile for two typical days for one apartment. The maximum activity distribution in the separate sectors of the apartment indicate different daily actions on weekends (left side) and weekdays (right side). The x-axis shows the time whereas the y-axis shows the sector-number. Data for this chart has been aggregated over a period of 6 month.

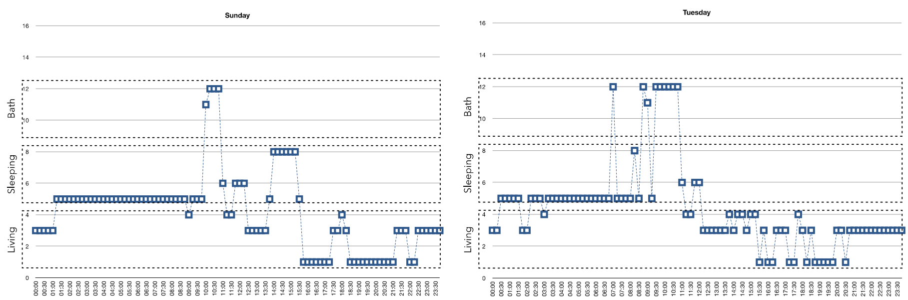


Fig. 4. PIR sector with maximum activity in a time slot of 15 minutes. A comparison between a typical Sunday and Tuesday (averaged over 6 months) show a shift in sleep time by have an hour earlier during weekdays.

Another interesting question was if there would be some detectable change within the daily activity profile between summer- and winter-months, as this is especially relevant in an application using light. Figure 5 shows a shift in sleep/wake times during two summer months and two winter months, meaning this system can in fact be used for detecting changes to the daily structure of a person, without the need for the person to wear or maintain any sensors at all.

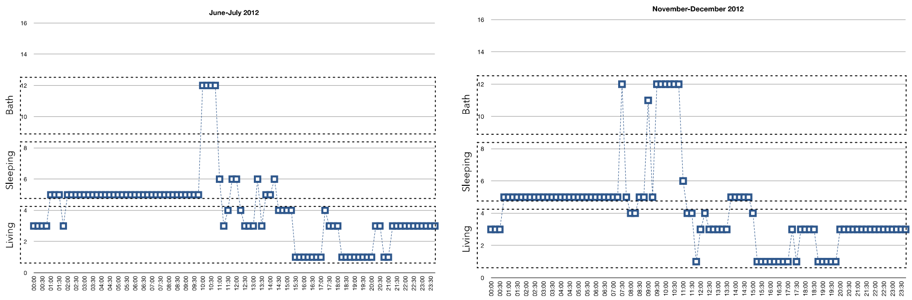


Fig. 5. PIR-sector with maximum activity in a time slot of 15 minutes. A comparison between summer and winter periods shows a shift of sleep-time by 15-30 minutes earlier in winter.

Among the parameters that can be deduced from activity data we recorded in this field study are sleep-/wake times (indicator of circadian rhythm), the time at which various sectors of an apartment feature the most presence (and therefore might be candidates for special light treatments), nightly actions like visiting the toilet or getting something from the kitchen, or the time at which a participant leaves the apartment or comes back (e.g. for getting lunch), and many more.

Despite these promising results, there are two obvious major drawbacks of this PIR-sensor based approach: First, it's only suitable for single person households without pets. Any additional heat-sources would cause additional actuations not related to the main person's actions. In our case, one workaround for this could be to automatically track situations where more than one person is present (e.g. by light barriers that allow for people counting at the entrance, or by observing other room parameters like CO2 levels) and exclude those from data evaluation.

Second, activity that's happening outside the apartment that might have a big effect on a person's health state cannot be tracked in this way. One way to overcome this issue would be to have optional wearable accelerometers that could deliver additional input when worn, but with the core system remaining still functional without them. In the Guiding Light project, we intend to use both solutions to these two drawbacks mentioned here.

4 Conclusion

Our pre-studies show, that we are able to detect individual daily structure within private residence by implementing standard passive infrared (PIR) presence sensors within rooms. Knowing about the individual daily structure of a person we are able to install a room lighting system, which follows the individual needs of an inhabitant. In the next phase we are able to discover whether our zonal and ambient lighting installation can help strengthening individual daily structure or is able to change structure of daily routines if required. Within the project "Guiding Light" we will evaluate whether these assumptions are valid in practice.

Acknowledgement. The project Guiding Light (no AAL-2011-4-033) is funded under European Ambient Assisted Living Joint Programme (AAL JP). Consortium partner are University of Applied Sciences Vorarlberg (coordinator), Tridonic GmbH & Co KG, Bartenbach Lichtlabor GmbH, myVitali AG, Institut für Sozialforschung und Demoskopie OHG and YOUSE GmbH.

References

1. Ancoli-Israel, S., Cole, R., Alessi, C., Chambers, M., Moorcroft, W., Pollak, C.P.: The role of actigraphy in the study of sleep and circadian rhythms. American Academy of Sleep Medicine Review Paper. SLEEP 26(3), 342–392 (2003)
2. Bosch, S., Marin-Perianu, R., Havinga, P.J.M., Horst, A., Marin-Perianu, M., Vasilescu, A.: A study on automatic recognition of object use exploiting motion correlation of wireless sensors. Personal and Ubiquitous Computing 16, 875–895 (2012)
3. Bieske, K., Dierbach, O.: Evaluation des Einsatzes von tageslichtähnlichem Licht in der gerontopsychiatrischen Pflege und Betreuung Hochbetragter. Vortrag auf dem 5. Symposium "Licht und Gesundheit" in Berlin (2006)
4. Eurostat. Harmonised European time use surveys. Official Publications of the European Communities (2009)
5. Figueiro, M.G., Plitnick, B., Rea, M.S., Gras, L.Z., Rea, M.S.: Lighting and perceptual cues: Effects on gait measures of older adults at high and low risk for falls. BMC Geriatrics 11, 2–10 (2011)
6. Huismann, E.R.C.M., Morales, E., van Hoof, J., Kort, H.S.M.: Healing environment: A review of the impact of physical environmental factors on users. Building and Environment 58, 70–80 (2012)
7. Hidayetoglu, M.L., Yildirim, K., Akalin, A.: The effects of color and light on indoor wayfinding and the evaluation of the perceived environment. Journal of Environmental Psychology 32, 50–58 (2012)
8. Lymberopoulos, D., Bamis, A., Savvides, A.: Extracting spatiotemporal human activity patterns in assisted living using a home sensor network. In: Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments, pp. 29.1–29.8. ACM, New York (2008)
9. Marianne, M., Sinoo, M.M., van Hoof, J., Kort, H.S.M.: Light conditions for older adults in the nursing home: Assessment of environmental illuminances and colour temperature. Building and Environment 46, 1917–1927 (2011)
10. Veitch, J.A., Galasiu, A.D.: The physiological and psychological effects of windows, daylight, and view at home: review and research agenda. NRC-IRC Research Report RR-325 (2012)
11. Wardono, P., Hibino, H., Shinichi, K.: Effects of interior colors, lighting and decors on perceived sociability, emotion and behavior related to social dining. Procedia Social and Behavioral Sciences 38, 362–372 (2012)

Usability an Important Goal for the Design of Therapeutic Games for Older Adults

Anne Collins McLaughlin[1,*], Michelle R. Bryant[1], John F. Sprufera[1], Jason C. Allaire[1], and Maribeth Gandy[2]

[1] North Carolina State University, Department of Psychology, Raleigh, NC, USA
{acmclaug,mrgoff,jfsprufe,jcallair)@ncsu.edu
[2] Georgia Institute of Technology, Interactive Media Technology Center, Atlanta, GA, USA
Maribeth.gandy@imtc.gatech.edu

Abstract. The importance of usability for older adults in therapeutic games has not been well explored. Aspects of game-related usability that go beyond typical considerations are a need for challenge, complexity, adoption by novices, motivation for extensive use, and enjoyment. Benefits to considering usability as it pertains to this special population may have long-term benefits for personal independence, maintenance of skills, and rehabilitation from injury. We outline areas we deem critical as a first step to utilizing what we know of older adult use of games for training purposes to facilitate a conversation between designers and researchers for creating and improving games for older players.

