AGEING AND HMI

Considering older adults throughout the development process – The HCD+ approach
Michael Sengpiel, Torben Volkmann, & Nicole Jochems

Practical experiences with different ways of eliciting information on ‘soft’ user requirements for assistive technology
Pamela Lindgren, MariAnne Karlsson, & Oskar Rexfelt

The role of mental workload in determining the relation between website complexity and usability: an eye-tracking study
Giovanni Serra, Federica De Falco, Piero Maggi, Rita Forsi, Antonello Cocco, Giancarlo Gaudino, Massimo Amendola, & Francesco Di Nocera

AUTOMATION

Contribution of Industry 4.0 to the emergence of a joint cognitive and physical production system
Philippe Rauffet, Clément Guérin, Christine Chauvin, & Eric Martin

Supporting supervisory control of safety-critical systems by optimizing visualizations on monitoring displays for fast and accurate perception
Marie-Christin Harre & Andreas Lüdtke

AUTOMATION IN TRAFFIC

Prediction of take-over time demand in conditionally automated driving – results of a real world driving study
Frauke L. Berghöfer, Christian Purucker, Frederik Naujoks, Katharina Wiedemann, & Claus Marberger

Drowsiness and fatigue in conditionally automated driving – Towards an integrative framework
Jonas Radlmayr, Anna Feldhütter, Alexander Frey, Oliver Jarosch, Claus Marberger, Frederik Naujoks, Veronika Weinbeer, & Klaus Bengler

Should I stop or should I go? Evaluating Human Capabilities of Cooperative Manoeuvre Strategies on Automated Driving
Kristin Mühl, Sandrina Büsch, Franziska Babel, Philipp Hock, & Martin Baumann

Design of time-to-lane-crossing based haptic steering guidance
Sjors Oudshoorn, Sarvesh Kolekar, David Abbink, & Sebastiaan Petermeijer

Perceived Safety: a necessary precondition for successful autonomous mobility services
Jan Grippenkoven, Zoë Fassina, Alexandra König, & Annika Dreßler

Designing for Automated Vehicle and Pedestrian Communication: Perspectives on eHMIs from Older and Younger Persons
Ina Othersen, Antonia S. Conti-Kufner, André Dietrich, Philipp Maruhn, & Klaus Bengler
TRAFFIC

Cooperation through communication – Using headlight technologies to improve traffic climate
Matthias Powelleit, Susann Winkler, & Mark Vollrath
Improving cooperation in traffic – Novel HMI approaches to enhance driver-driver interactions
Adrian Haar, Andro Kleen, Martin Schmettow, & Willem B. Verwey
Short-term effects of unnecessary collision alarms on driver behaviour
Christina Kaß, Gerald Schmidt, & Wilfried Kunde
Towards an error-tolerant HMI approach to counteract registration errors in augmented reality
Vitalij Sadovitch, Michael Wittkämper, Fares Wallis, & Ina Petermann-Stock
Effects of driving scenario on subjective workload and secondary task performance
Anna Pätzold, Rami Zarife, Michael Wagner, & Josef F. Krems
When does the driver benefit from AR-information in a navigation task compared to a Head-Up Display? Results of a driving simulator study
Kassandra Bauerfeind, Julia Drüke, Lennart Bendewald, & Martin Baumann
Occurrence of motion sickness during highway and inner city drives
Elisabeth Schmidt, Birte Emmermann, Joost Venrooij, & Klaus Reinprecht
Preferences of European cyclists towards passive and active systems with audio-visual and with handlebar vibration warnings
Marco De Angelis, Federico Fraboni, Gabriele Prati, Davide Giusino, Marco Depolo, Bruna Zani, Luca Pietrantoni, Jeremy Shires, & Daniel Johnson
Considering older adults throughout the development process – The HCD+ approach

Michael Sengpiel, Torben Volkmann, & Nicole Jochems
IMIS, Universität zu Lübeck
Germany

With the demographic change, the percentage of older adults increases while information and communication technology (ICT) becomes ubiquitous and often indispensable. However, many older adults using ICT encounter usability problems, particularly if the ICT was not designed with them in mind. If older adults are considered, their participation is often limited to the evaluation of a finished product. Our approach called “human-centered design for aging” (HCD+), considers older adults’ requirements and abilities throughout the development process, adapting established HCD-methods to accommodate the participation of older adults as experts for their own age group. This approach has been tested in a research project aiming to link older adults’ life stories to historical events and appreciating their life experience. By conducting interviews, focus groups, workshops and evaluations with older adults, meta-guidelines were identified and integrated into the HCD+ approach. Following this approach, older adults can be better served by ICT, fostering their participation in society.

Introduction

HCD+ is an integrated approach to develop Information and Communications Technology (ICT) for older people (see figure 1 for an overview). As the name suggests, HCD+ is based on the Human Centered Design (HCD) process and the principles described in DIN EN ISO 9241-210 and complements it with its focus on user characteristics and their impact on the design process and results, just like DIN EN ISO 9241-210 complements other existing design concepts. HCD+ is still in its formative stage, with a growing empirical base to evaluate, confirm and select existing design practices, as well as to create new design methods and recommendations with a focus on designing for older adults.

HCD+ methods have been put into practice and further developed in the research project “History telling” (HT). HT aims to link older adults' life stories to historical events, appreciating their life experience and fostering their social integration on a web-based platform. So far, a total of 183 people (aged M = 66.6; SD = 7.5) have participated within the HCD+ phases of HT (focus group) workshops, interviews and evaluations.
HCD+

HCD+ started as age differentiated design, with the “+” indicating older adults, which were often referred to as the age group 65+. This was in line with numerous research efforts for age-appropriate design of Man-Machine-Systems (MMS) that consider age as key feature for the classification of users (e.g. Charness et al., 2016; Coelho et al., 2015; Fisk 2009; Kurniawan, 2007). Yet it became apparent that age itself was rather a meta-trait embodying different user characteristics with enormous variability (Fisk, 2009) and did not predict the quality of human technology interaction as much as more specific age-related user characteristics (e.g. Sengpiel, 2015). Focusing on such age-correlated user characteristics seemed more relevant to usability and design, as they can be addressed more directly, e.g. in ability based design (Wobbrock et al., 2018; Wobbrock et al., 2011) or more broadly, design considering relevant user characteristics of which not all are abilities (e.g. attitudes, control beliefs, see Sengpiel 2016). Thus, the “+” became an indicator for the consideration of user characteristics in the design process with a focus on older adults. Of course, other age groups (e.g. children) might require special consideration as well. This is in line with concepts like Universal Usability, claiming that improving the usability for older adults will improve it for younger people as well (Shneiderman, 2000). Here, the History Telling (HT) project promises further insights, since it aims to bring together people across generations on one platform, e.g. grandparents and their children and grandchildren.

Overall, HCD+ addresses the design process as well as the resulting product. As Figure 1 shows, the iterative process follows DIN EN ISO 9241-210 with its focus on context of use, user requirements, design and evaluation (left side), and the goal of creating a universally usable product (right side). The green circle marks the interaction between Human and Technology to achieve a task or support an activity. One measure of the quality of that interaction is the usability, and in a broader sense, the user experience. Ensuring high quality of this interaction can be very challenging and calls for thorough research and design. Fortunately, there is a large body of knowledge, methods and design principles available to work towards that goal and
HCD+ aims to build upon these with a focus on age specific changes in user characteristics over the life span, trying to incorporate them into a research based design approach that appreciates older adults as users and designers (Jochems & Sengpiel, 2016).

Humans differ in many ways, yet some characteristics are more relevant to the interaction of Human and Technology, e.g. they can predict the quality of the interaction and inform design decisions more than others. Relevant user characteristics can inform the design process (especially if it is participatory) as well as the design of the technology itself (as product or service). Here it becomes apparent, that such user characteristics include but are not limited to abilities: Let us assume, we have learned throughout our research that older adults appreciated a comfortable user testing atmosphere with „coffee and cookies“ more than younger adults, and hence we try to provide it in testing and incorporate it into our guidelines as a direct consequence for the participatory design process. This preference is not an ability, but it is a user characteristic. Likewise, user characteristics that are not abilities, such as attitudes and control beliefs can inform product design itself, thus extending the concept of „ability based design“(Wobbrock et al., 2018; Wobbrock et al., 2011).

Finally, just as older adults have the right to have products made for their abilities and preferences, they also have a right to decide not to use a product that does not fit their needs. People usually have a choice whether to accept and use a product. Ideally, they know about available alternatives to accomplish their goal or support them in their desired activity and can assess their potential risks and benefits. Yet that alone is not an easy task, as can be seen in paper titles such as „Online T & Cs longer than Shakespeare plays—who reads them“ (Parris 2012) and „Big Data and The Phantom Public: Walter Lippmann and the fallacy of data privacy self-management“ (Obar, 2015). Once they have chosen a product, they need to know how to operate it. This distinction of knowing what to use and knowing how to use it is also reflected in the terms awareness and competence as constituent parts of computer literacy as defined by Mason and McMorrow (2006).

Many design solutions for older adults address the competence issue and age-related decline in sensory, cognitive and motor abilities, and it is important to consider them in the design process. However, ageing encompasses more than decline and life experience might also lead to changes in preferences and motivation to use a product. Yet, in particular for older adults, non-use is often viewed as a deficit that needs to be corrected with training and instruction, whereas older adults might have good reasons not to use (new) technology. On the one hand, these reasons may be found in the technology itself (e.g., bad usability or privacy risks of Big Data in social media) and, on the other hand in age-driven changes of cost-benefit-functions as described in the SOC-model of successful ageing by Baltes and Baltes (1989).

The HT project aims to fit older adults’ needs by compensating age-related deficits and by playing to their strengths when tapping into their wealth of life-experience as a means of social integration. The next section describes the History telling Project as well as the methods used.
The History telling Project

Figure 2: Screenshots of the current state of the History telling project. Top left: registration process; top right: selection of a story; bottom: story presentation

The History telling (HT) project aims to empower older adults by giving them a tool to tell life stories on a digital platform and share them with other people. Thus, it has implications on the personal and the societal level. On the personal level, HT offers a possibility for reminiscence and biography work and can strengthen the contact to one’s own story and to family and friends and new people. On a societal level, HT offers the possibility to experience history, first-hand from multiple perspectives and enriched with multimedia, spreading experience-based knowledge across generations.

At the core of the HT project is an interactive social network site, on which users can record their life story, enrich it with multimedia content and embed it in a spatial and temporal context. Furthermore, stories can be shared in a family or public space, so that vivid interaction between users of the platform can emerge through the stories, providing a powerful incentive to write more stories. See figure 2 for screenshots of the current state of development.

Participatory design within History telling
For the development of History telling, it was a key goal to integrate potential future users from the very beginning in the development process. Thus, 19 interviews with older adults were conducted in the ideation phase to establish acceptance by creating
something useful and to derive the project’s roadmap for a minimal viable product before the development itself began (Volkmann et al., 2016) and to ensure that a system was to be developed that suits the needs and interests of older adults.

Following the HCD+ approach, component-based development was used to realize the roadmap step by step. With the use of modern web technologies, it was possible to develop a usable prototype first and then add new functionality component by component. This approach also promised faster development of the whole system and replaceable parts of components for better and more focused evaluation. Every new component starts a new HCD+ cycle, in which users are considered and integrated as early as possible. To accommodate the challenge of consistency, which this software development approach has to deal with (Crnković, 2003), a living style guide was developed based on the CSS framework Bootstrap and age differentiated guidelines were established (e.g. Zaphiris et al., 2005; Hodes & Lindberg, 2002). At the current state of development, we have a frontend system with the following components: (1) **Registration** to assign published stories and comments to users; an (2) **input** component to write the stories; a (3) **feedback** component to display and write comments for stories; a (4) **timeline** component to display stories. Furthermore, a (5) **speech** component to interact with an avatar and to provide speech to text input and a (6) **stimulus material** component to stimulate the users’ imagination was developed. Finally, a (7) **backend** was developed to store and provide the necessary data.

**Participants and Recruitment**

So far, 183 participants took part in the different phases of HT development (125 female, 51 male, between the age of 46 and 93 (M = 66.6; SD = 7.5). They were recruited through managers of assisted living facilities, by speaking with local groups such as the “Frauenring” (Women’s ring) and “Landfrauen” (Country women) and through personal contacts. Because of different questionnaires used to assess affinity to technology, computer literacy and technology adoption in various contexts, there is not data for all participants, but most of the participants used a computer or a smartphone at least weekly. Computer literacy was assessed for 35 older adults which showed that the computer literacy of these participants was above average regarding older adults (Sengpiel & Dittbener, 2008). Also, there was a high range of computer literacy within these participants (min=11; max=26; SD=18.6; SD = 3.1).

**Atmosphere and procedure**

Whenever possible, the participants were free to choose the location. Thus, most of the interviews and evaluations were conducted in their private homes, and workshops were mostly conducted at the university. There were a few exceptions, e.g. for the observation study, which was conducted in the rooms of an adult education centre and one workshop which was held in a church room. We provided breaks that were longer than necessary to allow more time to socialize. For some of our focus group workshops, we used a moderator within the age of the target group.

**Methods**

Different methods in all stages of the HCD+ life cycle were used to develop HT components: semi structured interviews in the *ideation* phase, observation, group
interview, focus group workshop and semi structured interview in the *analysis* phase, focus group workshops in the *design* phase and field study, task completion with think aloud, wizard of oz and semi structured interview in the *evaluation* phase. Evaluations were conducted in two loops, conducting formative evaluations with low-fidelity prototypes first and summative evaluations with refined prototypes later (see table 1).

Table 1. Number of participants in the development of History telling assigned to HCD+ phases

<table>
<thead>
<tr>
<th>Phase</th>
<th>Participants</th>
<th>Gender</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#</td>
<td>f</td>
</tr>
<tr>
<td>Ideation Phase (semi structured interview)</td>
<td>19</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>Analysis Phase (observation, group interview, focus group workshop, semi structured interview)</td>
<td>43</td>
<td>34</td>
<td>9</td>
</tr>
<tr>
<td>Design Phase (focus group workshop)</td>
<td>19</td>
<td>16</td>
<td>3</td>
</tr>
<tr>
<td>Evaluation Phase (task completion and think aloud, Wizard of)</td>
<td>102</td>
<td>63</td>
<td>32</td>
</tr>
</tbody>
</table>

**Lessons learned**

This section presents the lessons learned for HCD+ working on the HT project.

**Recruitment**

It became apparent that it is important to invite participants to activities with a fixed date as early as possible due to full schedules, a point that Lindsay et al. (2012) also indicate, although it cannot be confirmed that it is important to recruit more older adults than required. It was easier to recruit groups of people and stay in touch with the most engaged persons in these groups as they could be considered as a pull factor for other group members. In the conducted methods it worked well to recruit older adults based on mutual give and take, although monetary incentives may also work (Voorberg et al., 2015). Therefore, focus group workshops were combined for example with a technology introduction part which sometimes became a key driver to participate in the events. The recruitment process showed that only those participants attended who had a higher technology adoption than the average older adults’ population.

