ORIGINAL ARTICLE

A theoretical framework for designing human-centered automotive automation systems

Husam Muslim[1] · Makoto Itoh[2]

Abstract

Increasingly sophisticated and robust automotive automation systems are being developed to be applied in all aspects of driving. Benefits, such as improving safety, task performance, and workload have been reported. However, several critical accidents involving automation assistance have also been reported. Although automation systems may work appropriately, human factors such as drivers errors, overtrust in and overreliance on automation due to lack of understanding of automation functionalities and limitations as well as distrust caused by automation surprises may trigger inappropriate human–automation interactions that lead to negative consequences. Several important methodologies and efforts for improving human–automation interactions follow the concept of human-centered automation, which claims that the human must have the final authority over the system, have been called. Given that the human-centered automation has been proposed as a more cooperative automation approach to reduce the likelihood of human–machine misunderstanding. This study argues that, especially in critical situations, the way control is handed over between agents can improve human–automation interactions even when the system has the final decision-making authority. As ways of improving human–automation interactions, the study proposes adaptive sharing of control that allows dynamic control distribution between human and system within the same level of automation while the human retains the final authority, and adaptive trading of control in which the control and authority shift between human and system dynamically while changing levels of automation. Authority and control transitions strategies are discussed, compared and clarified in terms of levels and types of automation. Finally, design aspects for determining how and when the control and authority can be shifted between human and automation are proposed with recommendations for future designs.

Keywords Human-centered automation · Human–automation interactions · Authority · Levels of automation · Automation driving system · Adaptive automation · Shared control

1 Introduction

The driving environment is a highly dynamic and potentially high workload domain that requires high human cognitive abilities and cognition to acquire and process surrounding information in a continuous manner (Inagaki 2011). Due to limitations of human information processing abilities,

drivers in such a complex domain are prone to errors (Mannering et al. 2009). Drivers' error, such as a failure to perceive risks and/or react appropriately, constitutes a major cause of road traffic accidents (Treat 1977; Green 2003). Road traffic accidents have a serious social impact in terms of both injury and death, in addition to the economic impact (Dingus et al. 2006). Technology has long been employed in addressing safety-related issues (Dickmanns 2002). Advances in the technologies enabled automation to be applied in all domains of human–machine systems including car driving. Automating driving tasks were mainly introduced to improve the traffic system and road capacity, support humans to control their vehicles more easily, and aid drivers in dangerous conditions (Coelingh et al. 2010; Wille et al. 2010). A wide range of driving automation systems (DAS) have been developed to support drivers

✉ Husam Muslim
hussam@css.risk.tsukuba.ac.jp

Makoto Itoh
itoh@risk.tsukuba.ac.jp

1 Graduate School of Systems and Information Engineering, Risk Engineering, University of Tsukuba, Tsukuba, Japan

2 Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba, Japan

in critical conditions such as imminent collision warnings and avoidance systems, and noncritical routine driving like adaptive cruise control and lane keeping assistance systems. Although the benefits of these systems have been repeatedly demonstrated in terms of safety, comfort and performance, several critical issues that have arisen when using ADS have also been reported (Parasuraman et al. 1993; Sarter et al. 1997). For example, drivers could be overly dependent on automation systems even when systems may not work as expected (Inagaki and Itoh 2013; Lee and See 2004). This may be caused by automation-induced complacency, which occurs when operators become more satisfied with their abilities when the manual task competes with the automated one (Parasuraman and Manzey 2010). In addition, increase of the likelihood of engagement in non-driving related activities has been observed (Lee et al. 2006). Furthermore, poor understanding of the exact capabilities of the assistance system can result in overtrust or distrust in automation (Inagaki 2008). These critical human factors related issues, perhaps, emerged from the fact that in addition to controlling the vehicle in a highly dynamic environment, drivers have to monitor and interact with the automation systems. This means that the drivers, when supported with automation and forced in the role of monitoring, supervising, and intervention that requires a higher level of skills, particularly when a driver has to deal with different levels of automation interventions (Katja et al. 2014), may not perform well. Such critical factors may not be totally eliminated or avoided by instructions or simple short-term trainings.

Automation implementation without sufficient consideration of design implications on human operators, especially in unpredictable and highly dynamic environments like the traffic environment, may lead to misuse, disuse, or abuse of automation (Parasuraman and Riley 1997). A misuse may occur due to lack of human understanding of the system limitations, i.e., a human uses an automation system to handle a task that is outside system design capacity. Drivers may think that an adaptive cruise control system (ACC) can respond to stopped vehicle as it has been identified by Itoh (2012). Disuse occurs when the human does not use the automation to handle tasks that are within the system design capacity. For example, drivers may not activate the lane keeping assistance, which helps the drivers keep the vehicle approximately in the center of the designated lane, where it could have been helpful. Some systems are designed in a way that permits human abuse of automation and affects human performance and skills, such as reduced attention to steering control and increased speed when supported by lane keeping assistance system (Van Der Wiel et al. 2015). These are mainly caused by lack of mutual understanding between human and automation as observed by Norman (1990), Billings (1997) and Abbink et al. (2012). How humans interact with, trust in and accept automation systems remains one of

the greatest challenges that affect the efficacy of assistive technologies not only in automobiles, but also in all aspects of human–machine systems.

Trust in and understanding of automation can be a vicious cycle. On the one hand, humans may not be able to fully understand an automation system until they can develop a certain level of trust in the system to reach their goal. On the other hand, the more the humans understand the automation system the more they can develop an appropriate trust in the system. Abbink et al. (2012) proposed design guidelines for human–automation interactions focusing on improving human understanding of automation systems by keeping the human always in charge of the control and provide the human with continuous feedback about the automation action, levels and boundaries. Other researchers discussed whether human control should be dependent on both agents' abilities in the given situation (Miller and Parasuraman 2007; Kaber 2017; Dekker and Woods 2002). For example, in automotive automation systems several researchers agree that control can be shifted from the humans to automation system in highly critical situations even without human directions (Inagaki 2000; Moray et al. 2000; Kaber and Endsley 2003; Miller and Parasuraman 2007; Prinzel et al. 2003; Wilson and Russell 2007). These arguments in the literature suggest the need for a dynamic control distribution between human operator and automation system, i.e., adaptive automation (Miller 2005; Inagaki 2003; Parasuraman et al. 1992). Adaptive automation provides a dynamic function allocation between humans and systems depending on interacting agents' abilities and limitations in the given situation (Inagaki 2003). Dynamic allocation of control and authority allows the systems to have more than one level of automation (Gao et al. 2006). This means that the authority of balancing the control between agents can be ranged depending on task requirements, risk evaluation, and agent ability. This raises the question of whether the automation should be allowed to retain the authority to trade the control from the human to the system without a human directive. Such control transition might cause automation surprises and bring distrust in automation (Bainbridge 1983; Abbink et al. 2008). An effective cooperation between human and systems can be used in resolving these problems caused by increasingly automated and authoritarian systems. Such cooperative human–machine relation is enabled by the concept of human-centered automation (HCA), which regards the human as the main element in the system (Billings 1997).