1 Overview of Therapeutic Games

Usability for older adults has been considered across many technologies, from ATMs [1] to home automation [2]. Improvements in these technologies often generalize to making many more systems and products easier to use. However, little attention has been given to age-related usability issues with digital games, despite the increasing use of game technologies in therapeutic and leisure environments.

In 2010, the number of adults aged 65 years or older represented 13.1% of the U.S. population. In the United States alone, one million individuals turn 65 each year and by the year 2020 almost 30% of the population will be over 65 [3]. Fortunately, these advancing years can be increasingly vibrant, with older persons living longer independently, often due to advances in technology. Such technologies include memory aids, such as medication reminders [4], home automation controls [2], and technologies that engender social connections and physical/mental activities, such as home sensors that communicate with loved ones [5]. The approach researchers and designers take to ameliorate the effects of age-related cognitive, perceptual, and movement changes through technology will impact an increasingly large older population. Usability of many of these technologies has made great strides, but this is not true with regard to the usability of digital games.

* Corresponding author.

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 358–364, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Games are a unique and important area of study for many reasons, one of which is their potential for therapeutic applications. For example, games have been used as an intervention to improve certain aspects of cognitive functioning in older adults [6]. Hoffman and colleagues found that the use of an immersive virtual reality (VR) could reduce pain in burn victims undergoing undergoing painful treatments [7]. Although the participants in this study tended to be younger, pain management is a concern for many older adults [8]. Games have also been used to rehabilitate patients with motor impairments [9]. Harley and colleagues created a system in which the affected arm of a stroke victim could engage in physical therapy sessions using motion tracking while playing a computer game. The goal of the game therapy was to encourage straightening of the limb, accomplished by matching that activity to the in-game goal of moving an object. The need to move a virtual object created a physical goal that allowed distraction from the effort and physical pain caused by trying to achieve that goal. Rewards built into the game promoted effort toward that goal.

Despite the positive health outcomes listed in these studies, games are not a panacea for the cognitive, perceptual, affective, and movement changes that tend to occur with age. Successful game design requires attention to the accessibility, usability, and playability of such interfaces for older adults. General guidelines exist for the design of systems for older adults [10][11] taking into consideration typical age-related changes. However, these guidelines are not specific to the creation of games or game interfaces as they do not consider the amount of challenge or the balance of skill and luck required for a game to be playable and fun. For example, most design guidelines are aimed at reducing effort, thinking, frustration, or indecision - yet these are all attributes that can be desirable in a game [12]. In this paper we consider past research on what makes games compelling and integrate this with general knowledge of older players and their needs.

2 Enjoyment, Motivation and Social Interaction

Motivation to continue play is a desirable design attribute as therapeutic games may require extensive use to produce benefits. Tyree and McLaughlin [13] examined the types of motivations older adults had for their everyday voluntary activities - these activities were not limited to the digital world though many participants mentioned digital games and online activities. Using the nine types of motivations in the Forbes [14] framework, older adults reported being most motivated for activities that gave them feelings of achievement, which often came from overcoming challenges or increasing skill to meet a challenge. This finding bodes well for game design as older players should not be thought of as passive recipients of therapy, but engaged players seeking challenges appropriate to their skill levels. In the same study, older adults reported being the least motivated by social interactions in their chosen activities. This was not to say they were not motivated by social interaction, only that most of their activities were not motivated by it. For particularly social activities they chose to list, social interaction was a significant motivator. This finding, that older persons did not choose activities based on social motivation was surprising, given the long-standing findings of socio-emotional selectivity theory [15] and underscores the complexity of social interaction as a motivator and the need for further investigation.

Other studies found that social interaction during gameplay has important protective functions. For example, it was found that cognitive abilities predicted game success, which was calculated using the scores provided by a commercial game [16]. This result alone might mean that the game should be made easier for older players, especially those with lower cognitive functioning. However, the study also included a social condition where the same game was used, but played with another person. In the social condition, cognitive ability no longer predicted game success - all players achieved success commensurate with the high ability individual players. Feelings of flow told a similar story. Flow has been defined as a "positive experiential state [that] occurs when the performer is totally connected to the performance, in a situation where personal skills equal required challenges" [17]. McLaughlin, Whitlock, and Allaire [18] found that feelings of flow suffered when players were given a complex version of a game to play compared to a simple version. However, those players assigned to the social interaction condition maintained higher reports of flow no matter what level of difficulty they were given. Other studies with older players found similar results using virtual competitors and human competitors [19] where human competitors increased feelings of immersion and positive affect. In conclusion, social games and activities were most motivating when the goal was to experience social interaction, protective of the rewarding feelings of flow when difficulties arise, and may support performance for older persons who have experienced age-related declines.

Outside of social interaction, McLaughlin, Gandy, Allaire, and Whitlock [20] examined unprompted qualitative data regarding likelihood of enjoyment. Their findings suggested that enjoyment relied both on affective experiences such as emotional selectivity as well as the usability of of the game. They suggested that experiencing increased well-being, positive emotions, and "flow" [21] during gameplay would increase the overall enjoyment for older adults. Motivation and effort were positively correlated to feelings of flow, but difficulties due to perceptual or movement deficits abilities negatively associated with flow [20]. Work examining development of flow in a variety of games and activities found that older players developed the most flow in games that could be adjusted incrementally for difficulty [22].

3 Utilizing Older Adult Capabilities

One tactic for game creation may be to take advantage of the strengths and capabilities that come with age to design games that are motivating and rewarding. For example, declarative knowledge (often called verbal knowledge or crystallized intelligence) increases for most of the lifespan [23]. Use of such knowledge in a therapeutic game could be rewarding and encourage continued play even in the face of difficulty with movement, perception, or more fluid abilities such as spatial ability or attentional control. For example, including an occasional need for vocabulary or trivia knowledge to advance in the game may be motivational and provide renewed energy for the more effortful portions of the games.

A second tactic may be to decrease the demands of the game linked to fluid abilities and fine motor skills. Whitlock, McLaughlin, and Allaire [24] found cognitive benefits for older adults in playing World of Warcraft, but the complexity of the game required extensive in-person and remote support to adhere to the training protocol [25]. For example, participants attended an initial in-person two-hour training session followed by a one-hour practice session. After the in-person training session, participants were provided remote support via e-mail, phone, and in-game experimenter support. Without the use of this assistance older adults may not have been able to overcome the complex nature of the game. However, the caveat must be made that many therapeutic games will require high effort and possible discomfort to be effective. Simplifying the game to a too-basic level could undermine the ultimate goal of gameplay and thus, high levels of support may be initially required.

If older adults spend more time feeling frustrated and unmotivated to reach task goals, they are likely to discontinue participation [20]. Fortunately, games can be designed to be rewarding under many conditions and for incremental improvements in skill. Including rewards where needed, particularly in the initial stages as the player is first introduced to the game, will likely increase adherence to therapeutic games. Incentives, such as bonuses or achievements, may help fulfill therapeutic goals by encouraging further participation and practice. By implementing design principles that take into account the abilities of older adults and produce initial rewards for small improvements, we can likely increase adherence and participation.

4 Shortcomings of Current Games

Although many older adults play digital games [26] it is not difficult to see why they might look at the games available today and choose other activities. We believe this is due to a focus on the younger audience as purchasers of games, with a consequence of this belief being the youth-oriented design of instructions, help, support, and lack of flexibility in gameplay. For example, older players experienced the highest flow in a version of one game for the Nintendo Wii where the experimenters deliberately removed complexity from the game [18]. This level of simplicity is not available in even the easiest levels of most commercial games. This is not to say older players cannot achieve high skill or deal with complexity, only that games to not provide an appropriate ramp to the higher complexity levels for these players. As mentioned previously, difficulty changes should be incremental and controllable to be most likely to produce flow [22].