**Atmosphere and procedure**

Participants often appreciated activities as social events. Thus, long breaks were provided to give enough time to socialize, which was well received (see also Lindsay et al., 2012; Massimi & Baecker, 2006; Ellis & Kurniawan, 2000; Newell et
al., 2007). Participants stated the importance of a comforting atmosphere within workshop sessions and evaluations. Lindsay et al. (2012) stress that this can also enrich the creativity and overall output. Participants used the sessions to learn about technology and asked many questions, especially in the breaks. Workshop leaders were included into the social group to provide information to the participant, a practice also stated by Newell et al. (2007). Because of these special considerations and the diverse group of participants, the exact timing was difficult and nearly all methods took longer than predicted. Figure 3 shows different situations within a workshop.

Figure 3: Atmosphere within a conducted workshop. Top left: introduction to technology; top right: simulation game as part of the workshop; bottom: socializing break

Methods

Older adults are a very heterogeneous group, especially regarding technology experience and adoption. Therefore, fall-backs had to be established if technology experience was crucial, e.g. using abstract description of technology or low-technology fall-backs. Also, parts of the conducted workshop were accommodated to participants’ wishes, such as addressing the question “how does communication with emails and other software on a smartphone work”. Finally, for some tasks the researchers had to explain repeatedly the scope and reasoning for both the particular method and the technology, e.g. to address privacy concerns.

Resulting Guidelines

Findings thus far resulted in the following guidelines, which are instrumental for the development of the HT Project can be found in table 2. It is important to note that
these guidelines are still work in progress and actively developed further within the overall process.

**Table 2: Resulting guidelines for HCD+ based development**

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engage with group leaders</td>
<td>Group leaders can pull the whole group to participate in activities</td>
</tr>
<tr>
<td>emphasise reciprocity when recruiting</td>
<td>As the elderly like to help, they also like to get help. Instead of monetary incentives, technology adoption workshops often work as incentives in itself.</td>
</tr>
<tr>
<td>Plan for social engagement</td>
<td>Giving time and space to socialize is always appreciated by older adults as is a comforting atmosphere (“provide cookies and coffee”).</td>
</tr>
<tr>
<td>overestimate the scheduled time</td>
<td>It is difficult to estimate time for methods like workshops and interviews because of older peoples’ diversity</td>
</tr>
<tr>
<td>Accommodate participants’ wishes</td>
<td>Within a workshop or evaluation setting, unplanned wishes, e.g. to get more information can occur. Accommodating these wishes is much appreciated.</td>
</tr>
<tr>
<td>Establish fall-backs</td>
<td>Sometimes activities do not work out as expected, e.g. because of little technology experience and adoption. Planning for low-technology fall-backs helps.</td>
</tr>
<tr>
<td>Use abstract descriptions of technology</td>
<td>To envision tasks, abstract descriptions of technology work best to stimulate the creativity of older adults.</td>
</tr>
</tbody>
</table>

**Discussion and outlook**

As shown within the History telling Project, the HCD+ approach can guide the development of software towards universal usability for older adults and their children and grandchildren. At the same time, HT provides feedback towards improving the HCD+ methodology. For the HT Project it will be important to broaden the usage of various technologies such as new input and output technologies and to broaden the context of use to assess the experience of the HT core idea within other settings. Also, a longitudinal evaluation should be conducted to assess real usage data over time, such as number and length of postings, time spent on the website and frequency and severity of usability issues. This might lead to the development of new components and to an improved utility, usability and user experience of the History telling system, increasing its value for the individual and
considering older adults throughout the development process

the society. One such development has been to find more stimulating storytelling and -sharing environments, which has already been realized as a student project with the museum “Günter Grass Haus” in Lübeck (see figure 4) and will be described in detail in future publications.

Figure 4: History telling at the Günter Grass Haus in Lübeck, consisting of flyer (left), mobile app (middle) and website (right)

As HCD+ is still in its infancy, there are various plans to improve it continuously alongside History telling and other projects. Central to these efforts is the fast, reliable and valid assessment of user characteristics and their impact on the design of products and services and the design process itself. These user characteristics encompass abilities and skills such as eyesight and computer literacy as well as attitudes and motivational aspects such as control beliefs and preferences. One of the process related outcomes will be the continuous revision and extension of the practical guidelines described in this paper. Naturally, HCD+ will also test, evaluate and incorporate applicable methods found in the literature to converge further towards its goal of universal usability with a focus on older adults and welcomes pointers and cooperation in this direction.

Acknowledgements

We thank all those involved in the development of HCD+ and the History telling project, especially the numerous unnamed students and study participants.
Literature


considering older adults throughout the development process


Practical experiences with different ways of eliciting information on ‘soft’ user requirements for assistive technology

Pamela Lindgren, MariAnne Karlsson, & Oskar Rexfelt
Design & Human Factors, Chalmers University of Technology, Sweden

Abstract

Users do not always adopt assistive technology (AT), but abandon it since it fails to meet their needs. Some users experience stigmatisation. Historically, AT design has focused on functionality, performance and safety. However, for AT users, other less tangible, ‘softer’ areas are also important, such as emotional or social needs. The focus on the more tangible aspects of AT is probably one explanation why it often does not meet users’ needs or fit into users’ everyday lives. Therefore, AT designers must also understand users’ ‘soft’ requirements. In this project, aimed to design a user-friendly AT device for short-distance individual transfers indoors, 36 user interviews were carried out. In the interviews, pictures of AT devices were used to visualise the AT concept under development. Fifteen of the interviews included questionnaires with semantic differential scales. The practical lessons learned point to the importance of actively triggering the users to reflect on issues beyond functionality and safety, and that the use of pictures and semantic scales had an effect on the character of the data elicited. The implications are that users need support in envisaging both the product being designed and its context in order to express requirements for soft qualities.

Introduction

Over a billion people, or about 15% of the world’s population, have some form of disability and the number is increasing, partly as a result of an ageing population and the increase in chronic health conditions (The World Bank, 2018). Many individuals with disability and impairment use assistive technology (AT) in their everyday lives, and the need for and use of AT are increasing (Smith et al., 2018). One category of AT comprises aids for transport (e.g. wheelchairs) and aids which facilitate transferring a person from one place to another, for example from bed to wheelchair (e.g. sling lifts). However, there are studies arguing that users abandon certain AT solutions as they do not fulfil users’ needs (Riemer-Reiss & Wacker, 2000; Alper & Raharinirina, 2006) and because users experience stigmatisation when using them (e.g., Louise-Bender Paper et al., 2002; Parette & Scherer, 2004).

One reason may be that the dominant perspectives in AT design are still functionality and performance whereas research has shown that users’ needs and
requirements for products do not only concern functional, or instrumental, aspects. Other, ‘softer’ needs and requirements must also be considered in the development process (Mallin & de Carvalho, 2015), including aesthetic qualities (e.g., visual, haptic, etc.) (cf., e.g., Desmet & Hekkert, 2007), hedonic aspects (e.g., pleasure, satisfaction and stimulus) (cf., e.g., Lai, 1997), and social consequences of AT use (see also Jordan, 2002; Norman, 2005). The implications are that designers need to develop better insight into these aspects to fulfil users’ needs and requirements and to overcome problems of AT abandonment and user stigmatisation. However, whereas user requirements for functionality etcetera may be relatively easy to identify and formulate (as they refer to more visible and conscious requirements), the ‘soft’ requirements require another type of elicitation methodology (cf. Karlsson, 1996; Visser et al. 2005). However, knowledge is limited on how to elicit more in-depth information from users on soft values overall, and in particular when it comes to AT.

Aim

The overall aim of the work is to contribute to the understanding of how to elicit information on ‘soft’ requirements, in this particular case for AT. By ‘soft’ requirements we mean requirements concerning product aesthetics, product meaning, and hedonic aspects with consequences in terms of appeal, satisfaction, self-esteem, and identity. The main research questions are ‘How does the design of a user study affect the elicitation of ‘soft’ user requirements?’ and in particular ‘How does the choice of data collection method and participants influence the type and character of the information elicited?’

The focus of this paper is the contents and outcome of two user studies in which semi-structured interviews and semantic differential scales (SDQs) were used to collect information from different user categories in order to elicit information on ‘soft’ requirements for an AT.

Research setting

A design project was carried out with the purpose to design a contemporary AT device to facilitate short-distance individual transfers, independently and/or with support of one person. The goal was a conceptual solution allowing for enhanced independence, accessibility, and mobility, and with an aesthetically appealing design that safeguarded users’ dignity and integrity during transfer.

The design of the first concept, model A, was based on a benchmark product analysis and available knowledge in the project team (industrial design, human factors, orthopaedics, medical technology, electronic engineering and manufacturing). Keywords guiding the ideation were "light", "clean", "contemporary" and "un-engineered" in contrast to the more technological appearance which characterises traditional aids. To create these product expressions (cf. Monò, 1997), the technology was hidden and the shape of the device was chosen to mimic an “organic living thing” as opposed to sharp, square, hard surfaces resembling something machine- or man-made. Additionally, innovative functions were added to enable independency including motorised power wheels, a
kneeing/stretching function to adjust seat height and provide different seat positions (incl. sitting astride), handlebars for steering and braking and the possibility to disassemble the device for transport. A user study was completed to elicit user responses, and based on the input, the initial design was modified, resulting in a new model, model B, which was evaluated in a second user study. The user studies constitute the empirical basis for this paper.

**The user studies: overview**

The two user studies, user study 1 and user study 2, were completed in partly similar, partly different ways (Table 1).

<table>
<thead>
<tr>
<th>Study 1</th>
<th>Method</th>
<th>Participant / User group</th>
<th>Product representation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interviews</td>
<td>A, B, C, D</td>
<td>2D and 3D visual representation of</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>AT model A</td>
</tr>
<tr>
<td>Study 2</td>
<td>Interviews</td>
<td>A, C, D</td>
<td>2D and 3D visual representation of</td>
</tr>
<tr>
<td></td>
<td>and</td>
<td></td>
<td>AT model B</td>
</tr>
<tr>
<td></td>
<td>Semantic Differential Scale Questionnaire</td>
<td>A, C, D</td>
<td>2D and 3D visual representation of AT model B; Photos of different use environments</td>
</tr>
</tbody>
</table>

The recruited participants represented four different user groups in terms of their respective role in relation to and use of the AT:

A. primary users, individuals with physical impairments (all wheelchair users),  
B. relatives as caregivers,  
C. professional home health caregivers, and  
D. orthopaedics ward hospital staff.

User study 1 followed a general procedure to search for users’ needs and requirements and to evaluate the design of model A. Hence, the interviews included a brief introduction of the conceptual AT before open-ended questions were asked on topics such as accessibility and usability e.g., of power wheels, kneeing/stretching function, and different seat positions, if/how the device would enable independency, as well as the design appearance and size of the AT with the support of visual representations. A large part of the interview revolved around five typical indoor transfer situation situations: from and to i) bed, ii) wheelchair, iii) chair, iv) shower and v) toilet.

User study 2 followed a similar procedure albeit with a clearer focus on evaluating the modified concept, i.e. model B. In addition to interviews, the study included a semantic differential scale questionnaire (SDQ), which had been developed based on the interview data collected in user study 1. The purpose of including the SDQ in the second user study was to prompt all participants to reflect on and identify some of the softer values of the visualisations of the concept. The participants in study 2 were a mix of informants from study 1 and new informants, which is why the
decision was made to hand out the questionnaire after the interview. The idea was that the face-to-face interviews would help build trust between the interviewer and the interviewees, and that the participants would be able to discuss the topics before answering the more structured questionnaire. Thus, this procedure was expected to result in more informed answers.

All interviews were carried out by a researcher (first author, who was also the industrial designer creating the design concept) and took place at a location according to the preference of the respective participant. As a consequence, the environments varied between private homes, cafeterias, offices, home care units and hospital wards. The interviews lasted between 30-45 minutes. In addition to written notes, the interviews were audio-recorded and the recordings later transcribed verbatim for further analysis. The transcripts were coded into categories and themes in a process inspired by Miles’ and Huberman’s (1994) three-stage approach of data reduction, initial coding, and search for themes.

All of the participants participated voluntarily with consent and the data gathered is protected by PUL and GDPR laws.

User study 1

Method and materials

Study 1 involved face-to-face interviews with 21 participants, representing the different user groups A, B, C and D. Almost all were experienced transfer device users.

Table 2. User study 1 involved altogether 21 participants, between 30 to 83 years of age

<table>
<thead>
<tr>
<th>Category</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary users with physical disabilities</td>
<td>3 men</td>
<td>4 women</td>
</tr>
<tr>
<td>Relatives as care giver</td>
<td>2 men</td>
<td>1 women</td>
</tr>
<tr>
<td>Home care giver</td>
<td>2 men</td>
<td>6 women</td>
</tr>
<tr>
<td>Hospital nursing staff</td>
<td>2 men</td>
<td>women</td>
</tr>
<tr>
<td>Occupational therapist</td>
<td></td>
<td>1 woman</td>
</tr>
</tbody>
</table>

The interviews were supported by visualisations of the concept, model A (see Figure 2). After the introduction, the participants were asked for feedback, to describe possible problems on the concept and to reflect on and describe if and how they were likely to use the AT device as well as their thoughts on different aspects of the design, including functionality, safety, interaction, appearance, etc.
Findings

The interviews provided information on ergonomical aspects, perceived usability, and technical features as well as on perceived problems associated with safety, functionality and efficiency. For instance, the motorised wheel function to operate the device was experienced as something positive from an independence perspective (as less physical strength was needed to accomplish transfers), but also as something negative from a safety perspective. Modifications of the design and solutions perceived were also suggested, e.g. an additional tilt restraint function, requested primarily by the caregivers (group C) to prevent the device from overturning.

‘Soft’ requirements, often related to aesthetics, were expressed primarily by the primary users (i.e. group A) who wanted the device to “melt into” their homes, to “look good” and “cool” as “…most aids are so ugly…” (female user, group A) and because “…that aids should look cool and all that have always had such a low priority” (male user, group A). User groups B, C, and D primarily responded to the safety, functionality and efficiency of the device, for example transfers to and from the device and between the different areas. The user categories also differed in response to some specific design solutions. Some primary users perceived the proposed astride position (i.e. sitting with legs on each side of the seat, see Figure 3) as positive, in that they could have eye contact with potential assistants but the dominant experience was a feeling of being physically and emotional exposed especially in situations associated with personal hygiene. User groups C and D responded to this same option mainly in terms of efficiency and functionality.
Reflections

The user study was constructive and effective from a design practice perspective as a way to evaluate and get feedback on an initial design concept at a low cost. Based on the user evaluation, the design concept was modified to fit the users’ needs and requirements better (e.g. the construction was changed to prevent the astride exposure, a remote control to enhance independence and a tilt restraint function were added to prevent the device from overturning and the user sitting on the device from falling backwards).