In this paper, two types of adaptive automation are distinguished to improve human–machine cooperation (Sheridan 1992): adaptive sharing of control and adaptive trading of control. In adaptive sharing of control, the system intervention gradually increases, but the level of automation remains the same and the human is always in charge of the overall control. For example, the system may compensate, increase

or decrease, the steering wheel friction torque and angle to ease the vehicle lateral control for the driver depending on internal factors, such as vehicle speed and steering angle input by the driver, and external factors, such as road curvature and spacing (e.g., critical or noncritical clearance distance to other vehicles). In adaptive trading of control, while the system intervention gradually increases, the level of automation changes accordingly and the human may not be in charge of the overall control. For example, a front collision avoidance system that is capable of autonomous steering and braking maneuvers may increase steering angle and/or brake pressure when driver's reaction is not enough to avoid front obstacles. Thus far, the driver remains in charge of the steering wheel and pedals. In case the driver fails to response appropriately, the system takes over the control of the steering and/or pedals and drives the car autonomously, at least for a short period of time, to avoid critical conditions. In this case, the driver might not be able to override the system. Adaptive automation has been found to be efficient and accepted more by the drivers comparing to other strict forms of automation. However, the questions as to when and how the control can be shared between the driver and the system, when authority must be handed over to either agent, and who is in charge of control distribution and authority transition strategy, still remain open.

This paper attempts to address these questions by reviewing and discussing research in automation systems, and providing a useful framework for designing automotive automation systems addressing human factors issues. The paper also aims to explain and discuss the meaning of the concept of HCA, which claims that the human must have the authority to intervene in the process of automation when necessary. First, it is essential to understand the types of automation functions and levels of automation to clarify how automation systems support humans. Second, control authority is defined in terms of levels of automation and the concept of HCA. Sharing and trading of control are defined and clarified with examples to understand modes of control and collaborations between human and automation. Finally, a set of design concepts and evaluation criteria are proposed for the future design of stable dynamic function allocation. The present study makes several noteworthy contributions to the knowledge and development of human–automation systems particularly in the automotive domain. It also clarifies the role of humans and automation in safety–critical situations and how to determine the final decision-making authority.

2 Types of automation functions

Human information processing refers to the cognitive process of information acquisition, and perception and manipulation of information to reach decisions and implement action accordingly (Baddeley 2012; Wickens et al. 2015). Figure 1 shows stages of human information processing (information acquisition, information analysis, decision and action selection and action implementation) and how various functions of automotive automation can be deployed to support each stage (Parasuraman et al. 2000; Inagaki 2011). The first stage of the model is related to the driver's ability to collect information from the vehicle, roadway and traffic environment. Advanced electronic technologies, such as night vision, and backward and side on-board cameras can be employed to enhance and extend driver's perception capabilities (Luo et al. 2010). The second stage deals with the driver's cognitive abilities in which a system may alert drivers' attention to potential risks in their vehicles, roadway and traffic environment. Pedestrians, intersections and traffic-light detection systems are typical examples of an assistance system that draws drivers' attention to potential risks in their path (Sotelo et al. 2006). In the third stage, the system encourages drivers to select the most appropriate action in the situation. For example, the system sets off a warning when the system detects an imminent collision, or provides a haptic guidance to help the driver keep in lane (Ho et al. 2006; Mars et al. 2014). Thus far, safety–critical decision-making and action implementation need to be done by the human driver.

In the fourth stage (action implementation), the system is designed to provide an automatic driving action. This action can be continuous driving assistance, such as ACC and LKA, or critical assistance, such as collision avoidance systems. SAE J3016 (2016) has defined six levels of driving automation (LDA) depending on driver's role and automation

Fig. 1 Human information processing model and roles of automation systems (Parasuraman et al. 2000; Inagaki 2011)
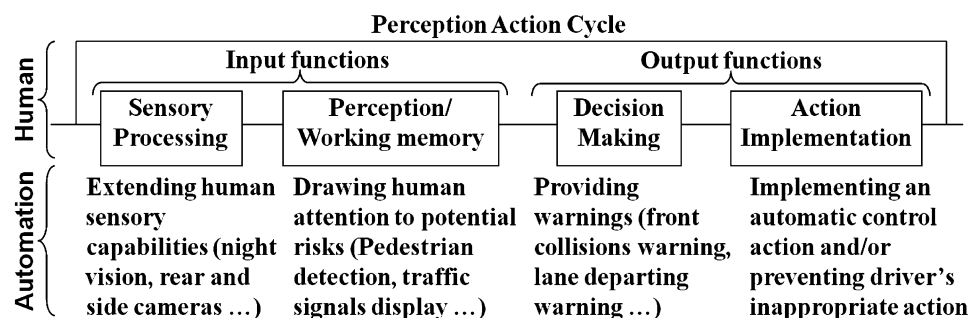
system functionalities and limitations. LDA mainly describe system capabilities and driving tasks undertaken by the system starting from no automation (level 0) where the human driver must perform all driving tasks manually to full driving automation (level 5) in which the system performs all driving tasks under all roadway and environmental conditions. However, LDA does not specify the drivers' ability to manage or regain the automatic control when they want to resume manual driving or in case of automation failure. The way in which the control is allocated between the human and the automation system depends on automation authority and type of action automation function. The design of automation function that supports human action implementation can be consistent or inconsistent with the concept of HCA in which the human has the final authority over the automation (Woods 1989; Billings 1997). Thus, DAS may automatically act in the situation or engage automatic driving mode even without the driver's directive or intervention. Three types of action automation functions are distinguished in the literature: compensation, prevention and relief (Inagaki and Sheridan 2012).

2.1 Compensation

Compensation includes systems that perform necessary actions that have not been done by the driver, such as automatic emergency braking systems (AEBS). The AEBS may have the ability to avoid or mitigate collision damages by applying an adaptive amount of brakes depending on the situation even when a driver fails to take any action (Coelingh et al. 2010). Another example of compensation can be an automatic lane change system (ALCS) that can detect and perform necessary lane change/merge maneuvers automatically (Kanaris et al. 2001). Such types of automation should be designed with care. A human can be surprised with automatic action, which may cause distrust in automation, especially when a mismatch occurs between the human intention and automation understanding of the situation (Inagaki et al. 2007). This may also occur when there is a lack of mutual understanding between human and automation (Muslim and Itoh 2017). However, even with an appropriate mutual understanding, the likelihood of automation complacency occurrence can be increased with the existence of automatic compensation systems. These conflicts can be minimized by encouraging cooperation between human and automation (Flemisch et al. 2008). Human–automation cooperation may refer to each agent's ability to control a process and cooperate with other concerned agents.