Previous studies have uncovered the lack of training support provided by current games. For example, Whitlock and colleagues [25] provided multiple avenues of support throughout their study and found that utilization of those supports facilitated game play. They concluded that in-person training was crucial, as well as scheduled remote follow-ups to encourage continued participation. Such training supports had two goals. The first was to allow older adults to function properly within the game space by providing a better mental models of virtual space and gameplay with numerous opportunities for practice and viewing of worked examples [27]. The second goal was to encourage exploration, increase computer self-efficacy, and reduce anxiety. IJsselstein, Nap, Kort, and Poels [28] found that feedback and in-game achievements reduce anxiety. Such a reduction in anxiety may free mental resources to engage in the task [29]. IJsselstein et al. also specified that emphasizing learning goals over performance goals was likely to increase engagement, especially when tasks are complex or require more skill than participants presently possess.

Last, current games may appear to have more costs than benefits to older players. Games, particularly therapeutic ones, need to provide evidence of their benefit to be adopted [30]. One way to do this is to ensure that therapeutic games provide scientifically evaluated benefits. Another way is by considering older adult perception of their need for therapeutic games [31]. If need for a game does not outweigh held values of privacy or independence (and the games appear to violate these values), older adults are not likely to engage. One way this could be applied to games would be to examine the way older adults feel about the utility of a game compared to the social desirability of participating in that game - the game should augment valued activities rather than replacing them [32]. Games need to provide proof of benefits that warrant the time and energy away from other valued activities or become parts of those valued activities to encourage acceptance.

5 Conclusions on the Importance of Usability

The issues raised thus far concerning games may be summarized as global usability concerns. Multiple areas need to be addressed in this regard including complexity, physical requirements, mental models, tutorials, support, and variability in abilities. Though usability principles are important general guidelines, more attention must be paid toward including how these principles are inextricably linked to older adult motivation, enjoyment, achievement, and acceptance of games as a therapeutic options. In circumstances when usability may not be amenable to change, social activity may be a reasonable stop-gap due to the potentially protective effects of social play. However, as these conclusions are based on few studies, more research in understanding social interaction and older adult flow should be performed.

Advancing the understanding of older adult needs with regard to ability, values, and usability in the design field has far reaching implications for the use of games for therapy and leisure. Increasing awareness of aspects of games that fail to meet the needs of the older adult population serves as a starting point for collaboration between researchers and designers. As our world both ages and becomes increasingly technological it is fundamentally important to attend to the special needs of older users and players.

References

1. Rogers, W.A., Fisk, A.D.: ATM Design and Training Issues: Human factors Input to Automatic Teller Machines Can Enhance - and Maybe Increase - Their Use. Ergonomics in Design 1997: The Quarterly of Human Factors Applications 5, 4–9 (1997)

2. Kientz, J.A., Patel, S.N., Jones, B., Price, E., Mynatt, E.D., Abowd, G.D.: The Georgia Tech Aware Home. In: Proceedings of CHI Extended Abstracts on Human Factors in Computing Systems, pp. 3675–3680 (2008)
3. U.S. Census Bureau: U.S. Interim Projections by Age, Sex, Race, and Hispanic Origin [Data file] (2004), `http://www.census.gov/` (retrieved)
4. Inglis, E.A., Szymkowiak, A., Gregor, P., Newell, A.F., Hine, N., Shah, P., Wilson, B.A., Evans, J.: Issues Surrounding the User-Centred Development of a New Interactive Memory Aid. Universal Access in the Information Society 2(3), 226–234 (2003)
5. Mynatt, E.D., Rowan, J., Craighill, S., Jacobs, A.: Digital Family Portraits: Supporting Peace of Mind for Extended Family Members. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 333–340 (2001)
6. Basak, C., Boot, W.R., Voss, M.W., Kramer, A.F.: Can Training in a Real-Time Strategy Video Game Attenuate Cognitive Decline in Older Adults? Psychology and Aging 23(4), 765–777 (2008)
7. Hoffman, H.G., Patterson, D.R., Seibel, E., Soltani, M., Jewett-Leahy, L., Sharar, S.R.: Virtual Reality Pain Control During Burn Wound Debridement in the Hydrotank. The Clinical Journal of Pain 24(4), 299–304 (2008)
8. Hanks-Bell, M., Halvey, K., Paice, J.A.: Pain Assessment and Management in Aging. The Online Journal of Issues in Nursing 9(3) (2004)
9. Harley, L., Robertson, S., Gandy, M., Harbert, S., Britton, D.: The Design of an Interactive Stroke Rehabilitation Gaming System. In: Jacko, J.A. (ed.) Human-Computer Interaction, Part IV, HCII 2011. LNCS, vol. 6764, pp. 167–173. Springer, Heidelberg (2011)
10. Fisk, A.D., Rogers, A.R., Charness, N., Czaja, S.J., Sharit, J.: Designing for Older Adults – Principles and Creative Human Factors Approaches. CRC Press, Boca Raton (2004)
11. Pak, R., McLaughlin, A.C.: Designing Displays for Older Adults, p. 211. CRC Press, Cleremont (2010)
12. Pagulayan, R.J., Keeker, K., Wixon, D., Romero, R., Fuller, T.: User-Centered Design in Games. In: Jacko, J., Sears, A. (eds.) Handbook for Human-Computer Interaction in Interactive Systems, pp. 883–906. Lawrence Erlbaum Associates, Inc., Mahwah (2003)
13. Tyree, R.M., McLaughlin, A.C.: Older Adult Engagement in Activities: All Motivations Are Not Created Equal. In: Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting. Human Factors and Ergonomics Society, Santa Monica (2012)
14. Forbes, D.L.: Toward a Unified Model of Human Motivation. Review of General Psychology 15(2), 85–98 (2011)
15. Carstensen, L.L.: Motivation for Social Contact Across the Life Span: A Theory of Socioemotional Selectivity. In: Jacobs, J.E. (ed.) Nebraska Symposium on Motivation: 1992, Developmental Perspectives on Motivation, pp. 209–254. University of Nebraska Press, Lincoln (1993)
16. Leidheiser, W., Juliani, A.W., McLaughlin, A.C.: Cognitive Ability Predicts Performance in a Novel Task but is Moderated by Social Interaction. Poster presented at the North Carolina Cognition Conference (February 2013)
17. Jackson, S.A., Marsh, H.: Development and validation of a scale to measure optimal experience: The Flow State Scale. Journal of Sport and Exercise Psychology 18(1), 17–35 (1996)
18. McLaughlin, A.C., Whitlock, L.A., Allaire, J.C.: Investigation into the Human Factors of Motivated Engagement in Complex Activity by Older Adults. In: Meeting of the American Psychological Association, Orlando, FL (2012)