The semi-structured interviews were supported by product representations (PRs) which were considered beneficial in the exploration of functionality and usability aspects across all user groups. They also helped focus the questions around the AT and context of use and triggered the participants to respond to the concept, with requirements and/or solutions of assumed future problems, as well as thoughts on the visual appearance of the AT. The PRs were also considered to elicit information regarding intimate situations without the participants being exposed to the physical situation, as opposed to using a physical prototype to try out functions.

Nevertheless, the user study was not considered to be fully satisfactory regarding the in-depth exploration the soft aspects of the AT. The majority of informants did not mention these aspects at all if not specifically asked, which indicates either that the aspects were not considered important and/or that the participants had difficulties in communicating their needs and wishes. A decision was therefore made to modify the second user study and introduce a semantic differential scale as a tool to support the participant in their evaluation of the modified model B.
User study 2

Method and material

User study 2 involved interviews with 15 participants (of whom some were the same individuals as in user study 1, i.e. all of the primary users) representing user groups A, C and D (Table 3). The interviews generally followed the structure of the interviews in study 1 but were supported by visualisations of the refined concept, model B (see Figure 4).

Table 3. Study 2 involved altogether 15 intended users as participants, ranging from 30 to 65 years of age

<table>
<thead>
<tr>
<th>Category</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary users with physical disabilities</td>
<td>2 men</td>
<td>3 women</td>
</tr>
<tr>
<td>Home care giver</td>
<td>2 men</td>
<td>4 women</td>
</tr>
<tr>
<td>Hospital nursing staff</td>
<td>2 men</td>
<td>1 woman</td>
</tr>
<tr>
<td>Occupational therapist</td>
<td></td>
<td>1 woman</td>
</tr>
</tbody>
</table>

Directly after the interview, all participants filled in the semantic differential scale questionnaire (SDQ) based on their impression of model B. Many of the participants commented on and explained their answers. As suggested by Wikström (2002), bipolar adjectives were chosen specifically for the product category and the target group. Hence, adjectives frequently mentioned in user study 1 were extracted and clustered according to similarity. In this way important themes could be identified. The questionnaire was organised into five different themes: themes 1-3 addressed to what degree the product communicated aesthetic, emotional and functional values (e.g. elegant – clumsy, dependent – independent), theme 4 how the participants estimated that the device would make them feel (e.g. dignified – undignified) and theme 5 how the participants perceived the AT to “fit” into three different use environments (e.g. fit in completely – not at all). These were visualised with photographs showing bathrooms with toilet and shower illustrating a private home, and a healthcare and hospital environment respectively.

Findings

As anticipated, the character of the information elicited from the interviews had large similarities with user study 1. The majority of comments focused on functionality, efficiency and safety aspects but there were some differences. For example, the primary users elaborated their requirements further to also include other use areas than the five situations presented in the interview, for instance transport, and additional requirements were posed: “…// to have something like a transport box to put it in”… (male user, group A). One explanation for their ability to describe their requirements in more detail could be the participants’ involvement in study 1 which had brough the design of ATs ‘to the surface’.
In addition, comments implied that functions were not only assessed based on functional values, but also on identity. One example concerned the tilt restraint function: “This is a chicken variant, I don’t operate like that” (male user, group A). Another comment concerned feelings assigned to the device’s appearance in a home context: “It does not have to scream ‘aid’, but if it has a ‘tougher’ design it becomes so much easier to live with it.” (male user, group A).

Overall, the responses to the questionnaire were very similar across the different user categories (Figure 5); a majority rated model B as expressing values such as “modern”, “dignified” and “practical”. Motives for the ratings were also verbalised: “Well, I think that it’s, it’s simple but also slightly futuristic almost .../... So, I don’t know if I’d use the word exclusive here.” (male user, group A).

---

**Figure 4. Visualisation of modified model B, with hand control**

**Figure 5. How the primary users and the personnel answered the SDQ**
Differences between groups mainly concerned the responses on how the AT would make them feel. Some users felt that it was just another accessibility aid: "And it’s just, like, kind of another aid". (male user group C). Others felt equipped and independent: "If a thing like that turns up then I can cope on my own" (female user group C). One participant explained that answering the interview and the SDQ had affected him emotionally: "I was touched somehow...//...The emotional aspects, they are an important thing." He also described many patients’ emotionally reactions of using transfer aids in the hospital: "Some patients think it's terrible having to sit in a wheelchair after an operation ...//...some understand that this is more or less temporary, because I will be able to walk again... // ...It's a little like that, being able to find acceptance in having to sit in one, in a wheelchair" (male user, group D).

There were also some differences in how the participants perceived the AT to fit the different use environments shown in the photos. Some participants did not think the environment mattered: "If I think that if it (the AT) fits or not in all of these (environments) doesn’t matter" (female user, group D). Other participants, in particular from group A, began to reflect on how the design would fit their own home: "So in my case, I could very well imagine having a thing like that, standing in my home without feeling any more handicapped so to speak" (male user, group A).

However, despite the interviews and SDQs, there were participants who found it difficult to express their thoughts: "What you think in this situation//...I have my values//... as a nurse and a fellow human, all these aspects //... I know what I think, but it isn’t that easy to put it into words" (male user, group D).

**Reflections**

The 2D and 3D representations in combination with the SDQ helped participants to focus on the design concept - it became evident to the participants that it was the specific concept that was the object of the evaluation and not AT in general.

Overall, the questionnaire had a considerable impact on the character of the data elicited. In comparison with the data from the interviews, the SDQ was considered more effective in gathering information on the soft factors (i.e. product expression or meaning and the resulting emotions) and from all categories of participants (in the interviews, many of the soft aspects were addressed only by the primary users). The SDQ not only provided ratings, but also triggered many of the participants to verbalise and elaborate their thoughts on the soft aspects implying that such an instrument could be used as a trigger also in a user requirement elicitation study (cf. User study 1). Nevertheless, while the SDQ data in the specific case did not result in specific design requirements, important insights and reflections were enticed by the SDQ concerning the affective connotations and emotional dimensions of an AT product. This information is vital for design practitioners to understand and relate to when making design choices, in order to e.g. develop an AT that does not contribute to user abandonment and stigmatization.
Discussion and conclusions

Methodological considerations in efficient requirements elicitation include e.g. the choice of data collection method, participants and mediating tools (stimuli) in order to enhance the data collection process (see e.g. Karlsson, 1996; Engelbrektsson et al., 2000). The importance of the choice of participants was evident in that the participants’ different roles in relation to the device (i.e., as primary users versus caregiver) and use situations (home versus workplace) affected their thoughts on soft product values per se. Primary users saw a value in this issue and expressed several needs and wishes while caregivers did not experience the same needs. Hence, the character of the information elicited varied from practical and work-oriented aspects to personal and emotional values.

The 2D and 3D images had two functions: first as product representations, i.e. representations of the new design, and second as mediating tools in eliciting information. The two AT solutions represented by 2D and 3D images supported the participants in envisaging a new AT under development regarding functionality, use purpose and context, making them suggest changes and solutions and also to express some requirements regarding soft product aspects to increase user acceptance. Earlier studies have demonstrated the impact of product representation and the type of information carried on the information elicited in user studies (Söderman, 2001). A full-scale prototype or mock-up may have resulted in additional feedback regarding e.g. the size and functionality of the aid whereas the elicitation of soft requirements might have gained more by introducing images of different design appearances, enabling comparisons of visual impressions.

If the 2D and 3D images made the participants focus on the object, the interior photos showing the different environments made them reflect on the AT in a broader perspective and context than the participants’ main use environment (e.g. hospital, home) and their role in relation to ATs (i.e., as primary user, caregiver or hospital staff). Additional photos showing other use contexts would probably have had a big impact on the information elicited as different situations play an important role in how individuals perceive usage and value a product and hence on the information elicited (cf. Correia de Barros et al. 2011).

Semantic differential scales (Osgood et al. 1957) are most often used to measure the connotative meaning of and emotional attitudes towards e.g. a product. Also in this study, the SDQ was used to acquire participants’ rating of product expressions (or meaning) as well as ratings of how the design made them feel. However, this appeared to support their awareness of the design beyond the merely practical functions, hence the SDQ also had a mediating role. For some this mediating role extended to helping them describe their thoughts on the themes included in the questionnaire in more detail as well as developing previous experience and memories. Mediation is thus essential to elicit users’ softer requirements for AT, its aesthetic and symbolic functions (cf. Mallin & de Carvalho, 2015) and the meaning assigned to AT which plays an important role in understanding the mechanism behind acceptance and rejection.
In conclusion, user studies provided information as a useful basis for design decisions regarding the new AT. However, several decisions had an impact on the character of the information elicited including the choice of participants, data collection methods, and product representations. Different user categories may, as in the case of AT, pose and/or emphasise different types of requirements for the product and these differences must be considered in the design of the user study. Furthermore, different requirements are accessible to different degrees and mediating tools are an important key in the elicitation process, especially information on ‘soft’ aspects which is more difficult to elicit than information on more instrumental product aspects. In the case of AT design, the role of product representations and mediating tools, including product representations, photos and semantics scales, is especially important as these can help the user to envisage a product and context of use, predict possible use problems, and help reveal design aspects which are essential for AT to be successfully integrated into individuals’ everyday lives and improve well-being.

References


The role of mental workload in determining the relation between website complexity and usability: an eye-tracking study

Giovanni Serra1, Federica De Falco1, Piero Maggi1, Rita Forse2, Antonello Cocco2, Giancarlo Gaudino2, Massimo Amendola2, & Francesco Di Nocera1

1Department of Psychology, Sapienza University of Rome
2Higher Institute of Communications and Information Technologies, Ministry of the Economic Development Italy

Abstract

Digital technology is now crucial for carrying out many activities and the drive for innovation in digitization has involved not only private organizations but also public administrations. However, despite a growing awareness of the importance of digitization of public services, usability issues have been addressed only recently. Terms like “Usability”, “User Experience”, and “Human-Centred Design” are surely becoming part of the vocabulary, but often they are used as empty keywords by policymakers, while there is a lack of specific research in this domain. One notable example is the lack of studies on cognitive load imposed by information abundant websites and its influence on both users’ performance and perceptions of usability. The main objective of the present study is to evaluate the mental workload of users navigating websites with different levels of complexity in their information architecture. Eye movements of twenty users were recorded during the execution of search tasks carried out on websites of three public agencies previously selected for their complexity level. Mental workload assessment was obtained by analysing subjects’ ocular behaviour. Results showed that low complexity websites were associated with better performance, lower mental workload and higher usability rates compared to high complexity websites.

Introduction

According to the European Commission (EC) “in less than a decade, most economic activities will be managed through digital ecosystems” making the digitization of all economic sectors mandatory. The European Commission has also published an “Action Plan” to guide Governments in this transition. The document establishes some basic principles to regulate and promote the digitization of Public Administrations (PAs). The Italian Government, through an internal policy document (Directive for Public Administration and Innovation of November 26, 2009, No. 8), required PAs to improve the quality and usability of their websites by applying guidelines inspired by “Usability principles”. In January 2017 the Italian Digitalization Agency (Agenzia per l’Italia Digitale: AgID) published the first
version of a web toolkit indicating style sheets, Javascript components and HTML examples to be applied for easily adhering to usability standards.

Usability has been defined by the ISO 9241-11 (updated by ISO 9241-210) as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use”. Effectiveness refers to the completeness and accuracy in the achievement of the objectives by the users, efficiency refers instead to the optimization of the user’s cognitive and temporal resources expenditure and satisfaction concerns the aspects of comfort and acceptability of the use of the product. Interface usability is closely related to another concept defined “Information Architecture” (IA), describing the structural planning of information contained within a website (Garrett, 2010, Rosenfeld & Morville, 2006). Several studies (Gullikson et al., 1999; McGills & Toms, 2001; Miller & Remington, 2004; Rosenfeld & Morville, 2006; Toms, 2002) found that highly complex information architecture is related to decreases in perceived usability. The ease of use of an interface directly affects the cognitive effort required for the interaction. Although the relationship between usability and cognitive effort can be easily deducted, the two concepts have been studied independently. Usability has been studied applying the User-Centred Design methodology with the objective to improve the Human-Machine Interaction. Instead, the construct of mental workload (MWL) has been investigated in the Human Factors research tradition to study human error and to increase systems’ security. The methodology used to measure interfaces’ usability and MWL are different but share a cognitive-behavioural approach. Eye movements have been used to study visual search strategy and have been applied both to usability studies and to measure the cognitive effort imposed by a complex task (May et al., 1990; Jacob & Karn, 2003; Pan et al., 2004; Poole & Ball, 2006; Majaranta & Bulling, 2014).

Recently, the availability of less intrusive eye-tracking systems allowed researchers to effectively use indices of ocular activity as a measure of the operator mental workload (see Van Orden et al., 2001). For example, frequency and duration of eye-blinks have been found to be inversely correlated to mental load (Brooking, Wilson, & Swain, 1996; Hankins & Wilson, 1998). Additionally, some studies (Bunecke, 1987; Ephrath et al., 1980) have shown that workload affects the duration of fixations, whereas others (Bellenkes, Wickens, & Kramer, 1997; Miller, 1973) recorded shorter and more frequent fixations in expert operators.

Studies from our research group (see for example Di Nocera, Camilli, & Terenzi, 2007 and Di Nocera, Ranvaud, & Pasquali, 2015) have introduced the use of an algorithm for the analysis of the scanpath called Nearest Neighbour Index (NNI). Using this algorithm, it is possible to compare different fixation patterns during the execution of a specific task. The result is expressed by a single value varying with fixation clustering-spreading. Also, as described by Camilli et al. (2008), the index is sensitive to the type of demand, and values increase as mental workload increases when the taskload is due to the temporal demand, while they decrease as mental workload increases when the taskload is due to the visuo-spatial demand.

Eye-tracking techniques have also been applied in usability studies (see Goldberg, 2003). Optimization of the interface features and functions has been studied both
using metrics based on scanpath elements (such as fixations and saccades) and metrics based on the analysis of the entire visual exploration path. Many authors (Fitts et al., 1950; Goldberg et al., 2002; Cowen et al., 2002; Ehmke & Wilson, 2007; Poole et al., 2005; Goldberg & Kotval, 1999; Kotval & Goldberg, 1998; Ehmke & Wilson, 2007; Habuchi et al., 2008; Wang et al., 2018) have observed that an incorrect arrangement of the interface elements, leading to poor usability, is associated with a greater number and of saccades and longer duration, as well as with visual search strategies characterised by transitions between non-contiguous areas of interest.

It is important to underline that the choice and interpretation of those metrics must not be rigid. In fact, it always depends on the researcher’s objectives: if the main goal of the study is to understand the effectiveness of a banner ad in capturing customers attention, a great number of fixations on it (and long durations) will generally be considered as a positive fact; on the contrary, if the purpose of the study is to investigate how easy or difficult it is to find a specific element on a web page, a great number of fixations may be associated with poor usability, as the website is likely to be more complex and difficult to understand and browse.