2.2 Prevention

A prevention system attempts to avoid or prevent drivers' inappropriate actions. A notable example of prevention is lane change collision avoidance systems (LCAS) that aims to avoid collisions with vehicles in the adjacent lane area. LCAS has a different objective than ALCS. Whilst ALCS supports the implementation of lane changing, LCAS prevents critical lane changing. Figure 2 illustrates a case where there are two vehicles, V_I and V_{II}, traveling on two different adjacent lanes at the same direction with a critical distance between them. Several accidents have been reported where the V_I's drivers decide to change lanes without realizing the presence of V_{II} or misjudge the speed and distance of V_{II} (Salvucci and Liu 2002). Researchers proposed several types of driver support systems to prevent such accidents. Itoh and Inagaki (2014) proposed two different steering intervention methodologies. One is to increase the steering wheel friction torque to resist the lane change initiation by the drivers, 'soft protection'. The other is to override the steering input by the drivers and provide a semi-autonomous driving, for a short period of time, to avoid the collision, 'hard protection'. In terms of safety, the hard protection has been found to be more efficient. However, the soft protection has been found to be accepted more by the drivers. This difference might come from the fact that the soft protection is compatible with the concept of HCA, whereas the drivers has the final authority over the system. In contrast, the hard protection system has the final authority over the driver. Given that the main goal of HCA is to provide more cooperative automation approaches (Billings 1997), it is difficult to judge whether the hard protection violates the concept of HCA. When the hard protection was activated, the steering was controlled by the system and the acceleration/deceleration was controlled by the driver. Thus, both driver and system cooperate to control the overall vehicle dynamics.

2.3 Relief

Relief automation assistance provides a continuous driving support during noncritical conditions, such as routine driving on a highway or limited access freeway, to reduce drivers' burden by assisting drivers' action for one or more driving functions, steering and/or headway and speed maintenance. However, the driver might need to supervise the operation of the system and the driving environment
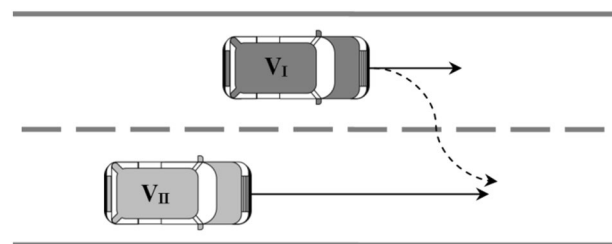

Fig. 2 An example of critical lane change scenario

depending on LDA requirements. This can be exemplified by the function undertaken by the ACC systems. An ACC system can maintain the host vehicle's speed and headway time to a preceding vehicle. The system aims to reduce drivers' workload by freeing them from frequent manual acceleration and deceleration (Corona and Schutter 2008). An automatic LKA that is capable of autonomous steering maintenance is another example of a relief control function. Such driver support systems are usually compatible with the concept of HCA. However, the main concern is how drivers might adapt to these systems in real driving. Integrating ACC and LKA systems can relieve the driver from all driving tasks in some driving scenarios expecting that the driver must intervene when necessary. The long-term interaction with such kind of assistance systems without expecting a system failure possibility or experiencing limitations of the system can lead to some critical human factors issues, such as overtrust and automation complacency. It might also lead to overreliance on the system, which can encourage drivers to engage more in non-driving related tasks (Llaneras et al. 2015).

3 Automation authority

Suppose an assistance system detects a driver's erroneous action, such as changing lanes toward a busy traffic stream, or an imminent front collision that requires the driver's immediate action. Such cases raise a number of questions. How can the automation effectively support car drivers to avoid hazardous situations? Which is better: warning the driver or executing an automatic action? For the automation to be able to determine how to support drivers effectively is dependent on the system ability to monitor and identify driver's status and state (the stages of information processing) and to what extent the authority of act is given to the system (Inagaki 2011).

Automation authority is generally defined as the delegating authority to act in a certain way from a human agent

to the automation system (Sheridan 1992). This study uses levels of automation (LOA) shown in Table 1 (Sheridan 1992; Inagaki et al. 1998), LDA (SAE 2016), and stages of information processing (Parasuraman et al. 2000) to discuss automation authority from two perspectives. The first is the extent of authority given to the system to act with or without human permission. The second is related to the extent of authority given to the human to ignore, modify or override the automation action. Supporting stages #1 and #2 of human information processing (information acquisition and analysis) requires LOA to be more than one. Whilst, the LOA for stage #3 (decision-making) should be more than three and for stage #4 (action implementation) is more than four. At LOA #5, the system initiates and implements an action only when directed by the human. Thus, the control authority is traded from human to the system in human-initiated manner. If the LOA is more than five, the system can act even without human approval, which is known as machine-initiated manner. Initiating an automatic action without human directive does not necessarily mean that the system has final authority over the human. For example, assume an ACC system is activated automatically when the host vehicle exceeds a certain speed limit and there is a preceding vehicle to follow. The system can autonomously slow down and speed up the vehicle according to traffic without driver's directive and intervention. Another example can be pre-crash safety systems, which apply the brakes automatically to avoid front collisions when the driver fails to hit the brake pedal. In some time-critical contexts in automotive domain, the need for automation intervention without human permission is indispensable for attaining safety (Inagaki 2006). In both examples, the control is traded from the human to the system in machine-initiated trading of authority. However, the drivers may override the ACC system and resume the vehicle cruise control, whilst they may not be able to override the automatic brake applied by the pre-crash safety system. It should be noted that the LOA describes the delegated authority of decision-making and

Table 1 Levels of automation (LOA) (Sheridan 1992; Inagaki et al. 1998)

Scale	Automation function
1	The computer offers no assistance, human must do it all
2	The computer offers a complete set of action alternatives, and
3	Narrows the selection down to a few, or
4	Suggests one, and
5	Executes that suggestion if the human approves, or
6	Allows the human a restricted time to veto before automatic execution, or
6.5	Executes automatically after telling the human what it is going to do, or
7	Executes automatically, then necessarily informs humans, or
8	Informs him after execution only if he asks, or
9	Informs him after execution if it, the computer, decides to
10	The computer decides everything and acts autonomously ignoring the human

action implementation, while the LDA describes machine abilities and responsibilities in the automotive domain only. Levels of automation and driving automation can specify the extent of authority given to the system to act with or without human directive, but they cannot specify the extent of authority given to the human to ignore, modify or override the automation action. Specifically, LOA and LDA may not exactly determine the final authority agent.

The second perspective of automation authority is related to the system design concept. If an automated system is designed in such a way that the human may override or change the automatic control, the system then is compatible with HCA. In contrast, if the human is incapable of managing the control action, the system design can be regarded as machine-centered automation in which the system remains the final authority and act without human directive. Whether or not the human is the final authority may not be specified by LOA only. When the LOA is six or above, the control can be allocated between the human and the machine following different strategies in which the human may or may not be able to regain or override the automatic control.

HCA that regards the human as the main acting agent in the system and maintains the human as the final authority is a widely accepted concept for automated system design (Billings 1997). According to this concept, the system may extend driver abilities of perception, support drivers' situation awareness, and select and execute an action under human directive in such a way that the driver has the authority to monitor and supervise automation and intervene when necessary. Whilst the effectiveness and human acceptance of these systems have been demonstrated repeatedly, safe task achievement cannot be assured because the human, as the final authority, may not perform well, such as ignore or override the system. Several longitudinal studies have argued that determining the final authority agent, i.e., human or automation, can be context-dependent (Moray et al. 2000; Wilson and Russell 2007; Miller and Parasuraman 2007). This means that in some safety–critical situations, automation may be given authority to act as an active and final control authority agent.