19. Gajadhar, B.J., Nap, H.H., de Kort, Y.A.W., IJsselstein, W.A.: Out of Sight, Out of Mind: Coplayer Effects on Seniors' Player Experience. In: Proceedings of the 3rd Annual International Conference of Fun and Games, pp. 74–83 (2010)
20. McLaughlin, A.C., Gandy, M., Allaire, J.C., Whitlock, L.A.: Putting Fun into Video Games for Older Adults. Ergonomics in Design: The Quarterly of Human Factors Applications 20, 13–22 (2012)
21. Hsu, C., Hsi-Peng, L.: Why Do People Play On-line Games? An Extended TAM with Social Influences and Flow Experience. Information Management 41(7), 853–868 (2004)
22. Belchior, P., Marsiske, M., Sisco, S.M., Yam, A., Mann, W.: Older Adults' Engagement with a Video Game Training Program. Activities, Adaptation and Aging 36(4), 269–279 (2012)
23. Baltes, P.B.: Theoretical Propositions of Life-Span Developmental Psychology: On the Dynamics Between Growth and Decline. Developmental Psychology 23, 611–626 (1987)
24. Whitlock, L.A., McLaughlin, A.C., Allaire, J.C.: Individual Differences in Response to Cognitive Training: Using a Multi-Modal, Attentionally Demanding Game-Based Intervention. Computers in Human Behavior 28(4), 1091–1096 (2012)
25. Whitlock, L.A., McLaughlin, A.C., Allaire, J.C.: Training Requirements of a Video Game-based Cognitive Intervention for Older Adults: Lessons Learned. In: Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting, pp. 2343–2346. Human Factors and Ergonomics Society, Santa Monica (2010)
26. Allaire, J.C., McLaughlin, A.C., Trujillo, A., Whitlock, L.A., LaPorte, L.D., Gandy, M.: Successful Aging Through Digital Games: Socioemotional Differences Between Older Adult Gamers and Non-gamers. Computers in Human Behavior (in press)
27. Paas, F.G.W.C., Van Merriënboer, J.J.G.: Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. Journal of Educational Psychology 86(1), 122–133 (1994)
28. Ijsselstein, W., Nap, H.H., de Kort, Y., Poels, K.: Digital Game Design for Elderly Users. In: Proceedings of the 2007 Conference on Future Play, pp. 17–22 (2007)
29. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory and instructional design. Educational Psychologist 38(1), 1–4 (2003)
30. Melenhorst, A.S., Rogers, W.A., Bouwhuis, D.G.: Older Adults' Motivated Choice for Technological Innovation: Evidence for Benefit Driven Selectivity. Psychology and Aging 21, 190–195 (2006)
31. Hanson, V.L., Gibson, L., Bobrowciz, A., MacKay, A.: Engaging Those Who Are Disinterested: Access for Digitally Excluded Older Adults. In: ACM CHI 2010, Atlanta, GA, USA (2010)
32. Sellen, A., Rogers, Y., Harper, R., Rodden, T.: Reflecting Human Values in the Digital Age. Communications of the ACM - Being Human in the Digital Age 52(3), 58–66 (2009)

Creating User-Friendly Healing Environments with Adaptable Lighting for Senior Citizens

Christoph Nedopil, Cornelia Schauber, and Sebastian Glende

YOUSE GmbH, Anzinger Straße 4,
81671 München, Germany
{Christoph.Nedopil,Cornelia.Schauber,Sebastian.Glende}@youse.de

Abstract. Identifying user-friendly use cases for technologies under development is often a difficult endeavor, especially when designing healing environments for the elderly, due to the absence of comparable technologies and the little technology experience in the target group. The principles of user centered design (UCD) have been successfully applied with professionals and lead users, but it is much more difficult for the development of healing environments like AAL-systems (ambient assisted living systems). This article describes the user-centered development of use cases for an innovative lighting system ("GuidingLight"[1]) aimed at increasing the independence and well-being of senior adults.

Keywords: GuidingLight, lighting system, ambient assisted living (AAL), healing environments, senior adults, use cases, activities of daily living (ADLs), expert interviews, user-centered design (UCD).

1 Introduction

Creating user friendly "Healing Environments" encompasses two challenges: creating user-friendliness, and the realization of healing and health environments. According to the salutogenetic approach by Antonovsky [1], healing environments support human health, functioning, and well-being on different levels. Research within the field of "Ambient Assisted Living" (AAL) can be interpreted as a realization of an Optimal Healing Environment Approach by means of technlogical aspects, as the AAL framework aims at improving and maintaining quality of life (QoL), health and well-being of (mostly elderly) people [2].

Many European countries and the European Union financially support the development of AAL technologies in order to handle the upcoming challenges and to realize the market potentials related to the demographic changes. If designed well, AAL technologies can indeed help to compensate for many age-related deficits and problems, and improve quality of life: for example by enhancing mobility, autonomy, or participation in social life, by providing emergency systems, communication channels with the family or friends, or by the constant monitoring of vital data parameters.

[1] This project is funded by the European Union and the German Federal Ministry of Education and Research under the AAL Joint Programme (AAL-2011-4-033).

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 365–373, 2013.
© Springer-Verlag Berlin Heidelberg 2013

However, the employment of technology for improving and maintaining the health and well-being of elderly people is a controversial topic: Seniors apprehend the replacement of human contact by machines, caregivers fear the loss of their jobs, and society struggles with an ethical dealing with the elderly. These and other reservations need to be taken into account to foster broad acceptance of AAL systems.

Moreover, senior users are often not as familiar with modern technology or interfaces as younger users. Handling technology means a much higher cognitive and emotional effort to them, so they are only willing to afford and utilize technology if it comes with a clear added value and is easy to use. Results from a review with 18 German AAL projects revealed that a lack of proven necessity and usability are among the main barriers of market success of AAL technologies [3]. In other words, it is extremely important to consider the requirements and expectations of the target group when creating innovations in general, and healing environments in particular.

The user-centered design (UCD) approach allows incorporating the users' and stakeholders' perspectives continuously into the innovation process. In the remainder of this paper, we will illustrate the methods and results of creating a user friendly healing and health environment by applying the UCD. We will employ the example of the EU-funded AAL research project "GuidingLight", in which an intelligent lighting system for senior homes is developed. The AAL project is executed by a research consortium consisting of various industrial leaders in light and sensor technology (Bartenbach Lichtlabor, Tridonic, myVitali), two research companies (apollis, YOUSE) and a higher education institution (University of Applied Sciences, Vorarlberg). The goal of the light system is to enhance the mobility, the spatial and temporal orientation, and the well-being of elderly people. The system includes motion sensors and controllable lamps with varying brightness and colour that support the individual throughout the day by optimal lighting conditions in the private home.

In the following, we will first introduce the process and the benefits of UCD (chapter 2), followed by a practical example of how UCD was applied to the use case definition for the "GuidingLight" project (chapter 3). In the last chapter (chapter 4), the advantages of UCD for AAL projects will be summarized and discussed.

2 The User-Centered Design Approach

User-centered design takes care of the fact that designers and technicians do not always know how the end-user experiences a product: Do the implemented functions support the main goals the user wants to achieve? Is the use of the product intuitive for laymen? Does the product design consider basic conditions of the context? UCD helps to develop products that are easy to understand, fun to use, and bring benefit (e.g. increased mobility) from a user's perspective. For developers, UCD can help to accelerate development cycles, save cost (e.g. less re-work at the end of the innovation cycle) and increase sales.

The paradigm of user-centered design aims at the integration of user requirements from the beginning throughout the complete development cycle by the employment of user integration (UI) methods such as interviews, field studies, or usability tests.

UCD helps to collect important information about the target users, their needs, and their goals. The findings are incorporated into decisions about the (re)design of the product. Opposite to classical linear development models, the UCD model allows for a user-centered focus in very early development phases, thus preventing the pursuit of product ideas that lack acceptance of the target group [4].

The process of user-centered design is described by the ISO 9241-210 standard [5]. YOUSE has adapted the UCD process model to research projects focusing on the special needs of senior users. The model comprises iterative stages of UI actions (see figure 1): The planning of the UCD process in phase 1 is a perfect occasion to enhance the awareness for user-centric aspects within the project team und to incorporate UCD methods into the process from the beginning. Second, characteristics of the user and the context are gathered and analyzed for a deeper understanding of the most prominent requirements (phase 2). Based on the insights from phase 2, ideas for solutions are created (phase 3). In phase 4, these ideas are discussed regarding the collected requirements, and the most appropriate ones are selected. In phase 5, the realized solutions are evaluated against the collected requirements and specifications. If the evaluation result is positive, the product is complete and can be launched (phase 6); otherwise, prior steps have to be re-run until a satisfying solution has been created.