The sequence of eye movements occurring while the user explores an interface may provide valuable information about her browsing experience and cognitive load. Behavioural metrics allow to objectively investigate user strategies and behaviours, avoiding potential distortions that are often related to the use of subjective metrics (such as self-report questionnaires and qualitative interviews).

With that in mind, considering the importance of usability and workload in optimizing the performance of individuals interacting with an interface, a clarification of the relationship between these constructs is strongly needed. Recent studies have made an attempt in this direction, however they showed mixed results. In a recent study Kokini and colleagues (2012) found that an increase in mental workload negatively affected perceived usability. Furthermore, Fedele and colleagues (2017) have found that positive interaction experiences were associated with lower mental workload. Differently, Longo and Dondio (2015) reported no relationship at all between the two constructs and suggested that they should therefore be considered separately. Nevertheless, resources allocation is an issue in any type of interaction, and the integration of mental workload assessment into usability evaluations could have important implications for improving interface design as well as for better defining the usability construct.

The study

The present study was carried out as part of an ongoing collaboration between the Italian ICT government agency (Istituto Superiore per le Comunicazioni e le Tecnologie dell’Informazione: ISCOM) and the Eye-Tracking Laboratory at the Department of Psychology, Sapienza University of Rome. The main objective of the project, named “WEBLOAD”, was to evaluate the mental workload associated with the navigation of PAs websites with different levels of complexity and understanding its effects on perceived usability.
The main hypothesis of this study was that a greater complexity of the information architecture structure would be related to higher mental workload and poor usability evaluations. Three large PAs websites (whose identity we are not allowed to disclose) were selected after a heuristic evaluation of their IA structures aimed at exploring the number of menu levels and their related categories. All websites were similar in the design and interaction features (menu, colours, aesthetic) but different in terms of information architecture complexity. From the less complex to the more complex, the identified websites will be referred as website 1, website 2, website 3 hereinafter.

Table 1: Information Architecture of the selected websites.

<table>
<thead>
<tr>
<th>Website</th>
<th>Information Architecture structure (Number of levels and number of categories per level)</th>
<th>Total</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1</td>
<td>Level 2</td>
<td>Level 3</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>44</td>
<td>134</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>53</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>45</td>
<td>109</td>
</tr>
</tbody>
</table>

Participants

Twenty volunteers participated in the study (7 females, average age = 57; SD = 6). All of them were native Italian speakers, were naive as to the aims, the expected outcomes, and the methodology of the experiment, and had normal or corrected-to-normal vision. All the subjects declared to use the Internet every day. This study was performed in accordance with the Helsinki Declaration of the World Medical Association.

Materials and method

The X2-30 eye-tracking system (Tobii, Sweden) was used to record eye movements during the interaction with the three websites. This is a standalone eye tracker that can be used in various setup by attaching it to monitors, laptops or for performing eye-tracking on physical objects with a sampling rate of 30 Hz.

Subjective measure of perceived usability and mental workload were collected at the end of the interaction with each website using the following scales.

- Net Promoter Score (NPS: Reichheld, 2003, 2006): a single item “How likely would you recommend this website/brand to a friend or colleague?”. Subjects can answer using an 11-point scale (0 to 10). This questionnaire is based on the fundamental concept that a user (customer) falls in one of the following three categories: Promoter (provides a score between 9 and 10), Neutral (provides a score between 7 and 8) and Detractor (provides a score between 0 and 6). The final value of the NPS is obtained by subtracting the percentage of detractors from the percentage of promoters. Sauro and Lewis (2010) found a strong correlation between perceived usability and NPS. Word of mouth is critical for
information complexity and usability: the role of mental workload

the success of PAs websites and digital services. Therefore, the NPS was included as a measure.

- **Usability Evaluation 2.0**: (Us.E. 2.0: Di Nocera, 2013) a multidimensional questionnaire for the evaluation of website usability. The questionnaire consists of 19 items subdivided into three subscales representing the framework users would adopt for evaluating the quality of their interaction with the interface: (Mental) Handling, Satisfaction and Attractiveness. Us.E. 2.0 allows fast assessment of website perceived usability identifying critical issues that could be eventually addressed by more extensive testing and re-design. Users are required to answer to all items along with a 5-point Likert scale (ranging from “strongly agree” to “strongly disagree”). The “(Mental) Handling” scale measures the interaction with the structure of the website (e.g. information architecture, layout). This is the “hard” usability aspect. Low scores in this scale would suggest the need to make changes to information architecture or page layout. The “Satisfaction” scale measures the perceived utility of the website. Low scores in this scale may indicate that the website does not meet users’ needs either because the users are not those expected by who created the website or because contents/services are not those expected by the users. Finally, the “Attractiveness” scale measures the interaction with the aesthetic features of the website (i.e. apparent usability). Low scores in this scale would suggest the need for a restyling.

- **NASA Task Load Index (NASA-TLX: Hart & Staveland, 1988)** has been administered at the end of a task. The respondent provides an evaluation of his/her perceive workload along six scales (from 0 to 100): mental demand, physical demand, temporal demand, effort, performance, and frustration. This self-report questionnaire is one of the most used to evaluate the mental workload experienced by individuals during the execution of a task (Wickens, Hollands, Banbury & Parasuraman, 2013; Hart, 2006).

**Procedure**

Five equivalent research tasks have been designed for each website. Subjects had to search specific information in different areas of the websites (e.g. “Transparent Administration”, “Open Data”, “Downloads”).

Participants performed the entire test in three separate sessions of the approximate duration of 50 minutes and at about 15 days apart from each other in order to limit effects related to fatigue and task duration. Moreover, with the aim of avoiding effects related to the order of presentation of the stimuli, the websites and the tasks were randomly assigned to the participants. Prior starting each navigation session, the experimenter asked participants if they had ever interacted with the website under testing. Participants had a free navigation session to familiarize with the websites before the calibration of the eye-tracker. Therefore, they were instructed on the activity and were asked to perform the task to their best.

At the end of each task, participants also reported their perceptions about the level of complexity of the task on a scale from 1 to 5 (1 = Not difficult at all; 5 = Extremely difficult). After completing all the five tasks participants reported their personal
information (gender, age, educational qualification, employment, the frequency of internet use) using an on-line form, and answered to the satisfaction, perceived usability and mental workload scales.

**Data analysis and results**

Success rate, completion time, perceived complexity, NPS, Us.E. 2.0 (Handling, Satisfaction, Attractiveness), NASA-TLX, and NNI scores were analysed in repeated-measure ANOVA designs using Complexity (website 1 vs. website 2 vs. website 3) as repeated factor.

Success rate was significantly different between websites ($F_{2,36} = 3.04; p < .05$). Duncan post-hoc testing showed that success rate for the high-complexity website (website 3) was significantly lower than the other two.

![Success rate per website.](image1)

**Figure 1. Success rate per website.**

Completion time was significantly different between websites ($F_{2,36} = 4.02; p < .05$). Duncan post-hoc testing showed that completion time for the low-complexity website (website 1) was significantly faster than the other two.

![Completion time (average) per website.](image2)

**Figure 2. Completion time (average) per website.**

Perceived complexity was significantly different between websites ($F_{2,36} = 3.92; p < .05$). Duncan post-hoc testing showed that perceived complexity of the low-complexity website (website 1) was significantly lower than the other two.
NPS score was significantly different between websites ($F_{2,36} = 4.52; p < .05$). Duncan post-hoc testing showed that proportion of the low-complexity website (website 1) was significantly higher than the other two.

Us.E. 2.0 scores were significantly different between websites ($F_{2,n} = 5.12, p < .05$) and between scales ($F_{2,n} = 17.77, p < .01$). No significant interaction between website and scale was found. Duncan post-hoc testing showed that 1) usability for the low-complexity website (website 1) was significantly higher than the other two and 2) handling was significantly lower than satisfaction and both were significantly lower than attractiveness thus indicating (Mental) Handling as the most awkward dimension.

NASA-TLX score was significantly different between websites ($F_{2,n} = 7.38; p < .01$). Duncan post-hoc testing showed that perceived workload for the low-complexity website (website 1) was significantly lower than the other two.
The Nearest Neighbour Index was significantly different between websites ($F_{2,36} = 6.41; p < .01$). Duncan post-hoc testing showed that the fixation pattern of the medium- and high-complexity websites (websites 2 and 3) were significantly more clustered than the low-complexity website.

Discussion and conclusion

Ensuring high usability of e-Governments tools, such as websites and web-based platforms plays an essential role in improving the efficiency in terms of costs and quality of services offered to citizens and businesses. Therefore, the evaluation of user performance and perceptions during the interaction with institutional websites has a key role both in the design phase of the interfaces and in their usability assessments. However, despite the growing interest of Public Administrations towards those topics, the relationship between complexity of information architecture, usability perceptions and mental workload imposed on the user has not been sufficiently investigated.

The objective of the present study was to evaluate the workload imposed on the user by “information abundant” websites and its relation with perceived usability. The idea that the complexity of information architecture can influence both the cognitive load imposed on user and his perceptions related to the pleasure of the user experience has been already suggested in the literature (e.g. Conklin, 1987). However, there is a lack of experimental research on this issue.

Here we have selected three existing Italian Government websites with different levels of information complexity to test the research hypothesis. Results indicated a consistency between satisfaction, usability and mental workload measures. Specifically, the websites associated to lower levels of mental workload (assessed by
both objective and subjective measures) received more positive usability evaluations were associated with a greater successes rate in the assigned search tasks, and participants reported their willingness in spreading the word about their usefulness.

The analysis of eye movements showed statistically significantly higher NNI values for the website 1, while lower values were reported for the websites 2 and 3. Considering the visual exploration strategies, this result highlights a less clustered fixations pattern for the website 1 and, on the contrary, a more clustered fixations pattern for the websites 2 and 3. Based on previous research (Camilli, Terenzi, & Di Nocera, 2008), when a task imposes high visual-spatial demand on the user-as in the case of an information search task- fixations clustering (i.e.: smaller NNI values) corresponds to greater mental workload experienced by the user. Subjects involved in this study have indeed experienced a greater mental workload while browsing the higher-complexity websites. This result is also supported by the subjective evaluations expressed using the NASA-TLX. Consistent results can be found in the usability evaluations expressed using the Us.E 2.0. scales, the “Handling” dimension (which is related to the information architecture) being the most problematic. Usability evaluations of a website are generally negative when users take too long to complete the task, make mistakes or fail in its execution (Nielsen & Levi, 1994; Nielsen, 1999; Palmer, 2002). In this study success rate was higher and completion time shorter for low-complexity website than the other two.

In conclusion, results obtained in this study underline the necessity and importance of integrating an estimate of users’ mental workload during the design and evaluation of usability of complex websites such as those of Public Administrations. Some limitations that may have influenced the findings of this study should also be reported. A first limitation can be attributed to the scarce heterogeneity of the experimental sample in terms of age range and occupation. In fact, the age of the participants was between 46 and 65. Moreover, all the participants were Government employees who may have, in some way, benefited from their knowledge regarding the PA’s websites structure and services. Although the type of processes investigated here can be considered common to all individuals, it could be useful to involve different types of users (e.g. experts vs. naives, typical vs. occasional users, etc.). Future studies could also include other measures (subjective and objective) in order to better understand the relation between the observed variables increasing the validity of results. In addition, the relationship between usability and mental workload could be influenced by both the information architecture and the layout of websites. A further objective could be to understand the specific role of these two variables. For example, conducting further studies contrasting the same information architecture with different layouts (e.g. the mobile version of the same websites), could allow greater generalization of the observed relationship between information architecture, mental workload and usability perceptions.

References


Garrett, J.J. (2010). *Elements of user experience, the: user-centered design for the web and beyond*. Pearson Education.


Contribution of Industry 4.0 to the emergence of a joint cognitive and physical production system

Philippe Rauffet, Clément Guérin, Christine Chauvin, & Eric Martin
Lab-STICC UMR CNRS, University of South Brittany, France

Abstract

The digital and technological revolutions of Industry 4.0 aim at increasing the flexibility of companies, the mass customization of products and the improvement of working conditions. Thus, Internet of Things (IoT) and biofeedback sensors become new sources of information on the production context and the state of resources, big data and cloud computing offer increased processing capacity (for learning and simulation), virtual and augmented realities, as well as vocal or gestural commands, make the interactions more contextual, natural, and customized. The integration of these Industry 4.0 technologies must however reconsider the place of the operator in the production, to tend towards a joint cognitive and physical system. Interference between operators and the cyber-physical production system must be optimized, by fostering the emergence of a real know-how to cooperate on the different production management activities (line operations, supervision, planning). In this perspective, the paper provides an overview on the Industry 4.0 technologies and their impact on the human-systems cooperation. A synthesis model proposes to position these technologies around the processes of building common frame of reference and distributing functions between humans and the cyber-physical system. This model is finally illustrated by a conditioning activity, shared between a human operator and cyberphysical components (cobot, augmented reality, etc.).

Introduction

The term Industry 4.0 emerged in the early 2010s and refers to the advent of a fourth industrial revolution (Drath & Horth, 2014) and the application of the Cyber-Physical Systems concept to the field of industrial production systems. For Wang, Wan and Zhang (2016), Industry 4.0 is one type of industry which incorporates technological advances for addressing major societal challenges: improving the quality of life with high-quality customized products, participating in the sustainable development by reducing non-renewable energy consumption, and offering better working conditions to employees.

Rüssman et al. (2015, Boston Consulting Group report) identify nine technology pillars related to the digital revolution of industry.

On the one hand, Internet of Things, horizontal and vertical integration of Information System, Big Data, Cloud computing, cybersecurity, will facilitate
management of companies’ information by linking or decentralizing data, making data reliable, or also gave rise new knowledge becomes available (Hozdic, 2015). These technologies contribute to the advent of cyber-physical systems, defined as interactive networks of physical and computer components, characterized by their evolutionary and decentralized nature, and their properties of openness and context sensitivity (Fantini, Pinzone & Taisch, 2018). In addition, they offer enhanced capabilities in monitoring, modelling, analysis, and calculations, using data mining and machine learning techniques (Longo et al., 2017).

On the other hand, simulation, virtual or augmented reality, improves production verification and supervision capabilities. Longo et al. (2017) argue that the industry 4.0 will allow the emergence of the virtual enterprise, thanks to the constitution of digital twins. According to Rosen, Wichert & Bettenhausen (2015), this notion of ‘digital copying’ can be used to reduce design costs, to detect errors early in the preliminary design phase of products or production systems. Digital twins could also help for supervision, displaying in real time in virtual environments the state and behaviour of machines and production. This immersion of the digital twin in reality could be done using augmented reality. Augmented reality, today often dedicated to one or a few specific tasks, will become ubiquitous (Grubert, Langlotz, Zollmann & Regenbrecht, 2017). It will be adaptive depending on the context (location, type of activity, situation encountered), to project operational information on the immediate environment of the task to be performed, but also more tactical information on the general environment of production. Longo et al. (Ibid.) believe that this combination of virtual and real worlds is carrier and could bring interesting applications for maintenance (remote assistance) or training to operational situations.