Authoritarian DAS might lead to trigger several negative human factor-related issues, such as human being out-of-the-loop performance problems, increase in the likelihood of human misunderstanding of the system, and automation complacency (Merat and Lee 2012; Endsley and Kiris 1995). These problems might come from a shift in the role of human drivers from that one direct control of the vehicle to one of monitoring and supervising the system and driving environment (Itoh 2012; Blaschke et al. 2009). It has been widely discussed that the negative consequences of automated driving may be addressed by keeping the human driver always in the direct control loop which can be achieved by sharing instead of trading the control between

humans and machines (Pacaux-Lemoine and Itoh 2015; Griffiths and Gillespie 2005). One of the most promising solutions to avoid substantive human factor issues associated with traded control are automotive shared control systems, which are also expected to increase safety and reduce human error (Abbink et al. 2012; Mars et al. 2014). However, traded control systems are more efficient at preventing human errors and may perform better than shared control systems when designed appropriately. The question that remains is how the characteristics of shared and traded control can be integrated in one system to combine advantages and eliminate disadvantages of both control modes.

4 Automation and human role of control

The roles of control in human–machine systems can be categorized into seven modes based on modes of control proposed by Sheridan (1992) and in terms of LDA by SAE (2016) as shown in Fig. 3. The first mode is manual control in which the control is determined by the human to perform the task entirely without system intervention. The system in this mode may extend drivers' perception, such as night vision and blind spot sensors, and support drivers' decision-making through some visual, audible and tactile warning systems. In the second mode, the control of a specific task is determined by the human with the assistance of the system to enhance human performance, such as haptic lane keeping assistance in which both the human and the system contribute to steering wheel control. Other systems that amplify or extend drivers abilities, such as electronic stability control (ESC) and the antilock brake system (ABS) can also be considered under the second mode of control. The first and second control modes are equivalent to level zero of driving automation in which the driver performs all aspects of the dynamic driving task, even when warning and intervention systems are available. In the third mode, some parts of the task can be implemented automatically by the system, while the human performs the remaining parts. This mode is equivalent to level one of driving automation in which ADS executes either steering or acceleration/deceleration using information about the driving environment and with the expectation that the driver performs all remaining aspects of the dynamic driving task. For example, the ACC system performs the acceleration/deceleration tasks and the driver performs the steering task. In the fourth control mode, which is equivalent to level two of driving automation, the control of task can be traded entirely between the human and the system depending on task requirement and agent ability. In automobile, one or more ADS execute both steering and acceleration/deceleration using information about the driving environment and with the expectation that the driver performs all remaining aspects of the dynamic driving task.
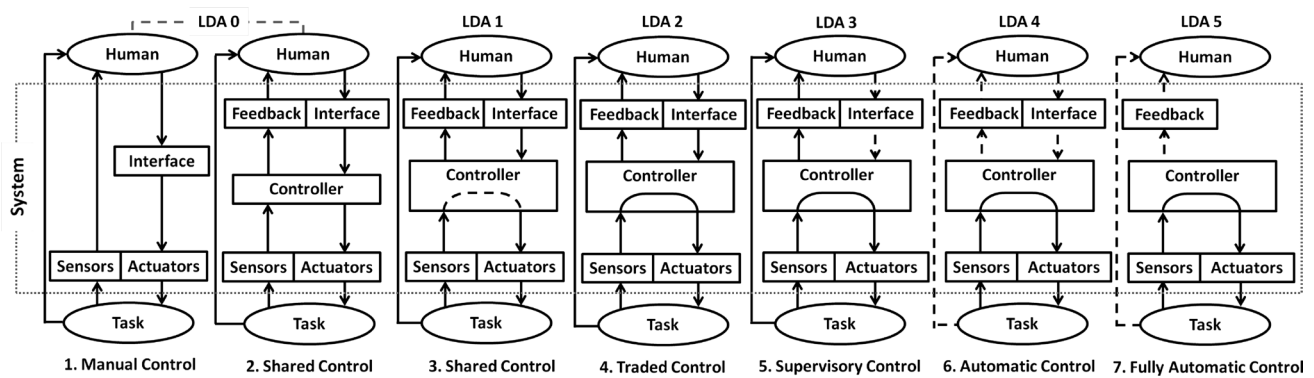
Fig. 3 The spectrum of control modes based on control modes proposed by Sheridan (1992) and levels of driving automation (SAE 2016). When the controller loop with a broken line is open, the task is performed by the human only, and when it is closed, the task is performed by human and system. When the controller loop with a continuous line is closed, the system performs the task entirely, and when it is closed, the task is performed by the human with or with automation assistance

For example, a vehicle is equipped with a limited ability system that is capable of steering and headway maintenance on highways. The system of the fifth control mode can perform the task entirely under all condition with the expectation of human intervention to achieve an adequate performance in the overall task when the system reaches a functional limit or failure. The fifth mode is equivalent to level three of driving automation in which the system performs all aspects of the dynamic driving task with the expectation that the driver will respond appropriately to a request to intervene. The sixth mode is an automatic control in which the task is supposed to be performed autonomously by the system without the need for human supervision or intervention. However, in complex and unpredictable environments like driving, the ability of the current systems is limited and their absolute reliability is not assured; therefore, human intervention might be needed to perform tasks for which the system is not designed. Such automatic control mode is equivalent to the level four of driving automation in which the system performs all aspects of the dynamic driving task, even if the driver does not respond appropriately to a request to intervene. The seventh control mode is equivalent to the level five of driving automation in which the system performs all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver. The system in this mode may have very limited interface, e.g., only display. Therefore, the human may not be able to intervene in the process.

Driving task can be divided into two main subtasks: (1) lateral control (LAC), which can be controlled with the steering wheel; and (2) longitudinal control (LOC), which can be controlled with gas and brake pedals. In traditional manual driving (control modes #1 and #2), a driver takes responsibility for controlling both the LAC and LOC subtasks during all times and under all conditions. Whilst the

driver can, from time to time, assume supervisory role with respect to one subtask and direct control of the remaining subtask in mode #3, the control of both subtasks can be traded between agents at the same time in mode #4. In automated driving modes #5–7, both subtasks, LAC and LOC, are autonomously performed by the system even without human directives. Following Sheridan (1992), modes #2 and #3 are called shared control systems in which the task is controlled by the human and automation simultaneously. When the task is determined entirely by either agent, human alone or automation alone, the system is called traded control just like control mode #4. Thus, shared control can be defined as a control distribution strategy in which both human and automation act simultaneously to determine the final output of the system. The most advantageous feature of shared control systems is that the human is always engaged in the process. Shared control systems are compatible with HCA, while some traded control systems might go beyond the boundaries of HCA. However, traded control systems represent the building blocks for future autonomous driving systems. Hence, shared control is central to this study to overcome human factors issues associated with traded control systems, particularly during authority and control transitions initiated by changes in either agent's ability (Flemisch et al. 2012). The literature identifies three types of shared control strategies (Sheridan 1992):