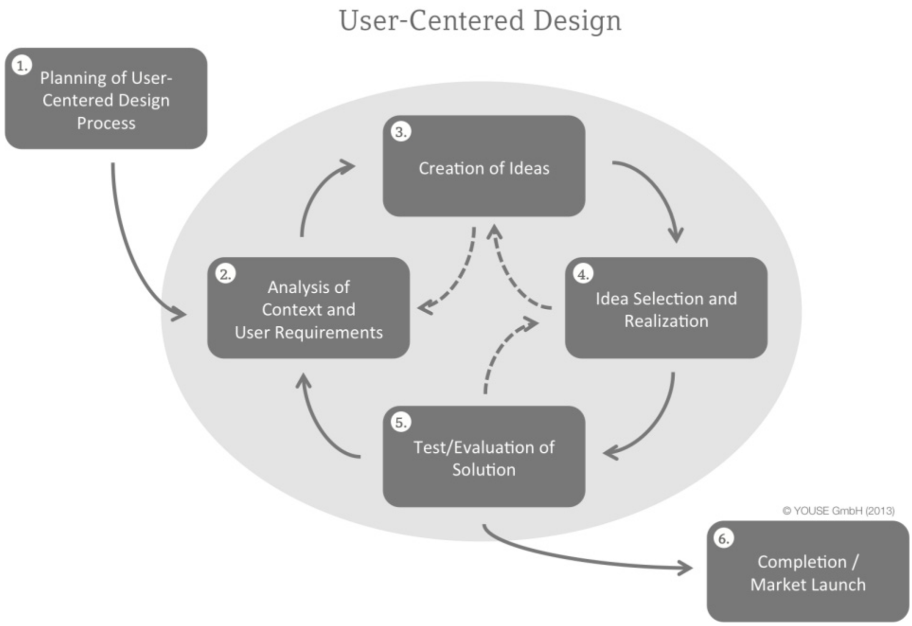


Fig. 1. User involvement into the development of innovative products or services according to the human-centered design approach (based on ISO 9241-210)

In the early phases (phase 2 and 3) of the UCD process, the focus is on researching needs of users and stakeholders, typical goals that the product ought to support, or characteristics of the usage context that the product should be equipped for. To this aim, qualitative information is gathered and refined in subsequent steps. Techniques like brainstorming or field observations, as well as the analysis of empirical data or customer feedback information are examples of frequently used qualitative UI methods.

The output of these early phases concerning the application ideas can be easily summed up in the form of use cases. Use cases describe a sequence of interactions between a user and a system in order to accomplish certain tasks. By defining each step of this interaction, use cases help to analyze the functional requirements or activities of a system that are necessary to support the user's goal achievement. In the case of a lighting system, a use case could refer to a senior getting up at night to go to the bathroom, and the type of light intervention that would support that activity.

With the help of use cases, the technical solution can be developed (phase 4) and evaluated by qualitative and quantitative methods (see phase 5). Prototypes should be developed for usability testing with users or experts, e.g. by sketching the solutions (e.g. paper prototyping, click-dummies) or by creating a mockup through the Wizard of Oz technique (i.e. a human simulating the system's response in real time). The analysis of performance or assessment parameters allows for the testing of assumptions about preferences, the comparison of alternative solutions, or the verification of postulated effects. If all goes well and the evaluation is positive, the usability (or lack thereof) clearly shouldn't be a hurdle for the market success.

After introducing the theoretical background of user-centered design, the following section will give a real-life example of applying the UCD process in the early development phases of the "GuidingLight" system, especially the development of use cases.

3 Defining User-Friendly Use Cases

According to the UCD process model, user-friendly use cases are based on a thorough analysis of the context requirements and the user needs. For the development of use cases for the research project "GuidingLight", we had to get insights into everyday problems of seniors, and to select those that can be solved or ameliorated by light interventions.

As mentioned before, use cases specify the interaction of the user and the system. In more detail, they are the core element of every innovation idea, since they imply the following aspects:

- User problems and appropriate technical solutions
- Variables to detect the occurrence of a problem
- Variables to measure the postulated effect/benefit of the solution

In other words, use cases combine aspects from business models (added value), the technical implementation (sensors and actuators), and the research hypothesis behind a project (measurable effect).

We realized the use case definition by taking the following steps (see the following sections for details): First, we analyzed the existing literature on activities of daily living (ADLs) and instrumental activities of daily living (IADLs) [6,7,8] to get acquainted with the lives of elderly people and typical activities that might cause problems and could be supported by enhanced light conditions. Second, we evaluated and prioritized the identified activities relevant to "GuidingLight" with medical experts (physiotherapists and occupational therapists) as well as caregivers during personal interviews. And third, we evaluated the use cases with regard to the financial and technical realities within the research consortium.

3.1 Analysis of Requirements: Daily Activities of Senior Adults

The idea of "GuidingLight" is to support elderly people with their everyday activities by an intelligent light system. So, what do seniors typically do during their days – and when do they experience difficulties that could probably be reduced by lighting interventions? Apart from biological regularities, lists with activities of daily living (so-called "ADL's") are a rich source of information giving insight into elderly people's everyday lives.

Activities of daily living (ADL's) origin from holistic care models from the 1960s and are based on Maslow's motivation theory [9]. They describe repeating human activities (e.g. eating, drinking, dressing/undressing) aiming at the fulfillment of a wide range of basic physiological and psychological needs, which might cause difficulties at older age or illness. ADL's are usually represented as lists of activities with grades of autonomy levels for each (e.g. "completely independent – needs help – unable"). ADL lists can be used for the diagnosis of care dependency, or for the planning or evaluation of care measures. The Barthel index, as a common example of an ADL-list, was originally developed to assess the severity of disability regarding personal care and mobility of stroke patients [10]. Scores are obtained using direct observation, self-report, or responses from family/friends.

Other screening instruments focus on different types of activities that are necessary for independent functioning in the community, referred to as "instrumental activities of daily living" (IADL's). They include tasks like meal preparation, cleaning, doing the laundry, shopping, or using the telephone (see e.g. [7]).

We analyzed the various lists with activities from ADL and IADL checklists for the development of use cases for "GuidingLight" with experts (see next section).

3.2 Idea Creation: Expert Interviews

In order to evaluate and prioritize the activities identified in the previous phase, experts in the fields of gerontology, physiology, physiotherapy and caretaking were contacted. Since the overall aim of "GuidingLight" is the enhancement of seniors' mobility in the first place, we recruited two nursing managers, one physiotherapist, and one gerontologist to learn more about the everyday problems of elderly people that impair their mobility and independence. The collected information about possible light intervention cases was corrected and amended together with the experts.

Based on the interviews (each of 1 to 2 hours length), a total of 13 use cases were created (see Figure 2). The use cases – as the original (I)ADL lists – relate to different need levels according to Maslow [9]. We deem this assignment extremely important for two reasons: first, the use cases did not only focus on deficit-oriented, low-level needs, but also on the enhancement of higher-order aspects like "self-esteem" or "self-actualization", giving the resulting system a more positive emphasis. And second, the inclusion of higher-order needs reduces the threshold for target groups to acquire such a system. In other words, the system is not only attractive for seniors suffering from heavy limitations, but also for rather independent seniors who want to profit from the system's comfort functions. The support of the light system could then be extended in a modular way, according to the user's needs.

The resulting use cases were presented to and discussed with the project partners in order to select the most appropriate ones.