Finally, additive manufacturing and self-adjusting robotics will support more flexible production. According Rüssmann et al. (2015), robotics becomes a more sustainable means of collaborative work. At the same time, additive manufacturing, now confined to prototyping, is becoming more industrial, and moving from mass production to “mass customization”. All of this contributes to the emergence of autonomous and self-organized manufacturing systems (Longo et al., 2017), making production more agile.

These opportunities, brought by the implementation of new technologies in manufacturing systems, raise new challenges (Hozdic, 2015; NIST, 2016; Longo et al., 2017). In particular, the cyber-physical production system must be resilient, responsive and flexible in the face of critical situations. Manufacturing must become intelligent and holonic (NIST, Ibid.) and rely on artificial intelligence that can automatically adapt to the changing environments and demands of the process. Above all, we must move from the era of intelligent products (for example with RFID), to that of intelligent machines and the optimization of joint systems, combining operator and cyberphysical systems. The emergence of these joint systems (Hollnagel & Woods, 2005) also called “symbiosis” between operators and digital assistances (Romero, Bernus, Noran, Stahre, Fast, & Berglund, 2016), is a major challenge of the industry 4.0. It addresses key questions about the future place of human in industry 4.0 (Hozdic, 2015, Hirsch-Kreinsen, 2014, Longo, Nicoletti, & Padovano, 2017).
The purpose of the paper is to contribute to the understanding and modelling of this joint industry 4.0 production system. The first part presents a review of the literature on the place of human in industry 4.0 and develops the concept of operator 4.0. A state of the art on human-machine cooperation models is also conducted in this section, in order to better understand the emergence of the joint, cognitive and physical system of production. The second part then presents a synthesis framework, to model the possibilities of man-machine cooperation between the operator and these new assistances of industry 4.0. This framework is applied to a use case (the activity of order picking). Finally, these proposals are discussed, in terms of the prospects for the design and adaptation of Industry 4.0 to future operators.

I. Related works

I.1. Towards operator 4.0, a joint cognitive and physical system of production

The place of operators in future smart factory

Several authors identify various issues related to the role of operators within these new sociotechnical systems. Hirsch-Kreinsen (2014) describes the impact of this new “industrial revolution” on the skills and qualifications of operators as well as the evolution of the organization of work

- **Skills and knowledge.** Operators will be required to demonstrate high adaptive capacity. They will act on complex, interconnected and autonomous technical systems that they will have to understand and control. Operators will manipulate more abstract information and will often be more “remote” from the process to be controlled, with an increased scope of supervision (Hoc, 1996).

- **Qualification.** Industry 4.0 will employ fewer unskilled workers than the traditional factory. Indeed, as Fantini, Pinzone and Taisch (2018) recall, it is the routine activities, characterized by a low level of requirements of manual dexterity or social interaction, which are most likely to be replaced by the technology. On the other hand, it will need qualified and well-trained workers (Lorentz, Russmann, Strack, Lueth, & Bolle, 2015).

- **Work organization.** Hirsch-Kreinsen identifies two extreme forms of organization and division of work between humans: a “polarized” organization with few disqualified tasks and a large group of highly qualified experts and specialists versus a “distributed” organization aiming for flexibility and based on a high level of qualification which enables operators to face unanticipated situations.

The place of human in the productive systems is thus to rethink in depth. It is in this sense that Hozdic (2015) calls for the updating of the operator model. The author specifies that it is necessary to reflect on new human activities, taking into account the new possibilities of the mixed reality combining the physical world and the digital / cybernetic world (through interfaces, techniques of representation of knowledge, augmented reality, etc.). In this perspective, Hozdic (Ibid.) also emphasizes the need to model new architectures of cyber-physical production systems (CPS), ranging from sensors to the presentation and contextualization of decision support, and that keep the human in the loop, or even increase its capabilities.
In order to better define the place of the human in the context of Industry 4.0, some authors have recently sought to build typologies of the operator, qualified as “operator 4.0” (Romero et al., 2016), to from the new technological revolutions (see §I.1). They have formalized in a detailed way the interactions with the cyber-physical system of production (Romero et al., 2016, Romero, Wuest, Stahre, & Gorecky, 2017), or proposed a framework to design and evaluate these interactions (Fantini et al., 2018).

**Operator 4.0: a symbiosis between human and machines**

Romero, Bernus, Noran, Stahre and Fast Berglund (2016) explain that the operator 4.0 is part of a new approach to designing adaptive production systems. Automation is considered as a means to increase human capacities at the physical, sensory and cognitive levels. The “symbiosis” between the human and the cyber-physical system, will allow emerging a new intelligent hybrid agent (human and artificial). This vision proposed by Romero et al. (2016) somehow joins the current of cognitive systems (Hollnagel & Woods, 2005). This approach follows and extends the concept of “Balanced Automation System” (Romero et al., 2016, Fantini et al., 2018): it is no longer only a question of finding a balance between manual tasks and automated tasks, but to design an adaptive and dynamic automation, based on a dynamic allocation of functions between the man and the machine.

According to Fantini et al. (2018), robots and algorithms are not intended to replace the human in industry 4.0, but are instead intended to become assistance that allows the operator to continue working, despite age, disabilities or inexperience, which favour the maintenance of human “in the control loop”, and which improve the performance and comfort of workers.

**A typology of operator 4.0, based on digital revolutions**

Romero, Bernus, Noran, Stahre, Fast and Berglund (2016) detail the different “increase” of the operator 4.0. They propose a typology based on the technological pillars of industry 4.0 listed by Rüssmann et al. (2015), and secondly on the nature of the assistance (physical, sensory or cognitive) provided to the operator (Figure 1).

The operator 4.0 can be physically increased (“super-strength operator” and “collaborative operator”), using cobots and exoskeletons. This will contribute to the increase in performance and the reduction of musculoskeletal disorders (the physical load can be shared between the man and the robot) and will also allow the insertion of disabled workers.

The operator 4.0 can also be increased on the cognitive level (to better treat and interpret the information, and to solve in a more optimal way the problems met) and on the sensory level (to better perceive the environment and to detect new signs):

- Biofeedback sensors will bring a sensory increase (“healthy operator”). Like the IoT sensors, they will allow a reflexive evaluation of the operator on the physical and mental effort that he provides, and thus contribute to the detection of risk situations, when the man is physically overloaded or cognitive
- New visualization interfaces, such as virtual reality and augmented reality, will contribute to sensory and cognitive enhancement of the operator (“augmented...
industry 4.0 and a joint cognitive and physical production system

operator” and “virtual operator”). The diagnosis of the situation can indeed be enriched, through the display of information from the cyber-physical system, and this contextualized (depending on the location, the role of the operators, the situation encountered). The perception of the environment will also be increased, thanks to the subjective vision proposed by these visualization means, and to the display of visual alerts or indexes highlighting the important elements of the environment.

![Figure 1. A typology of operator 4.0 (from Romero et al., 2016)](image)

Enterprise social networks will participate in a cognitive increase of the operator (“social operator”). They will notably improve the diagnosis of the real situation and help the problem solving thanks to the knowledge of other experts accessible online, synchronously (chat, video), or asynchronous (with the information “traces” left on the platforms of wiki type, forum). The analysis of big data and cloud computing will lead to a cognitive increase of the operator (“analytical operator”). They will extend the capabilities of calculation, classification, analysis and synthesis of information of the company, and thus enrich the analysis of the situation by the operator, whether at the strategic, tactical or operational level.

Finally, the operator 4.0 will benefit from increased interactions with the cyber-physical system (“smarter operator”), in support of the various increases of the operator listed above. Human-machine interactions will be improved by the use of personal assistants and artificial intelligence. The queries and commands can then be transmitted in natural language between the man and the machine, and the understanding of the needs of the operator will be contextualized and enriched (through a learning process where the personal assistant will consolidate a model of the operator).
Operator 4.0, a social agent

In addition to this vision based on the “symbiotic” assistance and increases of the operator, we must also approach human-machine cooperation with a perspective of social interactions (Fantini et al., 2018). In this sense, Romero, Wuest, Stahre and Gorecky (2017) refine their vision of the operator 4.0, placing it at the centre of a social network, made up of other social operators, but also of machines and software qualified them also of “social”. These authors propose a multi-agent architecture of the social factory to better formalize these social interactions and this holonic production system. This is based in particular on the introduction of two types of agents:

- The **interface agents** correspond to a set of rules and interaction conditions that support interactions between human or technical agents and the rest of the system. These interface agents are qualified as active, in the sense that they evolve continuously, by learning (based on observation, imitation, return of other agents or programming). They thus make it possible to personalize the assistance provided to the agent, making it dynamic and adaptive, and thus contributes to keeping the social operator or the social machine in the loop when they encounter difficulties. These agents correspond to the intelligence that drives the adaptive HMIs and manage the dialogue, adapting to the context or the operator. An example is the personalization and the customization in google maps, where trips can be proposed to a driver depending on the schedule, the driver’s location, and his/her travel history, and where the states of heaters or lights in driver’s home can be modified according to this trip.
- The **broker agents** correspond to the rules of allocation of functions and sharing or delegation of authority. They thus make it possible to adapt the level of automation to optimize man-machine cooperation.

Optimization of joint production systems: a human-machine cooperation perspective

Attempts to model the place of the operator in Industry 4.0 and its “social” interactions with other agents of the organisation highlight the emergence of a joint cognitive system, symbiosis of human and of the cyber-physical system. Obtaining a “symbiotic” system must go through an optimization of man-machine cooperation. This optimization is conditioned by:

- The creation of a common frame of reference (Clark, 1996, Hoc, 2001), including both a shared awareness of the environment, but also a representation of each agent and the team, seen as resources. It is in this sense that the interface agents allow adaptive assistance. This common frame of reference must also be constructed taking into account the questions of transparency, which will refine and guide the dialogue between the different agents (Chen, Procci, Boyce, Wright, Garcia & Barnes, 2014).
- The design of a dynamic allocation of functions, allowing a sharing of work between the man and the machine that responds to the hazards and keeps the operator in the loop.

I.2. Human-machine cooperation models to understand operator 4.0

Human-machine cooperation can be understood by structural approaches that define the structure of the relationships between cooperating agents, or functional
approaches that describe the cooperative activities that develop between agents (Chauvin & Hoc, 2014)

Human-machine cooperation models

Millot and Mandiau (1995) defined two generic organizational structures for cooperation:

- Vertical cooperation in which an agent at a higher level supervises and has authority. The lower level agent can provide advice;
- Horizontal cooperation where the agents are at the same hierarchical level and the tasks are shared by allocation, the sharing being defined by the higher hierarchical level.

Schmidt (1991) takes a functional-structural approach by distinguishing three types of cooperation, which can be observed when agents are mutually dependent to carry out their task (their individual activities must then be articulated in order to achieve the work objectives):

- Augmentative cooperation that aims to increase physical or intellectual abilities since additional agents with identical skills perform the task when the workload increases and cannot be managed by a single agent;
- Debating cooperation makes it possible to confront points of view between agents in order to make the solutions more reliable and reduce the errors. This type of cooperation requires some agents to check and control other agents;
- Integrative cooperation joins complementary skilled agents.

According to Hoc’s (2001) functional approach, two autonomous agents are in a cooperative situation if two minimum conditions are met. First, each agent pursues goals, and each may interfere with goals, resources, or procedures. Secondly, each agent strives to deal with these interferences in order to facilitate the accomplishment of the individual activities of each or the accomplishment of the common task if it exists. This interference management can be observed at three levels, defining three levels of cooperation:

- L1: cooperation in action (or execution). This level distinguishes between different operational cooperative real-time and short-term goal and procedure management activities: local interference creation (e.g. disagreement), local interference detection (e.g. redundancy), anticipation, and interference resolution.
- L2: cooperation in planning. This level is characterized by cooperative activities for developing or maintaining a common reference system: maintaining and developing a common objective, a common plan, a distribution of functions
- L3: metacooperation. It facilitates the two previous levels by developing a common code of communication and models of oneself and the partner

The central concept of common ground in human-machine cooperation

To move towards joint cognitive systems based on a symbiosis between operator 4.0 and technologies of industry, shared representations of situations, adapted to the contexts, must be supported (Salas, Prince, Baker, & Shrestha, 1995, Kaber & Endsley, 1998). Millot and Pacaux (2013) propose to extend the work of human-machine cooperation from framework of collective situation awareness (team-SA, shared-SA and distributed-SA) defined as the shared understanding of a situation
between agents. Although each individual has an individual representation of a given situation, one can identify a level of “recovery” of individual SA whose key elements facilitating the construction of the team-SA are to share the goals and future states of the system. (Salmon et al., 2008). The 4.0 technologies will thus facilitate a level of mutual transparency between the partner agents, by making it possible to build, feed, update and adapt the common repository of agents. The common frame of reference becomes the “linchpin” of cooperating agents since it allows: (a) cooperation in action and planning (Hoc, 2001) through effective coordination mechanisms (Schmidt, 1991). By generalizing the joint activity concepts developed by Clark (1996), Klein, Feltovich, Bradshaw, & Woods (2005) indicate that this coordination is facilitated when team members are inter-predictable (they can predict the actions of others) and have a sufficient common frame of reference; (b) enrichment of self and partner models. This enrichment of models helps to develop the “know-cooperate” component of cooperation (Rogalski, 1996).

**Know-How to Cooperate and processes for building common ground**

Millot and Pacaux (2013) propose a model of human-machine cooperation that articulates the coordination mechanism to the concept of knowing how to cooperate. Their model of cooperation is based on two principles: (1) the agents have know-how in the control of processes or the accomplishment of tasks (2) the agents have knowledge to cooperate with other agents. From the definition of cooperation proposed by Hoc (2001), the authors consider two processes contributing to the know-how to cooperate: (a) the coordination which makes it possible to manage the interferences between the aims pursued by the agents (detect, create, solve); (b) facilitating the goals of other agents. Their common reference concept, the Common Work Space (Pacaux-Lemoine & Debernard, 2002, Millot & Pacaux-Lemoine, 2013) then makes it possible to support cooperative activities.

**Industry 4.0, a support for the enhancement of human-machine cooperation**

Technologies of Industry 4.0 will help to build and adapt the common frame of reference of cooperating agents by supporting the processes of coordination and facilitation of goals. These adaptive mechanisms of the joint human-machine systems may vary according to spatial and temporal parameters, making it possible to define different modes of cooperation (Schmidt, 1991): close or distant cooperation, synchronous or asynchronous cooperation, collective or distributed cooperation, direct or indirect cooperation. The “state” or “quality” of the common frame of reference could then be a criterion allowing to favour real-time flexible human-machine cooperation (Hoc, 2001) and thus “activate”, according to the paradigm of the dynamic allocation of functions, modes of assistance to optimize resources or recover a degraded situation.