4.1 Extension

This type of shared control system aims to extend human ability either by reducing efforts required to perform a function, or by weighing up human input to cope with the situation. Figure 4 illustrates how the human act through the machine in this type of shared control systems. With respect to vehicle control functions, ADS may extend drivers'
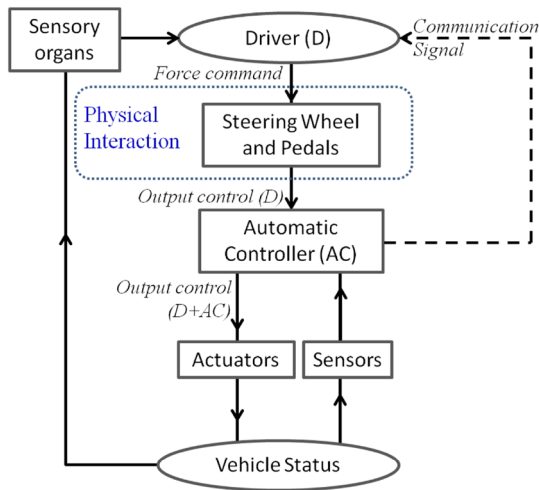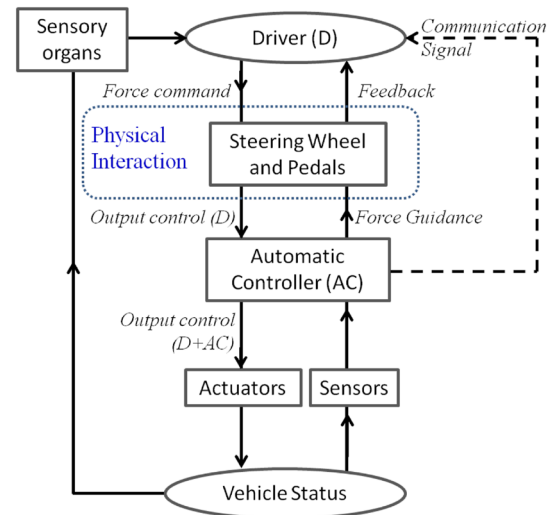
Fig. 4 Block diagram of shared control systems for amplifying and extending driver's action. The arrow with broken line indicates that the system may or may not provide the driver with a feedback about the system operation

abilities to control one or both driving functions, LAC and LOC, by generating a control action in consistence with the driver's intention. For example, a power steering system reduces the driver's effort to steer the vehicle. Recently, developed front collision avoidance systems (FCAS) may increase the driver's steering and/or braking input when the system detects that the control force by the driver is not enough to avoid the collision (Itoh et al. 2013). Thus, the human and the system contribute simultaneously and continuously to achieve a task that is difficult or critical to be achieved by the human alone.

4.2 Relief

This type of shared control systems generates a control force to guide the human action. The system attempts to assist the human in risky or potentially hazardous situations. For example, the haptic lane keeping assistance system can provide a continuous haptic guidance to the drivers to correct vehicle trajectory (Abbink et al. 2012; Mars et al. 2014). In this type of shared control, both human and system contribute to the same device input and the final output is usually determined by both agents as shown in Fig. 5. The main difference between extension and relief systems is that with extension, the system generates an assistance following the human's objective, whilst with relief type, the human and the system might have two opposite objectives. Thus, in relief type, the overall control of a task might depend on control distribution strategy between agents. With haptic lane keeping assistance, the human, as final authority, may override the control guidance of the system and proceed with lane departing.

Fig. 5 Block diagram of shared control systems for correcting driver's action

4.3 Partitioning

This type means that the system performs some parts of the task entirely while the human performs the remaining parts. Partitioning shared control is just like dividing the main task into complementary subtasks, e.g., LAC and LOC, so that the control of each subtask can be traded independently between human and automation system. Figure 6 shows control repartition of driving task between human and automation. A notable example of partitioning is found in ACC systems. On highways or long trips, drivers may let an ACC system maintain the vehicle speed, i.e., acceleration and deceleration, while they maintain the steering function.
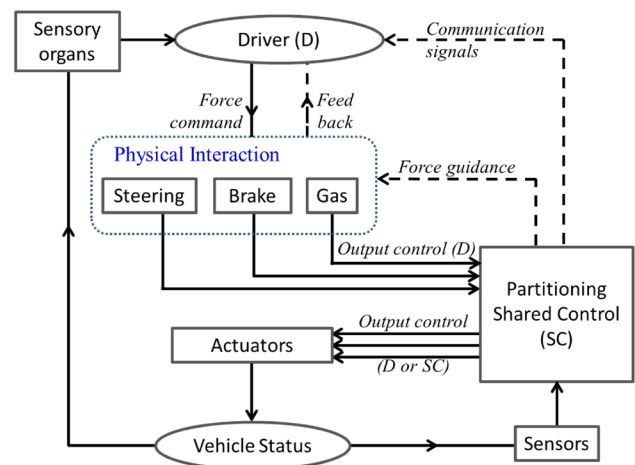
Fig. 6 Block diagram of partitioning shared control in which the task is divided into subtasks, and each subtask is equivalent to a specific function that can be traded between the human and automation system

Partitioning can be used not only to reduce human workload, but also to avoid inappropriate human actions. For example, a lane change collision avoidance system (LCAS) may take over the steering control to avoid a possible collision caused by a critical lane change initiation meanwhile the vehicle speed is controlled by the driver (Itoh and Inagaki 2014). Although the human continues to contribute to the main task by controlling other subtasks, partitioning may not be always compatible with HCA. In the former example with ACC, whether the longitudinal control is handed over to the system in human-initiated or machine-initiated manner, the driver may, at any time, regain the control of the LOC function. Such partitioning, which is usually used in routine (noncritical) driving tasks, is compatible with HCA. The latter example with LCAS, which is used to support drivers in critical situations, the LAC function is traded from the driver to the system in a machine-initiated manner and the driver may not be able to override the system. However, the overall task is determined by both agents. Therefore, such partitioning may violate some objectives of HCA.

5 Design aspects for human–automation interactions

This study recommends that future design of human–automation interactions need to take into consideration human and system information processing limitations and abilities, risk value, time criticality, and all potentially surrounding hazards and uncertainties. In what follows, four aspects for improving the future design of human–automation interactions are proposed.

5.1 Mutual understanding

Automation should be able to perceive driver status in comparison with traffic condition to understand driver behavior in a certain situation. At the same time, the system should be designed in a way so that the human driver may easily perceive the goals and capabilities of the automation system. Mutual understanding is a fundamental essence of the cooperation and positive interactions between the human and automation. Accurate understanding may help the human to reach an appropriate level of trust in automation and reduce automation-induced complacency.
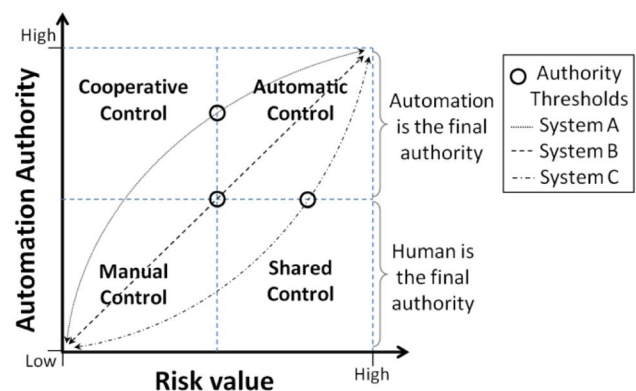
5.2 Control authority

It is recommended that automation system should be designed in a way that enables smooth control and authority transitions between human and automation. The control and authority transitions can be determined depending on agents' abilities, limitations and risk factors as shown in

Fig. 7. Where the authority is transmitted from the human to the automation is a point known as the authority threshold. Under this point, the system is compatible with HCA even when the control is traded from the human to the system. Beyond the point of authority threshold, it is difficult for the human to regain the control from the system. For example, a given system 'A' can be designed so that even when the control is handed over completely to the system, the human retains the final authority over the system. In a second system 'B', the control is transmitted from the human to the system in machine-initiated manner and the system becomes the final authority over the human. A third system 'C' comprises shared control as an intermediate stage in which the control of a specific task can be dynamically distributed between agents before reaching the authority threshold. In this case, adaptive sharing and trading of control are combined for a cooperative and smooth handing over and regaining the control by the human. The issue of determining the authority threshold is an intriguing one which could be usefully explored in the further research.