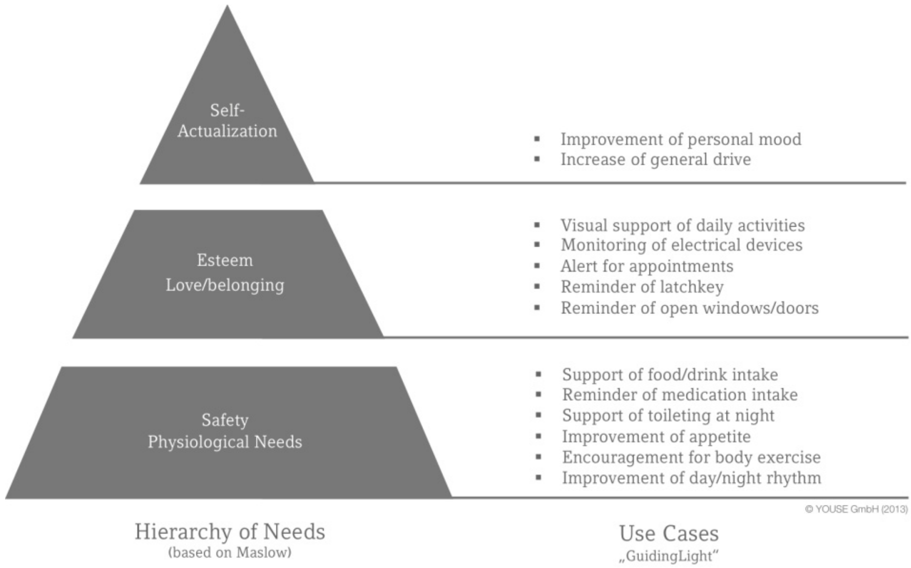


Fig. 2. Use Cases for "GuidingLight", developed from ADL analysis and expert interviews, assigned to Maslow's hierarchy of needs

3.3 Idea Selection: Fusion with Technical and Financial Aspects

The identified use cases were written and presented in form of user stories, so that they could be appropriately discussed within the project consortium. The evaluation was based on the following criteria:

- Compliance with the basic goals of "GuidingLight" (i.e. enhancement of mobility, orientation, and well-being)
- Potential for enhancing healing and health effects for a broad range of users
- Conformity with ethical aspects (e.g. no paternalism of users!)
- Technical feasibility (e.g. sensor restrictions, limits of data collection for light control)
- Research aspects (e.g. necessity to prove effectiveness of light interventions)
- Financial scope (e.g. cost of implementation)
- Market scope (e.g. possible market success)

The consortium finally decided on six use cases to be implemented and tested with senior users (see table 1): the improvement of personal mood, the increase of drive and mobility, the visual support of daily activities (e.g. reading, cooking), the support of regular food intake (pleasant and appetizing light), the enhancement of safety for night-time toileting, and the improvement of the personal day-night rhythm.

These use cases will be tested with 18 senior households in Austria, Italy, and Germany from summer 2013 on for several months.

Table 1. Final use cases for the "GuidingLight" system. It is assumed that positive light effects between use cases are indirectly related (e.g. visual support of daily activities will probably also improve the mood through the experience of enhanced self-efficacy).

Use Case	Description of light intervention	Needs (Maslow)	Target variable (GuidingLight)
Mood	Seniors often suffer from depression, due to the experience of progressing limitations, the death of relatives or friends, or increasing isolation. A bright ambient lighting of the main rooms supports the enhancement of the mood.	Self-actualization, Self-esteem	Well-being, Mobility
Drive/ Mobility	Independent from actual limitations or pain, seniors often lack the drive to engage in activities or to move about. The reasons can be a depressed mood, or feelings of insecurity and fears of falling. Ambient light clues support a regular day structure and activate the user.	Self-actualization, Self-esteem	Well-being, Mobility, Orientation
Visual support	Due to age-related changes in the eye, seniors experience various visual difficulties that hamper daily activities like reading, cooking, or shaving. Distinct spotlights for workspaces enhance visual performance.	Self-esteem, Safety	Mobility, Orientation
Toileting (at night)	Seniors need to go to the bathroom at night regularly. In order to safe electricity, or not to be dazzled, they seldom switch on the light. An automatic, motion-sensitive, anti-glare light guides them to the bathroom and back.	Safety	Mobility, Orientation
Sleep	Many seniors suffer from sleeping disorders like falling asleep lately, or awakening during night or early in the morning. Bright light during the day and smooth transitions in the evening and in the morning support a healthy sleep-wake-cycle.	Physiological needs	Well-being, Mobility, Orientation
Food Intake	The gustatory sense and appetite attenuate while ageing. A pleasant and appetizing light situation during meals supports a regular food intake.	Physiological needs	Well-being

4 Summary and Conclusions

The AAL project "GuidingLight" aims at the implementation of a healing and health environments for senior citizens, utilizing the benefits of light for the needs and hassles of elderly people. For the acceptance and success of this system, the requirements of senior users ought to be taken into account. The model of user-centered design (UCD) offers a proper solution to realize that, even in early development stages, long before a prototype has been designed.

We have demonstrated the procedure of defining use cases that reconcile the requirements of the users with the project framework including ethical or technical aspects. The effort of the analysis and the interviews remained within an acceptable range, yet the resulting input was extremely important for the future success of the project. We consider it an invaluable advantage that the consortium intensively addressed the overall aim of the system, the target group, and the added benefits of the light system for elderly users. In the course of the process, implicit assumptions were disclosed and explicitly discussed, which helped us to develop a common vision of what we want to accomplish. It also set the users' perspective against technical or scientific aspects.

We believe that Maslow's hierarchy of needs is an appropriate way of categorizing use cases for elderly and developing user-centered products: It provides a humanistic and psychological point of view to the technical development team; furthermore, this framework will help build a profound business model for "GuidingLight", as it categorizes human needs regarding different levels of physiological and psychological independence. This provides ideas on how to market the system – once developed – to different customer groups – something that has been poorly executed in many similar research projects so far (cf. [11]).

The next steps of the "GuidingLight" project will be (1) to put the use cases into practice technologically, (2) to validate the efficacy on the well-being, spatial and timely orientation as well as on mobility increases, and (3) to disseminate the knowledge and the technology, in order to make the benefits of "GuidingLight" available to elderly people.

Acknowledgments. The authors would like to express their appreciation to the partners of the "GuidingLight" project consortium in sharing their expert knowledge and discussing the perspective of the target users with an open mind. Our special thanks goes to Dr. Herbert Plischke who has provided us with his medical expertise and his invaluable contacts throughout the project. We would also like to thank the volunteer experts for their feedback and advice, as well as the German Federal Ministry of Education and Research (BMBF) and the European Union for the funding of the project "GuidingLight".

References

1. Antonovsky, A.: Health, Stress, and Coping: New Perspectives on Mental and Physical Well-being. Jossey Bass, San Francisco (1979)
2. Kohls, N., Ives, J., Sakallaris, B., Jonas, W.: Towards enhancing healing processes by developing and facilitatingtechnological aspects of Optimal Healing Environments - setting the stage. In: Bachmann, M., Ives, J., Kohls, N., Plischke, H. (eds.) Toward Optimal Healing Environments. Symposium on Assistive Systems for Social, Personal, and Health Interaction, vol. 1, pp. 18–23. Samueli Institute Conference Proceedings, Virginia (2012)
3. Glende, S., Nedopil, C., Friesdorf, W., Podtschaske, B., Stahl, M.: Nutzerabhängige Innovationsbarrieren im Bereich altersgerechter Assistenzsysteme. BMBF-VDE Innovationspartnerschaft AAL. VDI/VDE Innovation + Technik GmbH, Berlin (2010)
4. Glende, S.: Senior User Integration - Konzepte, Werkzeuge und Fallbeispiele. Suedwestdeutscher Verlag für Hochschulschriften, Saarbrücken (2010)
5. ISO 9241-210:2010 - Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems (2010)
6. Katz, S.: Assessing Self-maintenance: Activities of Daily Living, Mobility, and Instrumental Activities of Daily Living. J. Am. Geriatr. Soc. 31(12), 721–727 (1983)
7. Lawton, M.P., Brody, E.M.: Assessment of Older People: Self-Maintaining and Instrumental Activities of Daily Living. Gerontologist 9(3), 179–186 (1969)
8. Millán-Calenti, J.C., Tubio, J., et al.: Prevalence of Functional Disability in Acitivties of Daily Living (ADL), Instrumental Activities of Daily Living (IADL) and Associated Factors, as Predictors of Morbidity and Mortality. Archives of Gerontology and Geriatrics 50, 306–310 (2010)
9. Maslow, A.H.: A theory of human motivation. Psychological Review 50(4), 370–396 (1943)
10. Mahoney, F.I., Barthel, D.: Functional evaluation: The Barthel Index. Maryland State Med. Journal 14, 56–61 (1965)
11. Nedopil, C., Klaus, H., et al.: Mobile Apps für die Generation Plus. Deutsche Telekom Laboratories/YOUSE, Berlin (2011)

Aged People's Emotion Elicited by Touching Materials of Armrests

Tyan-Yu Wu and Chia-Ying Peng

Chang Gung University, Taiwan
tnyuwu@mail.cgu.edu.tw, pink_happyangel@yahoo.com.tw

Abstract. A chair with armrests is an important object required by aged people for sitting, when their physical strength decline gradually. Users' emotions can be evoked by touching material of the armrest. This study attempts to explore aged people's emotional responses as evoked by interaction with materials on armrests. An experiment was conducted to explore emotional responses evoked by touching six different materials of a armrest between aged people and young adults. The results indicate that a chair's armrests made from fabric can enhance users' pleasure senses better, as compared with the other five materials. The result also showed that aged people were willing to give a higher mean score on pleasure than young adults were.