II. A framework for modelling joint cognitive and physical production systems

II.1. Development of a human-machine cooperation framework for operator 4.0

The functions of production to model

Operators perform functions of different natures to perform their tasks. In the current definition of jobs of operators, there is indeed a combination of several aspects of production management that an operator must manage. We can find for example:
The scheduler, which does both short-term planning and supervision, possibly correcting the planning of the manufacturing orders of the day according to the encountered hazards (related to the demand, or the shortcomings of the production system provoked by machine breakdowns or the absence of employees);

In empowering organizations (e.g. in Michelin or French Post Office), flow facilitators do both production operations and supervision. They will carry out support activities (quality control, training) or managerial activities (point 5 minutes, transmission of directives and feedback of field problems) in addition to their implementation tasks.

Rather than focusing on roles, we will discuss in terms of production functions, by distinguishing operational, planning and supervisory functions. These different functions can be analysed according to the level of control that operators can have on operations. This control can be characterized following two axes (cf. figure 2):
- Control can be reactive (during production) or anticipatory (before production).
- The control level depends on the margins that operators can have on production load and capacity (in other words, on the trade-off that an operator will be able to find between the demand of production and the implementation of the human and technical resources).

**Figure 2. Mapping of functions of production and existing cyberphysical components**

*During production (reactive control), we will therefore distinguish two kinds of function: (a) production operations (product manufacturing, maintenance, etc.) consist in implementing resources to meet the demand. Production capacity and load are fixed, and the operator must remain in this constraint field, with a reflexive control of the action in progress; (b) production supervision consists of monitoring and reconfiguring the production system, if necessary, to ensure that the resources (capacity) can adequately meet the demand for production (load). Capacity can therefore be redefined (redistribution of tasks, prioritization of manufacturing or intervention orders), but demand is a constraint.*
Before production (anticipatory control), planning tasks aim at monitoring and adjusting the load / capacity ratio in the management of production operations. Operator plays on load AND capacity, with a more or less distant horizon (at the level of Industrial and Commercial Planning or Master Production Scheduling).

**Categorization of industry 4.0 technologies**

The technologies for industry 4.0 can be grouped into four categories (Tech1 to Tech4), according to their impact on information processing or implementation of the action (Table 1).

**Table 1. Characterization of industry 4.0 technologies**

<table>
<thead>
<tr>
<th>Technologies 4.0</th>
<th>Impact for industry 4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet of Things &amp; biofeedback sensors</td>
<td></td>
</tr>
<tr>
<td>« connected &amp; healthy operator »</td>
<td></td>
</tr>
<tr>
<td>Bigdata, Cloud computing, Machine Learning</td>
<td></td>
</tr>
<tr>
<td>« analytical operator »</td>
<td></td>
</tr>
<tr>
<td>Enterprise social network</td>
<td></td>
</tr>
<tr>
<td>« social operator »</td>
<td></td>
</tr>
<tr>
<td>Tech1</td>
<td>Raw information collected by new sensors</td>
</tr>
<tr>
<td>Tech2</td>
<td>Information processing</td>
</tr>
<tr>
<td>Tech3</td>
<td>Situation information, adapted to context</td>
</tr>
<tr>
<td>Augmented reality</td>
<td></td>
</tr>
<tr>
<td>« augmented operator »</td>
<td></td>
</tr>
<tr>
<td>Virtual reality</td>
<td></td>
</tr>
<tr>
<td>« virtual operator »</td>
<td></td>
</tr>
<tr>
<td>Personal Digital Assistant</td>
<td></td>
</tr>
<tr>
<td>« smart operator »</td>
<td></td>
</tr>
<tr>
<td>Tech4</td>
<td>Implementation of action</td>
</tr>
<tr>
<td>« supers-pen strength operator »</td>
<td></td>
</tr>
<tr>
<td>« exoskeleton »</td>
<td></td>
</tr>
<tr>
<td>« collaborative operator »</td>
<td></td>
</tr>
<tr>
<td>Assistance for constrained actions (Norm NF EN 1005)</td>
<td></td>
</tr>
<tr>
<td>Assistance for constrained actions (lift, carry, push-pull / Norm NF X35-109)</td>
<td></td>
</tr>
</tbody>
</table>

**A framework for modelling operator 4.0**

The figure 3 below synthesizes the human-machine cooperation mechanisms that should be developed for designing future operator 4.0, by positioning the different industry 4.0 technologies and the different levels of functions of production. The four categories of technology 4.0 in Table 1 (Tech1 to Tech 4) can be positioned according to the two main levels of industrial activity (Figure 3): at the upper level, with tactical activities of planning and supervision, and at the lower level, the operational activities of product manufacturing.

Industrial activities at the operational level are performed by a combination of human agents and machines (H or M or H-M or M-M). At this level, sensor technologies (Tech1) can dynamically collect information related to the state of material resources (IoT: temperature, hydrometry, presence, etc.), to the operator functional states (biofeedback: heart rate variation) or to the characteristics of the production processes. At this level too, the human can be assisted in performing physical tasks (Tech4).

The collected data of the situation of production are treated at the tactical level by human or artificial agents, which manage the diagnosis and the adaptation of the operational level. This process is performed by comparison with models (individual’s long-term memory, organizational memory and computer database). This comparison process, supported and enriched by big data and machine learning (Tech2), allows:
industry 4.0 and a joint cognitive and physical production system

Figure 3. Framework for modelling operator 4.0 with a human-machine cooperation perspective according to industry 4.0 technologies and functions of production.
To evaluate the current situation (diagnosis), by measuring the quality of the COFOR (“interface” property of supervisory agents 1 and 2) and the relevance in the distribution of the functions at the operational level (“broker” property of supervision agents 1 and 2). The diagnosis is based on the mutual understanding of goals and the compliance with procedures, and focused on short term operational activities of production (corresponding to cooperation in action of Hoc, 2001);

To generate system adaptations from the results of these assessments. The adaptation criteria must satisfy the maintenance of a common reference system (Hoc’s cooperation in planning, 2001) and may lead to re-allocations of functions at the operational level (broker property of supervisory agents);

To infer new rules to improve models and the know-how to cooperate (Hoc’s meta-cooperation, 2001) through a learning process.

At this tactical level, decision-making on adaptations of the industrial system also involve a dialogue between supervisory agents, which feeds a Common Work Space or CWS (for more details on the construction dynamics of the CWS, see Millot & Pacaux-Lemoine, 2013, Fig. 11). This dialogue between supervisory agents then makes it possible to determine the cooperation structure, horizontal or vertical (see Millot & Mendiaux, 1995, Millot & Pacaux-Lemoine, 2013), which will define the role and authority of each agent for the operational regulation of the industrial system (information sharing, task sharing). The results of this “horizontal” dialogue, resulting in the delegation of authority to one or the other of the supervisory agents, will then be transmitted “vertically” at the operational level, in order to:

- Coordinate the operating agents with each other according to a mode of cooperation adapted to the situation (augmentative, debatable or integrative, see the Schmidt taxonomy);
- Maintain COFOR by facilitating perception, understanding and projection of future states of the situation, including the context of the task in which the information can be transmitted (Tech 3), based on the same information transparency rules.

This model of operator 4.0, with a human-machine cooperation perspective, replaces the new technologies of industry 4.0 as input or assistance for managing the dialogue between agents and coordinating the joint activity. It aims at supporting the definition and the analysis of new activities of operators 4.0. The next section will notably study the order picking activity. This specific activity will include both planning and organization tasks for the different production batches to be performed, as well as physical production tasks related to batch packaging.

II.2. Application of the framework to the specific activity of order picking

We consider the activity of an order picker, who can be assisted by all the different kind of technologies presented in table 1. Tech 1 will be composed of biofeedback sensors (smart watch on operator) and IoT sensors (RFID, smart sensors for detecting the presence of empty parcel boxes in storage). Tech 2 consists in calculation capabilities provided by the digital twin (predicting breakdown of production agents from Tech 1 sensors) and the integration of other data (related to
industry 4.0 and a joint cognitive and physical production system

customer demand). Tech 3 will support dialogue and communication with augmented reality glasses, pick-to-light and voice-to-pick technologies. Finally, Tech 4 corresponds to a cobot, able to interact with a human agent in a parallel or a sequential process.

The activity of an order picker 4.0, considered as an agent in symbiosis with the four technologies, could be described through the following scenario, designed with the aid of the framework proposed in Figure 3:

- Let consider that the customer demand is increasing. The activity of conditioning (filling and closing parcel boxes according to the customer orders) could be therefore shared between human and cobot (Tech 4) with a parallel organization (in the augmentative mode of Schmidt’s cooperation taxonomy, where each agent is making a complete box at the same time).
- The detection of production problems is enhanced with the aid of Tech 1 sensors. A Problem can be related to operator overload, or wrong cobot behaviour for specific production to fill in the boxes.
- Based on the history of incidental situations experienced at this picking workstation, machine learning and simulation capabilities (Tech3) will help for deciding between different improvement strategies. Considering a problem of wrong cobot behaviour on the filling of specific products into boxes, a choice could be proposed between:
  - Switching to a sequential organization (the human agent, in addition to his/her own parcel boxes, will also fill the boxes of cobot with the specific products, transforming this task into an integrative cooperation),
  - Stopping the production, to teach the cobot with the correct behaviour
- The data mining achieved of the factory database thus help for enrich situation knowledge, and it can be therefore considered a component of the metacooperation level of Hoc’s HMC model.
- Finally, the operator can be alerted with augmented reality (smart glasses) that there is a performance decrease of the cobot, or that a new function allocation is proposed between agents. It is also assisted in his/her manual actions by voice or by light to choose the correct products or the rightsized parcel box.

Discussion

The conceptual proposals in this paper aim at complementing the rather technocentric approach of recent research works (Romero et al., 2016 and 2017, Fantini et al., 2018) on the place of humans in future smart, connected and agile factories. The proposed modelling framework aims to better position, or even design, the new activities of operators 4.0, seen as a joint physical and cognitive production system. This notion of symbiosis is crucial in the emergence of smart factories: the new technologies should not replace human operators but help the transformation of workstation to allow operators to gain higher flexibility and power of decision.

However, we could go further in this approach, by analysing in more detail the cognitive dimension of the joint activity. The question of the “social” organization of the operator 4.0 (Romero et al., 2017) could thus be deepened with methods such
as the CWA of Rasmussen, and in particular the phase of socio-organizational analysis.

The management of human-machine dialogue should also be further explored, so that the common ground can be optimized between the two agents, thus guaranteeing the performance of the joint system and the quality of decision-making. In this sense, it would be interested to work on an enrichment of the model presented here with the concept of informational transparency, developed in particular around 2 models (Lyons, 2013, Chen et al., 2014):

- The first model concerns the transparency of the autonomous and intelligent agent towards the operator (“robot-to-human transparency”) and can be detailed using the SAT of Chen et al. (2014). Situational Awareness-Based Transparency (SAT), uses Endsley’s Situation Awareness model to define three levels of agent-to-human transparency, ranging from disclosure of basic information (reporting of actions and goals), sharing of explanations on the reasoning used (constraints considered, methods chosen), until the transmission of elements on the consequences of actions in progress (projection, uncertainty, risk);
- The second model concerns the level of information that the agent has and can communicate about the state and the behaviour of the human operator (“robot-of-human transparency”). It must include modelling and assessment of the operator’s or team’s condition (stress, fatigue), goals and behaviour (social purpose), and must also be based on an understanding of environment and the structure of the performed tasks.

The level of transparency would also influence the trust that humans would place in the machine; the higher the level of transparency, the higher the trust level would increase (Chen, 2014). On the contrary, a lack of access to data can lead to forms of “circumvention” of cooperation and a decline of trust that built up gradually by observation of the behaviour of the agent can lead to misuses. The design of future work must consider these drifts by proposing adjustments of the interface/broker agents according to the use (on the degree of transparency to bring, on the degree of automation, etc.).

**Conclusion**

Ultimately, the reflection on the industry of the future is in full swing. It mobilizes a lot of research, to address both the technical challenges of implementing new technologies in the plant, but also to consider the sociological, cognitive and ergonomic aspects of this transformation. While some authors have recently reflected on the place of the man in the industry 4.0, the existing frameworks however remain rather technocentric. This is why the current paper attempts to combine these recent works on operator 4.0 with more cognitive approaches, by especially using well-known models of human-machine cooperation.

**Acknowledgements**

Authors acknowledge the French National Agency of Research, which supports and funds Humanism Project (ANR-17-CE10-0009) and the current research works.
This project involves three laboratories in Automatics Sciences and in Human Factors, CReSTIC (Reims), Lab-STICC (Lorient), and LAMIH (Valenciennes).

References


Supporting supervisory control of safety-critical systems by optimizing visualizations on monitoring displays for fast and accurate perception

Marie-Christin Harre & Andreas Lüdtke
OFFIS Institute for Information Technology
Germany

Abstract

The amount of automation in technical systems continually increases e.g. autonomous driving or automatic failure detection in control rooms. The monitoring of a high amount of information becomes a major task of the human operator. Due to the rising quantity of information, this task becomes increasingly complex. Especially in safety-critical systems, it has to be ensured that the operator assesses critical system states fast and accurate to initiate countermeasures in time. Therefore the HMI has to be designed carefully. For this, we developed the Konect method that allows systematic derivation of efficient visualizations. The method combines knowledge from the human factors domain (e.g. task analysis, ecological interface design), the information visualization domain and knowledge about human perceptual skills. We validated the method in the automotive domain (for truck platooning) and in the maritime domain (with 17 designers creating designs, and 51 participants testing the designs in different laboratory studies). The results revealed that Konect designs performed significantly better than designs constructed with conventional methods (p<0.001) which reduced the amount of accidents by 81.8% (tested in driving simulator).

Introduction

The amount of automation in technical systems continually increases. Autonomous driving or automatic failure detection in control rooms represent just two examples. It is one of the ironies of automation, that an increased level of automation lies in higher demands for the human operator in monitoring the system instead of replacing the human operator by automatic devices: “The increased interest in human factors engineers reflects the irony, that the more advanced a control system is, so the more crucial may be the contribution of the human operator.” (Bainbridge 1983). The human operator has to assess the state of the system continuously to estimate if the system is working correctly or if an anomaly occurs. Especially in the safety-critical domain, this monitoring task is a time-critical task, as countermeasures have to be taken as fast as possible in case an anomaly is detected. To monitor the system state, the human operator observes relevant information on a Human Machine Interface (HMI). To allow fast detection of critical system states, this HMI has to be designed carefully. For this, we developed the Konect method.
that allows systematic derivation of efficient visualizations. The method combines
knowledge from the human factors domain (e.g. task analysis, ecological interface
design), the information visualization domain and knowledge about human
perceptual skills. The method is described in detail in the upcoming paper. The
upcoming section starts with presenting related work about methods for visual HMI
design. Thereafter, the method is described with a small example. The Konect
method was evaluated in two different domains (automotive, maritime). The results
revealed that HMI designs developed with Konect are significantly faster and more
correct perceivable.