5.3 Time criticality

All safety critical situations are about timing. Timing determines how risky the situation is and when it is necessary for automation system to intervene as well as the means of interference. In Fig. 7, time to contact is essential to determine the authority threshold and when the human is unable to handle the situation encountered. Time criticality can be the main factor that makes HCA a domain-dependent concept. In some cases of automobile, machine-initiated trading of authority is accepted more than in aircraft.

Fig. 7 Evaluation criteria for authority and control transitions. Examples of systems with different strategies of control and authority transition are also shown

5.4 Risk value

Selection and implementation of an autonomous action requires high information collection and analysis of performance abilities. The system should be able to collect and analyze all surrounding information when it is about to perform an action. For example, when the system autonomously controls the vehicle to avoid a potential hazard, the system should take into account not only the detected hazard, but also how to safely guide the vehicle in the new situation during and after avoiding that hazard.

6 Application examples

6.1 Operational driving

Human performance of a task can vary depending on several factors, such as task complexity and requirements, working environment, and workload. During operational (noncritical) driving, driver performance can be subject to driver workload, road type, driving requirements, timing, and traffic status. For example, even operational driving can be critical when the driver's workload is high (Kaber and Endsley 2003). Assume that an automation system can determine when the driver performance does not fit with the situation and system intervention is needed. The question is how to support the drivers without affecting their driving performance and triggering undesired interactions with automation. Following the proposed design aspects, the system should be designed in such a way that the human can understand the operation of the system and the system can understand human behavior (mutual understanding). To understand human behavior and predict the level of performance, the system should be able to perceive the main task and divide it into subtasks (e.g., steering, acceleration and deceleration) to compare human performance of each subtask with task complexity, such as traffic status and visibility conditions, and requirements, such as speed limits and safety distance to other vehicles, in the given situation.

How the system can support human performance during noncritical situations (operational driving) using the characteristics of sharing and trading of control can be exemplified in Fig. 8. Suppose the system has the ability of controlling the LAC and LOC functions autonomously and is able to identify driver performance (within or under the required level for safe task achievement). First, the system structures the main driving task into subtasks, steering and pedals control. Second, the system evaluates driver performance of each subtask and compares it with the required level of performance in the given situation. The system creates a model of performance criterion that can be updated depending on the evaluation criteria shown in Fig. 7 above.
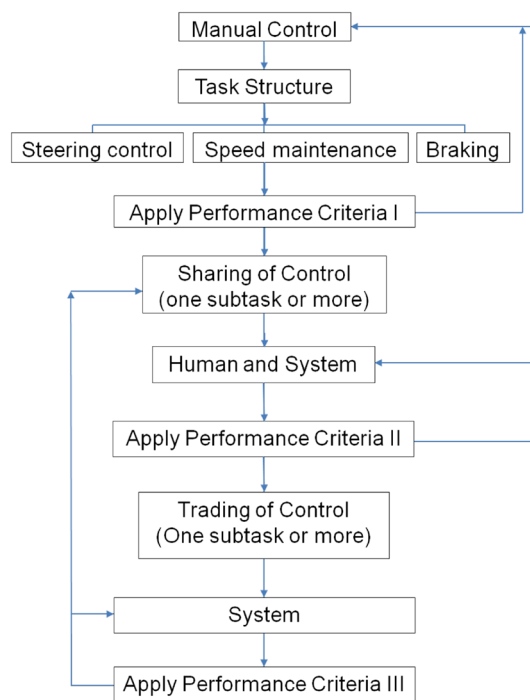

Fig. 8 Flowchart of authority and control transitions of ADS in performance and situation dependent manner

Based on the criterion, the system may decide that driver performance of each subtask is either well or needs to be automated. If the system recognizes that the driver control input of any subtask is inappropriate for the situation, such as the driver being unable to keep in lane, keep safe headway distance or inappropriate and rapid braking reaction, the system, first, provides control guidance to assist the driver cope with the situation. The aim of the control guidance is to adjust driver's performance and alert the driver that his or her performance is lower than the required level. The force feedback guidance may gradually increase or decrease based on the driver's cooperation with the system. So far, human and machine share vehicle control and the human is in charge of the entire driving task. In the case that the control guidance was not enough to improve driver performance, the system may increase automation intervention to perform one subtask or more autonomously. This means that a task or subtask can be automatically traded from the human to the system.

Machine performance is also evaluated by the system and the driver. During the automatic operation of the system, the driver may resume the control of any automated subtask by starting to input control actions via the specified interface i.e., steering wheel, and gas and brake pedals. Upon the driver's request for control or the system reaches a functional limitation, the system hands over the control of each subtask gradually to the driver. To avoid human-out-of-the-loop
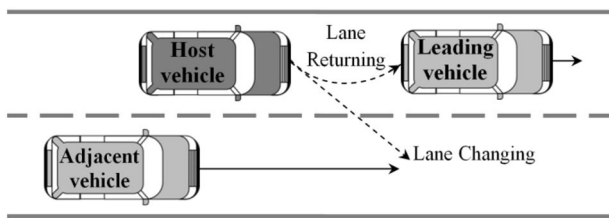
Fig. 9 An example of a critical lane change scenario
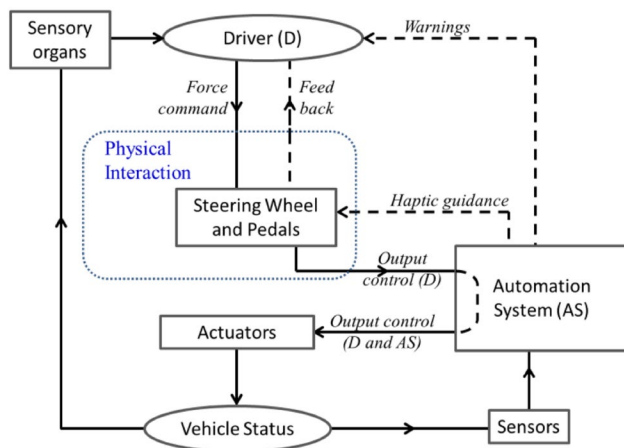
Fig. 10 Diagram of cooperative adaptive collision avoidance system; the system provides an adaptive support action based on driver performance and risk value, such as TTC between vehicles, during critical maneuvers

performance problems, the control should be shifted gradually to the driver with respect to driver readiness. As can be seen in the flowchart, it would be better not to trade the control directly between the human and the system.