Keywords: Aged people, Sense of touch, Emotion.

1 Introduction

As is well known, armchairs are essential objects for elderly people (Jonsson & Sperling, 2010) for sitting. It is because aged people's physical strength and visual abilities gradually decline as they age. However, their tactile abilities are not affected greatly with age. Thus, they depend on tactual sensation to feel objects frequently during their daily lives. Sitting on a chair, elderly people do not only enjoy the comfort of sitting, but also feel a sense of security through tactual sensations fed by surface materials of the armrests. They experience a sense of uplifting satisfaction and delight in armchair materials (e.g. armrests) that meet their physical needs. Therefore, how aged people sense and response to material used on a chair armrest is becoming a key issue to be considered when designing a comfortable chair. The aim of this study is to explore aged people's emotional responses as evoked by interaction with materials of armrests (see Fig. 1).

2 Sense of Touch

Chang (1995) defines sense of touch as a type of skin sense, sensory receptor of skin, which adopts stimuli from outside world. It is a process comprising multiple senses of touch, pain, temperature and so on. Touching is being in physical contact and, as

D. Harris (Ed.): EPCE/HCII 2013, Part II, LNAI 8020, pp. 374–380, 2013.
© Springer-Verlag Berlin Heidelberg 2013

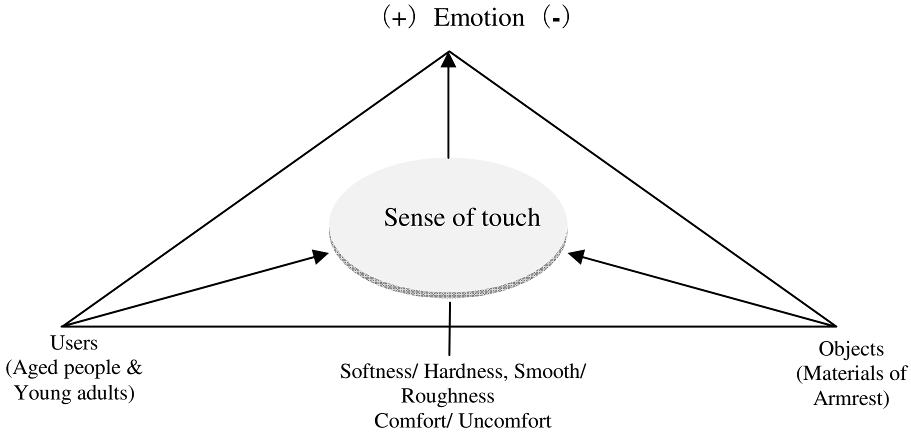


Fig. 1. Research structure

such, is the basis for the feeling of being in contact. Within this contact, touch is a strong basis for the development of feeling of affection (Sonneveld & Schifferstein, 2008; Fields, 2003; Montagu, 1971). Touch also induces human affection. In general, touch includes touch sensation, pressure sensation, and vibration sensation. Touch sensation can be drawn from two types of tactual actions: 1) active touch, a skin sensation derived from using physical body to touch objects actively, 2) passive touch, a skin sensation derived from being touched by other objects. This study on sense of touch is focused on active touch as people actively use their palms and fingers to contact surface of stimuli and to feel intuitively. To control the variables in our research, the sensation derived from the shape and volume of armrest will be excluded.

3 Emotions through Touch Sensations

Both Jordan's (2000) physio-pleasure and Norman's (2004) visceral level of emotion emphasized users' pleasure and emotions which are elicited through a sensitive interaction with products. de Rouvray, Bassereau, Duchamp, Schneider, & Charbonneau (2008) stated that a user pays more attention to emotional responses when sitting on an office chair. This intention may add an extra emotional value (Desmet, Overbeeke, & Tax, 2001) to the product offerings. Thus, designing a chair with pleasure should enhance sensation qualities in terms of comfort. Zuo et al., (2004) proved that touching a rough surface, users may respond with a negative emotion. According to emotional appraisal theory, users may respond positively if touching a surface made with proper material. Conversely, users may respond negatively when touching a surface made of poor material.

Different materials convey different characteristics. Klatzky et al., (1987) and Lederman et al., (1986) addressed that a product's properties, for instance, hardness, roughness, softness and slink can affect users' appraisal to a product and further affect their purchase decisions. Therefore, through touch, users feel safe (Sonneveld & Schifferstein, 2008) and they often use adjective terms (e.g. soft/ hard, smooth/ rough and un / comfort) to describe the feeling when touching materials skins.

Thus, understanding how material sensation affecting a user's emotion is very important when developing a product with emotions. This paper attempts to explore aged people and young adults' emotions (i.e. pleasure and arousal) evoked through touching materials and the relationship between materials properties (i.e. softness/hardness, smooth/roughness, and un/comfort) and emotions.

4 Aging Difference towards Touch

Jonsson and Sperling（2010）interviewed 18 aged people regarding their requirements to a furniture design. The result concluded that aged people preferred having furniture constructed with comfortable materials; indeed, aged people expected to buy a product which can make them feel pleasure when it is in used. Specifically, some research found that middle-aged and older adults have more intense and subjective emotional experiences than younger adults have, even though physical responses to emotions were dampened in the two older groups (Kryla-Lighthall N. & Mather M., 2008; Burriss, Powell, & White, 2007). Moreover, the enhanced sense of well-being among aged people reflects both decrease in negative affect and increase in positive affect. Thus, we assume that aged people have greater satisfaction and positive emotional response than young adults do, when touching a material.

5 Methods

An experiment was conducted to explore six materials used frequently in making armrests of a chair. In the experiment, we investigated if the senses of touch derived from materials affect aged people's and young adults' emotions differently.

Subjects

20 aged subjects (11 male, 9 female; average age = 72), and another 20 college students subjects (9 male, 11 female; average age = 25) in ~~a~~ good health condition participated in this experiment.

Stimuli

Firstly, we collected materials used frequently on armrests of a chair from different sources such as furniture books, website, and magazines. Through group discussion, six materials used on armrests more frequently were screened out and adopted for the experiment. They include wood, fabric, leather, metal, plastic, and rubber. These six materials were constructed with 210×297 mm plane sheets and used as stimuli in this experiment (see Pic. 1).

Questionnaire

Questionnaire items includes Mehrabian and Russell's (1974) assessment (i.e. pleasure and arousal), physiological assessment (i.e. uncomfort/comfort, softness/hardness, and smooth/roughness), and behavior assessment (i.e. likeness and purchase motives). The items were designed with 5-point Likert Scale for the experiment.

Experimental Installation Setting

A table with a chair was settled in an isolated and quiet space. There is a black box on the table. On the box, a hole was created for a hand to accessing the materials during the test (see Pic. 1). Each stimulus was located in the black box randomly. Participants' hands are able to go through the hole and touch stimulus without any hindrance as active touch was adopted for the engagement with stimuli in this study.