Related work

The optimization of HMIs to the extent that they are optimally adapted to the tasks
of the operator has a long-lasting background in the HF domain:

The human centred design process as introduced by the international organization
for standardisation (ISO) represents a well-established process for designing an HMI
consisting of four different activities: (1) understand and specify the context of use,
(2) specify the user requirements, (3) produce design solutions to meet user
requirements and (4) evaluate the designs against requirements. Especially in the
first step, task analysis methods are commonly applied. There exist different
approaches for conducting a task analysis as proposed by Stanton et al. (Diaper &
Stanton, 2003).

Ecological Interface Design (EID) is a framework introduced by Kim Vicente and
Jens Rasmussen (Vicente & Rasmussen 1992). EID shifts the focus from the user
centred perspective to a work domain-oriented perspective. This is based on EIDs aim
of supporting the operator’s problem-solving capabilities in unforeseen problem
situations. HMIs based on EID provide knowledge about the functioning of the system
that is monitored so that the operator understands reasons for upcoming problems and
can derive suitable countermeasures (Burns & Hajdukiewicz, 2013). There exist also
approaches combining task analysis and EID (Jamieson et al., 2007).

The approaches previously presented focus on the derivation of information to be
shown and on the structuring of information. The visualization of information plays
a minor role. Konect combines both aspects – the well-established derivation of
information and structure enhanced with systematic derivation of visual appearance
of information elements.

With regard to the visual appearance, Petersen and May (2006) presented an
approach for scale transformations and information presentation in supervisory
control. Scales are used to describe the type of information to be presented (e.g.
nominal, ordinal, interval, and ratio). To define the scale type, the authors present rules
for mapping data being presented to the operator to a certain type of scale. To convert
the information into a visual form, the authors apply a theoretical framework
developed by Zhang et al. (1996). Zhang et al. presented a framework for describing
properties and structures of relational information displays. Compared to Konect, the
framework does not focus on fast and correct perception but is more oriented towards
expressiveness.
The Konect Method

The Konect Method involves four basic steps as shown in Figure 1. The first step is the information determination, the second step the idea box specification, third, the glyph sketching and the last step is the design composition. Details for all steps are described in the upcoming text.

![Steps of the Konect procedure.](image)

**Information determination**

The first sub step is the information determination. This step ensures that the information shown on the HMI optimally supports the task of the human operator. In this regard it is important that relevant information needed during monitoring is shown and no essential information is missing. At the same time irrelevant information that overcrowds the interface and produce visual clutter should be avoided. This is often referenced as information fitness (Denker et al., 2014). Therefore, the essential information as well as the structure of the information (e.g. which information elements should be grouped together in order to avoid an extensive visual search of information within one task) is systematically derived. In this step, the connection between existing HF methods as task analysis and EID is established. These methods offer an engineering way for specifying the information to be shown on the HMI as well as the structure. For this, task analysis focuses on analysing the activities (tasks), the human operator conducts and grouping information elements according to these tasks (e.g. information that is needed in one task should be grouped together). EID shifts the focus from the human operator and his tasks to the work domain. Thus, a work domain analysis is conducted and structured according to an abstraction hierarchy. This is exemplarily shown in Figure 2 for a small example of a propulsion system: On the highest level, the functional purpose is specified. This is the main aim of the system analysed. In case of the small example, the main aim is that the propulsion system operates in safe mode. On the second level, the abstract function is described. This includes laws and principles existent in the underlying system e.g. principle of conservation of energy. After this, processes are described that are involved in the system (e.g. combustion) and the entities relevant in these processes (e.g. engine). On the lowest level, the physical form and parameters of these entities are described.

Before entering into step 2 activity of the Konect method, a decision has to be made which elements should be grouped together. Prior work has shown that information
can be perceived faster and remembered better in case certain visual attributes (e.g. colour, length, shape) are integrated in one integrated visual form (Ware, 2004). These integrated visual forms are called **glyphs**. To allow fast perception of all information elements at one glance, the Konect method applies the concept of glyphs. The decision on which information to integrate can be done on the basis of a task analysis tree or the abstraction hierarchy (as common in EID and exemplarily shown in Figure 2). E.g. to estimate the engine state with one glance, the information about engine revolutions, torque, temperature and oil pressure should be grouped together in one glyph. For each glyph, one idea box is specified in the upcoming step 2 – the **idea box specification**.

![Figure 2. Extract of work domain analysis for engine monitoring example.](image)

**Idea Box specification**

The second step is the idea box specification. The idea box is a core concept of the Konect method and provides a suitable format to collect all data relevant for informing the design process so that an optimization for fast and correct perception of chosen visual forms can take place. The idea box is a table with predefined columns. The columns are shown in Figure 3.

![Figure 3. Idea Box – Exemplary extract.](image)
In the first column the information is added in the idea box. This is taken from the abstraction hierarchy (e.g. revolutions, torque, temperature, oil pressure). When defining which visualization is most efficient the so-called insight is important. The insight specifies why the human operator needs this information (e.g. the human operator wants to perceive if the value is ok or the human operator wants to perceive the quantitative value). We specified a predefined list of insights. Each insight is mapped to a visual efficiency ranking containing visual attributes e.g. colour, length. The full list of insights and the mapping to efficiency rankings of visual attributes can be found in Harre et al. (2018). The rankings are derived based on well-accepted empirical knowledge: Cleveland and McGill (1984) presented an empirically verified accuracy ranking for visual attributes for quantitative data (e.g. position is more accurate perceivable than volume). This was later extended by Mackinlay (1986) for different data types (nominal, ordinal, quantitative). Both works focus on accurate perception. To include the time factor, we added knowledge about preattentive perception (Healey et al., 1996): Some visual attributes e.g. colour can be perceived preattentively – in millisecond range. All insights that have this time factor (e.g. perceive quantitative value (fast)) do only consist of visual attributes that are perceived preattentively. Most insights can be perceived on low level vision. Low level vision focuses on the perception of single visual attributes as colour or length. But there exist also insights in which various data elements are included simultaneously (e.g. compare). This insight is only perceivable on higher level vision. High level vision deals with the perceptual organization of single visual attributes. The efficiency rankings for these insights are derived based on prior empirically verified knowledge about high level vision (Kosslyn et al., 1990), (Rensink, 2000), (Rensink, 2002).

The specification of insights is important as different visualisations are different efficient for different insights e.g. if the human operator wants to perceive if the engine revolutions are ok, colour is an appropriate choice. In case he or she wants to perceive the quantitative value of the engine revolutions, a length might be more efficient. The insight specification allows a mapping to different efficiency rankings as described in the previous paragraph for possible visual attributes for the information. The full ranking with mapping to the different insights is shown in Figure 4.

As described previously, all information elements specified in the idea box should be grouped together in one glyph. To offer possibilities on how to integrate visual attributes in one integrated form, the last column – the combination – offers possibilities on how to combine visual attributes. This is done via the well-established gestalt laws (e.g. symmetry, proximity) (Ware, 2004).

In a last step, all entries in the idea box should be ordered according to their importance – starting with the most important information and ending with least important information. In this respect the importance of an information depends on the influence this parameter has on the achievement of the overall goal (functional purpose). This can be either derived based on the work domain analysis in which the impact of parameters on the functional purpose is estimated or within an interview with a human operator who directly states which information is more important than
another one for his or her monitoring task. This step is necessary to ensure that most important information is visualized most efficient with regard to human perceptual skills.

<table>
<thead>
<tr>
<th>Insight</th>
<th>Visual Efficiency Ranking</th>
<th>Scientific Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>perceive quantitative value</td>
<td>position(0), length(1), angle(2), slope(3), area(4), volume(5), density(6), color saturation(7), color hue(8)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1996)</td>
</tr>
<tr>
<td>perceive summary</td>
<td>length (0), slope (1), volume (2), color hue (3)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1986), (Healy et al., 1996), (Ware, 2004)</td>
</tr>
<tr>
<td>perceive quantitative value (fast)</td>
<td>position(0), color hue (1), texture (2), connection (3), containment (4), density (5), color saturation (6), shape (7), length (8), angle (9), slope (10), area (11)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1986)</td>
</tr>
<tr>
<td>perceive if value is ok</td>
<td>color hue (0), shape (1), length (2), slope (3), volume (4)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1986), (Healy et al., 1996), (Ware, 2004)</td>
</tr>
<tr>
<td>perceive ordered category/mode</td>
<td>position (0), density (1), color saturation (2), color hue (3), texture (4), connection (5), containment (6), length (7), angle (8), slope (9), area (10), volume (11)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1986)</td>
</tr>
<tr>
<td>perceive ordered category/mode (fast)</td>
<td>color hue (0), length (1), slope (2)</td>
<td>(Cleveland &amp; McGill, 1984), (Mackinlay, 1986), (Healy et al., 1996), (Ware, 2004)</td>
</tr>
<tr>
<td>perceive pattern</td>
<td>edges/depth/orientation at multiple scales (0), size/location (1), categorical relation (2), coordinate relation (3)</td>
<td>(Kosslyn et al., 1990), (Rensink, 2000), (Rensink, 2002)</td>
</tr>
<tr>
<td>perceive relationships</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perceive trade-offs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>compare</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perceive clusters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>perceive paths</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Mapping of insights to efficiency rankings.

**Glyph sketching**

The idea box specified in step 2, is now applied in step 3 to inform the design of glyphs. The designer of the HMI works with the idea box as follows: He or she starts with the first information (e.g. revolutions as shown in Figure 3). The information should be visualized once to estimate the quantitative value and additionally to estimate if the value is ok or not. According to the visual efficiency ranking, the most appropriate visual attribute for the first is length, and colour is the most appropriate for the latter. Similar applies for torque (length for quantitative value and colour for state (critical or not)), oil pressure and temperature. To create a glyph, these visual attributes have to be combined. Therefore, symmetry represents a suitable possibility to group all length variables in one visual form. This is exemplarily shown in Figure 5: The length of the quarter-circles represents the quantitative values of the parameters and the colour indicates if the parameter is
within its acceptable limits (ok) or if it exceeds these limits (critical). Thereby, the overall state of the engine is fast perceivable.

![Propulsion System Diagram](image)

**Figure 5. Example Glyph for engine.**

**Design Composition**

In the glyph sketching phase different glyphs are designed separately from each other. As described previously, fast perception strongly relates to preattentive perceptual skills of the human operator. In this regard, there exist some prerequisites to obtain so that critical states produce a preattentive recognizable “pop-out” effect:

It has to be ensured that glyphs – especially as they were designed separately – do not visually interfere with each other. For this the designer has to verify that the following guidelines are met (Harre et al., 2018):

*Consistency* Use the same visual attribute for the same kind of insight for similar important information elements.

*Simplicity in shapes* Choose simple shapes and visual forms, choose non-accidental visual forms with regard to orientation.

*Simplicity in Colours* Reduce colours for elements that do not carry any information besides from structuring the interface.

**Evaluation**

The Konect method was evaluated in different domains (first, in the automotive domain for a truck platooning use case and second, in the maritime domain for
vessel performance monitoring). The studies had a different focus: The truck platooning study was primarily used to compare designs derived with the Konect method compared to designs created with conventional methods, while the vessel performance monitoring study compared designs created after step 3 of Konect to designs after step 4 of Konect to further evaluate the guidelines for design composition. As conventional method, the human centred design process as described in the related work section was applied. This departs from the Konect method as no integration of specific prior knowledge about human perceptual abilities in a systematic step-by-step process is offered. Nevertheless, the human centred design process is a well-established standard that is widely applied for deriving HMI designs.

In both studies, a workshop was conducted as starting point. In this workshop, participants (HF experts with different background knowledge, UX professionals and designers) applied the method to create design solutions. In case of the truck platooning scenario, 5 designs have been created and in the maritime use case 12 designs were created. These designs were then tested in a laboratory study in front of the computer applying the following procedure: As starting point, the participants got a detailed explanation for the scenario and the different designs. This was followed by a training phase and a subsequent measurement phase. In the measurement phase (see Figure 6), the participants were asked to press the enter-Button as soon as they felt ready for estimating a situation. After pressing the button, a random design was shown presenting either a critical or a non-critical situation and a time-measurement started. The participants were asked to estimate as fast and accurate as possible if a shown situation was critical or not. As soon as they estimated the situation, they pressed the enter-Button and the time measurement stopped and the design disappeared. After this, the participant had to enter in a text-field if the situation shown before was critical or not. This answer has then been evaluated as correctness indicator for the percept. The study applied a within-subjects design to cancel out individual differences with regard to reaction times.

![Figure 6. Procedure for measuring mean RT and mean assessment failure.](image)

As these experiments were quite artificial, a driving simulator study was conducted additionally in which the participants were asked to drive in a simulated truck platooning scenario. Each participant drove about 40 minutes without having to react to any occurrence. After these 40 minutes drive, an ACC failure was simulated.
to which the participants react via braking to prevent an accident. For conducting this study, a between-subjects design was chosen (one group drove with a Konect design, the other group had a design created in a conventional user-centred design process). This was done to avoid a learning or accustoming effect for an ACC failure and to reproduce the real conditions with rare ACC failures as realistically as possible. The meters left to the truck in front when braking as well as the amount of accidents were measured.

Table 1. Overview for method evaluation.

<table>
<thead>
<tr>
<th>Method Evaluation</th>
<th>Participants</th>
<th>Results</th>
<th>Further Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truck Platooning (laboratory study)</td>
<td>33 (21f)</td>
<td>Conventional 14.59 (SD=7.9), Konect 5.43 (SD=3.77)</td>
<td>T-Test (p&lt;0.001) (Ostendorp et al., 2016a)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean RT 2122.86 (SD=951.65), Konect 1731.03 (SD=772.31)</td>
<td>p&lt;0.001</td>
</tr>
<tr>
<td>Truck Platooning (driving simulator study)</td>
<td>23 (12f)</td>
<td>Conventional 1.21 (SD=1.4), Konect 3.45 (SD=1.74)</td>
<td>T-Test (Friedrichs et al., 2016) p =0.002 &lt; 0.05</td>
</tr>
<tr>
<td>Vessel Performance Monitoring (laboratory study)</td>
<td>18 (14f)</td>
<td>Konect (after Step 3) 7.72 (SD=6.24), Konect (after Step 4) 3.09 (SD=2.73)</td>
<td>T-Test (Harre et al. 2018) p &lt; 0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean RT 1214.5 (SD=222.15), Konect 1041.5 (SD=174.47)</td>
<td>p &lt; 0.001</td>
</tr>
</tbody>
</table>

All results of the different evaluation studies are summarized in Table 1. The results of the Truck Platooning laboratory study revealed that conventional designs led to a mean assessment failure of 14.59% (meaning that in nearly 15% of the cases, a situation was falsely assessed as being either critical or non-critical, although the opposite was correct). This was reduced to 5.43% with Konect designs. The T-Test showed that this difference is significant with p<0.001. With regard to the time needed to assess the situation, the participants needed around 2122.86ms with the conventional designs, compared to 1731.03ms with the Konect designs, which is again a significant result (p<0.001).