6.2 Critical condition

Suppose that a driver has to make an unexpected lane changing maneuvre to avoid a rear-end collision with a vehicle ahead and dangerously closes in on another vehicle located in the adjacent lane area as shown in Fig. 9. Proceeding with lane change initiation by the host vehicle's driver without realizing the vehicle in the adjacent lane may lead to a side-impact collision. In case the host vehicle's driver could realize the adjacent vehicle and return to the initial lane, the driver needs to respond appropriately to avoid colliding with the vehicle ahead. To support the driver in such complex and time critical scenario, the system should have the ability to evaluate the situation by detecting surrounding hazards, driver reaction and compare risk values to provide the driver with an efficient collision avoidance assistance.

Figure 10 shows the proposed design of automation system in which the driver can act through the system in

cooperative manner that is compatible with HCA. The system provides assistance depending on driver's input and risk value, such as distance and time to collisions between vehicles. However, the human, when supported by automation under the concept of HCA, may not perform well as recognized by Sheridan (1995). Thus, the control of a task can be shifted partially or entirely from the human to system for safe task achievement. An effective cooperative human–machine relationship does not necessarily mean that the automation should always agree with human action to be consistent with HCA. It is necessary for the system to provide the driver with an adequate and continuous feedback to avoid misunderstanding of automation boundaries and limitations, and misinterpretation of automation elements.

7 Conclusions

The main goal of the current study was to identify and address issues in human–automation interactions that have negative impacts on human performance, efficiency of automation systems and overall safety. The study has presented roles of human and automation in terms of authority and control transitions strategies with costs and benefits of each strategy. In terms of the proposed design aspects and discussed examples, four recommendations for future research are suggested to improve human–automation interactions:

1. The system should provide human with proper and continuous feedback to understand the automation action in the given situation. An appropriate humans' understanding of the system encourages them to accept and cooperate with the automation action to improve performance for safe task achievement. In order for the system to be able to determine the appropriate action for the situation, the system should be able to compare human reaction and behavior with the situation encountered

2. Human perception of risk can be negatively influenced by the automation assistance. The existence of automation assistance seems to encourage risk tradeoff and complacency by influencing the tendency of human to check their environment, such as drivers may become reluctant to check the driving environment when supported by ADS. The chance of automation abuse can be reduced by designing a system with clear boundaries that make it easy for the human to understand how the system perceives a situation, makes a decision and implements an action. However, automation-related complacency, which has long been reported as a leading factor in aviation accidents, cannot be avoided even with well-designed systems and skilled operators (Wiener 1981). System designers, therefore, must not only design a good system, but also propose the most effec-

tive systematic approach for training the human operator to use the system as expected.

3. Humans' ability to avoid hazards that are outside the system design capacity can be affected considerably by the authority and control transition strategy. Human-centered automation is a cooperative approach to reduce these conflicts between human intention and automation limitations, which influence human performance as an automation backup in case of failure.

4. The concept of human-centered automation could be viewed from two perspectives. One could focus on how the human interact with, trust and accept the automation. This issue is important for the development of adaptive automation for critical driving support systems. The other perspective could focus on investigating whether the previously shown benefits of adaptive shared control are still present, when the control is traded from human to the system and automation is assumed as the final authority agent. This is important for the development of automated driving systems during critical and noncritical driving.

The generalisability of these recommendations is subject to certain limitations. For instance, an experimental validation of the proposed design aspects and recommendations is lacking. It would be interesting to validate proposals using empirical and case studies. Although the study covered important issues of human factor in automotive automation systems under the framework of human-centered automation, further experimental investigations are needed to evaluate the long-term effects of improving human–automation understanding on human–automation interactions.

References

Abbink DA, Boer ER, Mulder M (2008) Motivation for continuous haptic gas pedal feedback to support car following. In: Intelligent vehicles symposium, 2008 IEEE, pp 283–290

Abbink DA, Mulder M, Boer ER (2012) Haptic shared control: smoothly shifting control authority?. Cogn Technol Work 14:19–28

Baddeley A (2012) Working memory: theories, models, and controversies. Annu Rev Psychol 63:1–29

Bainbridge L (1983) Ironies of automation. Automatica 19:775–780

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

Blaschke C, Breyer F, Farber B, Freyer J, Limbacher R (2009) Driver distraction based lane-keeping assistance. Transport Res Part F Traffic Psychol Behav 12(4):288–299

Coelingh E, Eidehall A, Bengtsson M (2010) Collision warning with full auto brake and pedestrian detection—a practical example of automatic emergency braking. In: Proceedings of 2010 13th international IEEE annual conference on intelligent transportation systems, Madeira Island, Portugal, pp 155–160. https://doi.org/10.1109/ITSC.2010.5625077

Corona D, Schutter BD (2008) Adaptive cruise control for a smart car: a comparison benchmark for MPC-PWA control methods. Trans Control Syst Technol 16(2):365–372

Dekker SW, Woods DD (2002) Maba-Maba or abracadabra? Progress on human–automation coordination. Cogn Technol Work 4(4):1–13

Dickmanns ED (2002) Vision for ground vehicles: history and prospects. Int J Veh Auton Syst 1(1):1–44

Dingus TA, Klauer SG, Neale VL, Petersen A, Lee SE, Sudweeks J, Knipling RR (2006) The 100-car naturalistic driving study: Phase II—results of the 100-car field experiment. Report DOT-HS-810-593. National Highway Traffic Safety Administration, Washington, D.C.

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

Flemisch FO, Kelsch J, Löper C, Schieben A, Schindler J (2008) Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation. In: de Waard D, Flemisch FO, Lorenz B, Oberheid H, Brookhuis KA (eds) Human factors for assistance and automation, no. 2008. Shaker Publishing, Maastricht, pp 1–16

Flemisch FO, Heesen M, Hesse T, Kelsch J, Schieben A, Beller J (2012) Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. Cogn Technol Work 14(1):3–18. https://doi.org/10.1007/s10111-011-0191-6

Gao J, Lee JD, Zhang Y (2006) A dynamic model of interaction between reliance on automation and cooperation in multi-operator multi-automation situations. Int J Ind Ergon 36(5):511–526

Green M (2003) Skewed view: accident investigation. Occup Health Saf Can 2003:24–29

Griffiths PG, Gillespie RB (2005) Sharing control between humans and automation using haptic interface: primary and secondary task performance benefits. Hum Factors J Hum Factors Ergon Soc 47(3):574–590

Ho WL, Cummings ML, Wang E, Tijerina L, Kochhar DS (2006) Integrating intelligent driver warning systems: effects of multiple alarms and distraction on driver performance. TRB Annual Meeting

Inagaki T (2000) Situation-adaptive autonomy for time critical takeoff decisions. Int J Modell Simul 20(2):175–180

Inagaki T (2003) Adaptive automation: sharing and trading of control. In: Hollnagel E (ed) Handbook of cognitive task design. Erlbaum, Mahwah, pp 147–169

Inagaki T (2006) Design of human–machine interactions in light of domain-dependence of human-centered automation. Cogn Technol Work 8(3):161–167

Inagaki T (2008) Smart collaboration between humans and machines based on mutual understanding. Annu Rev Control 32(2):253–261