Experimental Produce

Each participant has to go through four steps to complete the experiment. On Step1 sitting in front of table still, a participant's eyes were covered with a blindfold and instructed to relax, while one of their palms was guided to place on the stimulus. On Step2 a participant was told to start to move the palm forward slowly within six seconds. On Step3 a participant was asked to open his/her eyes and fill up a questionnaire. Then, taking a 20-second break, participants repeated the same processes till finishing six stimuli (see Pic. 2).

Fig. 1. Six stimuli and a black box for the experiment

Fig. 2. Participant engaged with stimuli and checked questionnaire

6 Results and Discussion

6.1 Internal Reliability of PA Questionnaire

With regard to pleasure and arousal (PA) questionnaire, the statistic result shows Cronbach's Alpha > 0.9 for pleasure dimension and over 0.68 for arousal dimension (see Table 1). The result demonstrates that questionnaire has an acceptable internal consistency (Nunnally, 1978).

Table 1. Cronbach's Alpha of PA questionnaire

	Metal	Leather	Rubber	Plastic	wood	Fabric
Pleasure (N=6) (Cronbach's Alpha)	0.972	0.955	0.949	0.958	0.948	0.930
Arousal (N=6) (Cronbach's Alpha)	0.682	0.801	0.687	0.709	0.683	0.798

6.2 Pleasure Elicited through Touch Sensation

The result showed that pleasures of participants including older and young adults evoked by six materials had the following descending order of pleasures: fabric (M=23.23) > wood (M=22.45) > plastic (M=21.90) > leather (M=20.45) > metal (M=20.03) > rubber (M=19.15) (see Fig. 2). The results indicate that a chair's armrests made from fabric can enhance users' pleasure senses better, as compared with the other five materials. It is suggested to have a further experiment to explore the difference types of fabric textures which may evoke participants' emotions differently, due to the different degrees of touch sensation derived from different fabric textures.

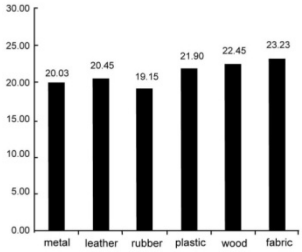


Fig. 3. Means of pleasure among six materials

6.3 Emotional Responses on Six Materials between Aged and Young Adult

In this study, pleasure and arousal (PA) were used to assess participants' emotional responses between aged people and young adults. In Fig. 3, top three materials (i.e. fabric, wood and plastic) elicit higher pleasure responses than others. Among them, both aged people and young adults agreed that fabric can convey the highest degree of pleasure consistently. The result also showed that aged people were willing to give a higher mean score than young adults were. This fact confirms prior research results, which demonstrated that aged people have a greater emotional response on sense of touch, as compared with young adults.

In Fig. 4, among six materials, both groups showed a highest arousal on metal, compared with other six. Unlike other five materials, young adults give a greater arousal score on metal than aged people do. It implies that young adults can sense metal more agilely on arousal and are willing to give a higher score than aged people do.

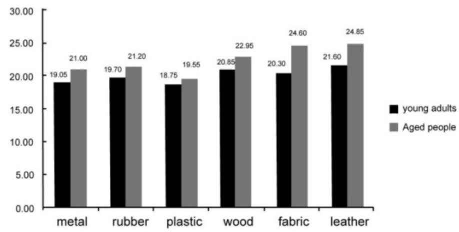


Fig. 4. Pleasure derived from aged and young adults towards touch sensations

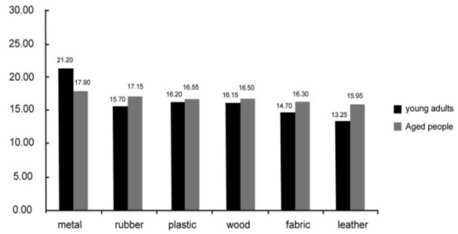


Fig. 5. Arousal derived from aged and young adults towards touch sensations

6.4 The Comparison of Fabric Sensation between Aged People and Young Adults

T test was used to examine the tactile sensation difference between aged and young adults towards six materials. The results demonstrate that there are significant differences on the reaction to materials by two age groups. Table 2 showed that there are significant differences on pleasure, arousal, softness/ hardness, smooth/ roughness, comfort/ safeness, and purchase motivation, except likeness. The result demonstrates that aged people have a greater emotional response (i.e. pleasure & arousal) than young adults do, when touching fabric. The result also showed that aged people have greater sensational responses on softness/ hardness, of fabric than young adults do. In contract, young adults have a greater sensational response on comfort/ safeness than aged people do, as touching fabric. Regarding purchase behavior, aged people are willing to buy an armchair covered with fabric better than young adults do. However, there is no significant on likeness to both aged people and young adults. In sum, the result shows that aged people seem to prefer fabric armrest and are willing to buy it, as compared with young people.

Table 2. T-test for fabric sensation of touch between aged and young adults (single)

Fabric	Young adults		Aged people			
	Mean	sd	Mean	sd	T value	P value
Likeness	3.35	1.268	3.60	1.095	-.667	.254
Purchase	3.05	1.276	3.70	1.218	-1.648	.054^{+}
Pleasure	21.600	5.614	24.850	4.648	-1.994	.027*
arousal	13.250	4.876	15.950	4.084	-1.898	.033*
Soft/ Hard	3.100	1.119	4.050	1.605	-2.171	.018*
Smooth/ Rough	9.750	1.832	10.750	2.712	-1.366	.090^{+}
Comfort/ Safe	11.000	3.387	8.550	3.546	2.234	.016*

*** p ＜.001; **p ＜.01; * p ＜.05; + p ＜.10

References

1. Chang, Z.C.: Modern psychology. Tung Hua Books, Taipei (1995)
2. Desmet, P.M.A., Overbeeke, C.J., Tax, S.J.E.T.: Designing products with added emotional value: Development and application of an approach for research through design. The Design Journal 4(1), 32–47 (2001)

3. de Rouvray, A., Bassereau, J., Duchamp, R., Schneider, J., Charbonneau, S.: Perception and Deception: How Quantity and Quality of Sensory Information Affect Users' Perception of Office Chairs. The Design Journal 11(1), 29–50 (2008)
4. Fields, T.: Touch. MIT Press, London (2003)
5. Jordan, P.: Designing pleasurable products, vol. 7. Taylor & Francis, London (2000)
6. Jonsson, O., Sperling, L.: Wishes for Furniture Design among Persons in the Third Age. In: Jonsson, O., Sperling, L. (eds.) 7th International Conference on Design and Emotion, Chicago, USA (2010)
7. Klatzky, R.L., Lederman, S., Reed, C.: There's more to touch than meets the eye: The salience of object attributes for haptics with and without vision. Journal of Experimental Psychology: General 116(4), 356–369 (1987)
8. Lederman, S., Thorne, G., Jones, B.: Perception of texture by vision and touch: Multidimensionality and intersensory integration. Journal of Experimental Psychology: Human Perception and Performance 12(2), 169–180 (1986)
9. Montagu, A.: Touching. Columbia University Press, New York (1971)
10. Murphy, P., Neill, W., Ackerman, D.: By natures design. Chronicle Books, San Francisco (1993)
11. Mehrabian, A., Russell, J.A.: An Approach to Environmental Psychology. The MIT Press, Cambridge (1974)
12. Norman, D.: Emotional design: Why we love (or hate) everyday things. Basic Books, New York (2004)
13. Nunnally, J.C.: Psychometric theory. McGraw-Hill, New York (1978)
14. Sonneveld, M.H., Schifferstein, H.N.J.: The tactual experience of objects. In: Schifferstein, H.N.J., Hekkert, P. (eds.) Product Experience, pp. 41–67. Elsevier, London (2008)
15. Zuo, H., Hope, T., Jones, M., Castle, P.: Sensory interaction with materials. Design and Emotion: The Experience of Everyday Things, 223 (2004)

Author Index