Inagaki T (2011) To what extent may assistance systems correct and prevent 'erroneous' behaviour of the driver? In: Cacciabue PC (eds) Human modelling in assisted transportation. Springer, Milan, pp 33–41

Inagaki T, Itoh M (2013) Human's overtrust in and overreliance on advanced driver assistance systems: a theoretical framework. Int J Veh Technol 2013:8 **(Article ID 951762)**

Inagaki T, Sheridan TB (2012) Authority and responsibility in human–machine systems: probability theoretic validation of machine-initiated trading of authority. Cogn Technol Work 14(1):29–37

Inagaki T, Moray N, Itoh M (1998) Trust, self-confidence and authority in human–machine systems. In: Proceedings of the IFAC man–machine systems, pp 431–436

Inagaki T, Itoh M, Nagai Y (2007) Support by warning or by action: which is appropriate under mismatches between driver intent and traffic conditions?. IEICE Trans Fundam Electron Commun Comput Sci E90-A(11):2540–2545

Itoh M (2012) Toward overtrust-free advanced driver assistance systems. Cogn Technol Work 14(1):51–60

Itoh M, Inagaki T (2014) Design and evaluation of steering protection for avoiding collisions during a lane change. Ergonomics 57(3):361–373

Itoh M, Horikome T, Inagaki T (2013) Effectiveness and driver acceptance of a semi-autonomous forward obstacle collision avoidance system. Appl Ergon 44:756–763

Kaber DB (2017) Issues in human–automation interaction modeling: presumptive aspects of frameworks of types and levels of automation. J Cogn Eng Decis Mak 12(1):7–24

Kaber DB, Endsley MR (2003) The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theor Issues Ergon Sci 5(2):113–153

Kanaris A, Kosmatopoulos EB, Ioaunou PA (2001) Strategies and spacing requirements for lane changing and merging in automated highway systems. IEEE Trans Veh Technol 50(6):1568–1581

Katja K, Annika L, Jonas AH (2014) Tactical driving behavior with different levels of automation. IEEE Trans Intell Transp Syst 15(1):158–167

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

Lee JD, McGehee DV, Brown TL, Marshall D (2006) Effects of adaptive cruise control and alert modality on driver performance 1980. Transp Res Rec. https://doi.org/10.3141/1980-09

Llaneras RE, Salinger J, Green CA (2015) Human factors issues associated with limited ability autonomous driving systems: driver's allocation of visual attention to the forward roadway. In: Proceedings of the 7th international driving symposium on human factors in driver assessment, training, and vehicle design, 25, pp 52–55

Luo Y, Remillard J, Hoetzer D (2010) Pedestrian detection in near-infrared night vision system. Intelligent vehicles symposium (IV) IEEE2010, ISSN 1931-0587, pp 51–58

Mannering F, Washburn S, Kilareski W (2009) Principles of highway engineering and traffic analysis, 4th edn. Wiley, New York

Mars F, Deroo M, Hoc JM (2014) Analysis of human–machine cooperation when driving with different degrees of haptic shared control. IEEE Trans Haptics 7(3):324–333

Merat N, Lee JD (2012) Preface to the special section on human factors and automation in vehicles: designing highly automated vehicles with the driver in mind. Hum Factors J Hum Factors Ergon Soc 54:681–686

Miller C (2005) Using delegation as an architecture for adaptive automation. Technical Report SIFT-TR 05-TUSC-01

Miller C, Parasuraman R (2007) Designing for flexible interaction between humans and automation: delegation interfaces for supervisory control. Hum Factors 49:57–75

Moray N, Inagaki T, Itoh M (2000) Adaptive automation, trust, and self-confidence in fault management of time-critical tasks. J Exp Psychol Appl 6(1):44–58

Muslim H, Itoh M (2017) Haptic shared guidance and automatic cooperative control assistance system: performance evaluation for collision avoidance during hazardous lane changes. SICE J Control Meas Syst Integr 10(5):460–467

Norman DA (1990) The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. Human factors in hazardous situations. Philos Trans R Soc Lond Ser B Biol Sci 327(1241):585–593

Pacaux-Lemoine MP, Itoh M (2015) Towards vertical and horizontal extension of shared control concept. IEEE international conference on systems, man, and cybernetics (IEEE SMC), pp 3086–3091

Parasuraman R, Manzey DH (2010) Complacency and bias in human use of automation. Atten Integr Hum Factors 52(3):381–410

Parasuraman R, Riley V (1997) Humans and automation: use, misuse, disuse, abuse. Hum Factors 39:230–253

Parasuraman R, Bahri T, Deaton JE, Morrison JG, Barnes M (1992) Theory and design of adaptive automation in aviation systems. Progress Rep. No. NAWCADWAR-92033-60. Naval Air Development Center Aviation Division, Warminster

Parasuraman R, Molloy R, Singh IL (1993) Performance consequences of automation-induced 'complacency'. Int J Aviat Psychol 3(1):1–23

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

Prinzel LT III, Freeman FG, Scerbo MW, Mikulka PJ, Pope AT (2003) Effects of a psychophysiological system form adaptive automation on performance, workload, and the event-related potential P300 component. Hum Factors 45(4):601–613

SAE J3016 On-Road Automated Vehicle Standards Committee (2016) Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems. SAE International, Warrendale

Salvucci DD, Liu A (2002) The time course of a lane change: driver control and eye-movement behavior. Transp Res Part F 5:123–132

Sarter NB, Woods DD, Billings CE (1997) Automation surprises. In: Salvendy G (ed) Handbook of human factors and ergonomics, 2nd ed. Wiley, Oxford, pp 1926–1943

Sheridan TB (1992) Telerobotics, automation, and human supervisory control. MIT Press, Cambridge

Sheridan TB (1995) Human centered automation: oxymoron or common sense?. In: Systems, man and cybernetics, 1995. Intelligent systems for the 21st century, IEEE international conference on, vol 1, pp 823–828

Sotelo MAL, Parra I, Fernandez D, Naranjo E (2006) Pedestrian detection using SVM and multi-feature combination. Intelligent transportation systems ITS, IEEE, pp 103–108

Treat JR (1977) Tri-level study of the causes of traffic accidents: an overview of final results. In: Proceedings of American Association for Automotive Medicine Annual Conference, 21, pp 391–403

Van der Wiel DW, van Paassen MM, Mulder M, Mulder M, Abbink DA (2015) Driver adaptation to driving speed and road width: exploring parameters for designing adaptive haptic shared control. In: Systems, man, and cybernetics (SMC), 2015 IEEE international conference, pp 3060–3065

Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) Engineering psychology and human performance. Psychology Press, Portland

Wiener EL (1981) Complacency: Is the term useful for air safety. In: Proceedings of the 26th corporate aviation safety seminar, 117, pp 116–125

Wille JM, Saust F, Maurer M (2010) Stadtpilot: driving autonomously on Braunschweig's inner ring road. Intelligent vehicles symposium IV 2010 IEEE, pp 506–511

Wilson GF, Russell CA (2007) Performance enhancement in an uninhabited air vehicle task using psychophysiologically determined adaptive aiding. Hum Factors 49(6):1005–1018

Woods DD (1989) The effect of automation on the humans role: experience from non-aviation industries. Flight deck automation: promises and realities, pp 61–85, NASA CR-10036, 61–85