The results measured in the laboratory study were confirmed in the driving simulator study. When participants brake in case the ACC failure occurred, the mean meters left to the truck in front were 1.21m for conventional designs compared to 3.45m with the Konect designs. This strongly indicates that the increase of accuracy and
time to perceive the situation lead to a better braking performance in critical situations. This is also reflected in the amount of accidents occurred. With the conventional designs every second participant had an accident, while with the Konect designs only 9.1% had a rear-end collision.

In the Vessel Performance Monitoring laboratory study, the influence of step 4 was analysed: The results showed that the mean assessment failure for estimating if a situation was critical or not was 7.72% with a design after step 3 compared to 3.09% after step 4. The reaction time for assessing the situation was reduced from 1214.5ms to 1041.5ms.

Summarizing, it can be said that the results revealed that Konects Design Composition step led to a significant improvement (p<0.001) and that Konect designs were significantly faster and more correct perceivable compared to non-Konect designs.

**Conclusion**

In the underlying paper, the Konect method was presented. Konect is a method for optimizing visualizations on monitoring displays for fast and accurate perception in the safety-critical domain. The method consists of four steps to be conducted: Information determination, Idea Box Specification, Glyph Sketching and Design Composition. For validation purposes, the method was applied in workshops in two different domains (automotive and maritime). The workshops revealed that Konect designs were significantly fast and more accurately perceived compared to designs created with conventional methods. In a driving simulator study, this led to a decrease of accidents from 50% to 9.1%.

**References**


Prediction of take-over time demand in conditionally automated driving - results of a real world driving study

Frauke L. Berghöfer¹, Christian Purucker¹, Frederik Naujoks¹, Katharina Wiedemann¹, & Claus Marberger²
¹Würzburg Institute for Traffic Sciences GmbH, ²Robert Bosch GmbH, Germany

Abstract

In conditionally automated vehicles the driver still functions as a fall-back, responsible to take back manual vehicle control at any time. This study examines how driver’s behaviour and characteristics may influence the time required for such a take-over. 34 participants took part in the study. During conditionally automated driving (SAE L3) in real traffic that was simulated using a Wizard-of-Oz vehicle, drivers engaged in Non-Driving Related Tasks (NDRTs) until a Request to Intervene (RtI) was issued. The assessed driver characteristics were: type of visual behaviour, self-rated sleepiness right before the RtI, attitude towards automated driving, previous experience with Adaptive Cruise Control (ACC) systems, individual reaction speed, as well as age and gender. Based on their visual behaviour, drivers could be classified into four groups with distinct visual gaze patterns. Further analysis indicated that a “medium off-road” visual behaviour was associated with increased take-over times. Additionally, individual reaction speeds and previous experience with ACC systems influence the take-over time. Sleepiness, as it occurred in the study, did not affect take-over time. Although further research is needed to verify the influence of driver characteristics on take-over times the current study gives important insights. Furthermore, the chosen Wizard-of-Oz approach extends previous research and helps to bridge the gap between simulator studies and naturalistic driving.

Introduction

Automation in automobiles improved greatly in recent decades. However, at the moment and in the near future, the driver will still need to take back manual vehicle control at system limits. It is therefore important to take a closer look at the take-over process and to examine possible factors involved.

The process of taking back manual vehicle control

The process of manual vehicle take-over after a period of automated driving can be subdivided into several phases (see Figure 1). While driving with a conditionally automated driving system, the driver’s state needs to remain compatible with system requirements determined by the system limits. That is, for example, the driver is
allowed to engage in Non-Driving Related Tasks (NDRT) such as reading a book or using a smartphone but needs to be receptive for a Request to Intervene (RtI). In this case the driver has to switch from the NDRT to the driving task, which is described as the “driver state transition”. In this phase, depending on previous activities before the RtI, the driver may need to redirect his or her attention back to the driving scene, may need to free his or her hands and put them back on the steering wheel and place the feet on the pedals. For these reasons, this transition may elicit varying demands regarding the sensory, motoric and cognitive state of the driver. Time demand for the driver state transition is expected to vary with the number and complexity of the underlying sub-processes. After the automated driving function is deactivated by the user the driver performs the required intervention, such as keeping the lane, initiating a lane change or braking. Take-over time is defined as the duration of the transition phase, that is, the time between RtI and the driver’s intervention (Marberger et al., 2017).

**Figure 1. Model of the transition process from automated driving to manual control (Marberger et al., 2017).**

**Factors influencing the take-over process**

Several factors are assumed to affect the take-over process. These include the complexity and criticality of the situation as well as the state of the driver right before an RtI is issued. Previous research for example reported a reduced level of take-over performance when the driver previously engaged in NDRTs (Radlmayr et al., 2014; Zeeb et al., 2016). Also sleepiness or fatigue are generally associated with poor vehicle guidance (Matthews & Desmond, 2002; Philip et al., 2005) which may also deteriorate take-over performance. Besides the state of the driver, various individual driver characteristics may also influence the take-over. Just a few studies considered such individual driver characteristics like trust in automation (Hergeth et al., 2016), individual reaction time (Körber et al., 2015), age (Körber et al., 2016; Warshawsky-Livne & Shinar, 2002), or gender (Warshawsky-Livne & Shinar, 2002). Another study by Zeeb et al. found that driver’s gaze behaviour with respect
to NDRTs can be considered as a stable factor as it revealed classified groups of gaze patterns. These gaze patterns in turn affected specific aspects of the take-over performance (Zeeb et al., 2015). As a limitation, most of the studies were conducted in driving simulators and just very few examined take-over situations in the real world. Moreover, the NDRTs used in automated driving studies often consist of standardized tasks, which are either cognitively or visually demanding, but often lack ecological validity.

The purpose of this study is to examine drivers’ behaviour in take-over situations in conditionally automated driving on the real road. Of particular interest are the effects of naturalistic non-driving related tasks on take-over time and the prediction of take-over time by the drivers’ gaze behaviour pattern and other driver characteristics.

Method

Test vehicle and automated driving system

To simulate a conditionally L3-automated system (SAE, 2014) on the real road, a Wizard-of-Oz setting was used in the study (Figure 2). When the driver activated the system, the co-driver (driver wizard) took over full longitudinal and lateral vehicle guidance via dedicated control units on the armrest, the seat and secondary foot pedals. The Human Machine Interface (HMI) was managed by the experimenter on the backseat (HMI wizard). The participants were fully informed about the characteristics of the simulated L3 system and the role of the driver wizard before the drive.

Figure 2. Automation related HMI (left); co-driver as driver wizard (middle); experimenter as HMI wizard on the back seat (right). Figure adopted from Naujoks et al. (2018).

The automation system could be activated and deactivated with two levers on both sides of the steering wheel. The participants were instructed to use only those two levers and not the steering wheel or the pedals for deactivation. The actual state (available, active, off) and the RtI were displayed on the cluster display and the Head-up Display (HUD), and were accompanied by acoustic signals. The displayed information for activation and deactivation of the system can be seen in Figure 3.
Procedure and NDRTs

At the beginning, the participant performed a familiarization drive to get used to the functions and the handling (e.g., the take-over procedure) of the automated driving system. The test drive on a German freeway was separated into six sections (see Figure 4). Every section started with the activation of the system. The participant was then instructed to engage in one of the five NDRTs. After approx. 5 or 15 minutes an RtI was issued and the driver had to take-over manual vehicle control. After a short interview the participant reactivated the system and started with the next NDRT block. Each test drive additionally consisted of a section of manual driving as an individual baseline. The order of the sections was pseudo-randomized and counterbalanced across the drivers. A more detailed description of the study procedure can be found in Naujoks et al. (2018).

Figure 4. Procedure of the test drive. Figure adopted from Naujoks et al. (2018).
prediction of take-over time demand in conditionally automated driving

Subsequent to the whole test drive, participants went through a short simple reaction time task on a computer where they were instructed to press the space bar as fast as possible when a displayed red traffic light turned green.

The chosen NDRTs affected the sensory, motoric and cognitive state of the driver in different ways and consisted of an audio listening task, a reading task, a backseat-searching task, playing Tetris on an installed tablet and a reference task (supervision of the vehicle environment). See Figure 5 for a more detailed description of the tasks; a more detailed assessment of the workload demands of the task can be found in Purucker et al. (2018).

Variables

For a general overview of the take-over time, several time components were measured such as the time to the first gaze to front, the time to free the hands, the time to turn the body to front, the time to pull and release the levers as well as the time the vehicle control is handed over by the driver wizard. For classification and regression analysis the take-over time (TOT) as dependent variable was then defined as time between start of the RtI message and start of deactivation procedure (pulling of both steering wheel levers). Furthermore, to assess the drivers’ gaze behaviour, the driver was recorded on video during the drives, and off-road glances were coded after the study. Off-road glances include glances at the NDRT, at other passengers or glances inside the vehicle within an interval of 60 s before the RtI. From the video coding, number and duration of off-road gazes were used as metric independent variables. For modelling the TOT, the drivers’ sleepiness shortly before the RtI, the attitude towards highly automated driving (HAD), previous experiences with ACC-systems, results of a simple reaction time task, as well as age and gender of the drivers were used as predictor variables. The driver’s sleepiness in a time interval of 10 s before the RtI was rated by trained observers using the 5-point rating scale for observer ratings, ranging from “0 – Not Drowsy” to “4 – Extremely Drowsy” (Appendix A; Höfling, 2017; Wierwille & Ellsworth, 1994). The attitude towards highly automated driving (HAD) was assessed via a questionnaire with 7-point likert.
scales. The given ratings were averaged for each driver and were contributed to the multiple regression as a metric variable. The items of the questionnaire can be seen in Appendix B. The individual reaction time was measured by a simple reaction time task in seconds. Results from five trials were averaged, outliers excluded. Sleepiness, attitude and individual reaction time were used as metric variables. Previous experience with ACC-systems, age and gender of the drivers were assessed also via a questionnaire and were handled as categorical independent variables.

Results

Sample

34 participants (6 females) with a mean age of 54 years (SD = 14 years) took part in the study. Of those, 21 had previous experience with ACC. Each drive consisted of 5 take-over situations. In one case, only 2 take-over situations were recorded due to technical problems, resulting in 167 take-over trials.

Effects of NDRT on take-over

Figure 6 shows the effects of specific NDRTs on different take-over time components. The values for “lever pulled” correspond to the definition of take-over time (TOT). The average TOT is the longest for the search and the reading task (~5 s), shortest for the reference task, and a bit longer for the audio book task (~3 s). The average TOT after playing Tetris is in between (~4 s). The mean time for the “first gaze to front” component is roughly equally long for the search task, the reading task as well as the Tetris task, whereas the time to free the hands is the longest for the reading task an the shortest for the Tetris task.

Figure 6. Average duration of different take-over time components in combination with specific NDRTs.
Classification of drivers’ gaze behaviour

To classify the drivers’ visual gaze behaviour a hierarchical cluster analysis with Ward’s minimum variance method (Ward, 1963) was conducted using the maximum duration of off-road glances in combination with the number of off-road glances for each trial and for each driver 60 s before the RTI. Figure 7 shows the results of the analysis. Considering the total number of trials the analysis revealed four groups with different gaze patterns described as “long off-road”, “medium off-road”, “short and often off-road” and “short and seldom off-road” (Figure 7 left panel). When TOTs are averaged across all NDRTs in a separate analysis, still four groups emerge, but with less variance (Figure 7, right panel). To examine whether the gaze behaviour is indeed an individual driver characteristic and not just determined by the specific NDRT, the results from the initial cluster analysis were plotted separately for each NDRT in Figure 8. It can be seen that specific NDRTs influence the driver’s gaze behaviour and, therefore, the classification based on maximum duration of off-road glances and number of off-road glances. While the reference and the listening (audio book) task provoke mainly “short and seldom off-road” glances, the search task, the reading task and the Tetris task tend to involve at least three of the classified groups. Note that this pattern seems to relate to the observed TOTs.
To examine how the TOT can be predicted from the observed gaze patterns and the obtained driver characteristics, a first multiple regression model was calculated with the classified gaze behaviour, the attitude towards automated driving systems, previous experiences with ACC-systems, the individual reaction time as well as age and gender of the drivers were additionally used as predictors. 18 take-over trials were excluded because drivers were not engaged in NDRTs at the time of the RtI, resulting in 149 valid take-over trials. Results are shown in Table 1. The total model produces an adjusted $R^2 = .182$, ($F(9,148) = 4.669$, $p < .001$). Significant predictors are medium off-road gaze patterns, the result of a simple reaction time test and previous experience with ACC systems. Neither the rated sleepiness, nor the attitude towards highly automated driving, age and gender reach significance. Concerning the $\beta$-weights of the significant predictors, only previous experience with ACC systems has negative impact on take-over time, indicating that drivers with experience in ACC systems need shorter time to take back manual control than drivers without previous experience. Average reaction time and medium off-road gaze behaviour, however, show positive $\beta$-weights, that is, longer individual reaction times and medium off-road gaze behaviour lead to longer take-over times. Among those three significant predictors, medium off-road gaze behaviour shows the largest $\beta$-weight and thus, the strongest impact.

Table 1. Result of the multiple regression with take-over time as dependent variable.

<table>
<thead>
<tr>
<th>DV: Take-over time</th>
<th>Non-standardized coefficients</th>
<th>Standardised coefficients</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.319</td>
<td>.034</td>
<td>39.162</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Short &amp; often off-road gazes</td>
<td>.037</td>
<td>.066</td>
<td>.063</td>
<td>.566</td>
</tr>
<tr>
<td>Medium off-road gazes</td>
<td>.205</td>
<td>.055</td>
<td>.394</td>
<td>3.751</td>
</tr>
<tr>
<td>Long off-road gazes</td>
<td>-.024</td>
<td>.062</td>
<td>-.042</td>
<td>-.385</td>
</tr>
<tr>
<td>(Constant)</td>
<td>.775</td>
<td>.360</td>
<td>2.151</td>
<td>.033</td>
</tr>
</tbody>
</table>
Discussion

The aim of the study was to measure the effect of specific NDRTs when human drivers are requested to take-over manual control from conditionally automated driving. This was pursued by classifying the drivers’ gaze behaviour and by examining, which of various specific driver characteristics additionally influenced TOTs in a multiple regression.

In a first step, the measured time components of the control transition were regarded separately for each NDRT. The TOT was the longest for the search task and the reading task and the shortest for the reference task. The time the drivers needed to redirect their gaze back on the road was comparable for searching, reading and playing Tetris. Regarding the time the drivers needed to free their hands, it was the longest for the reading task, followed by the search task and the Tetris task.

In a second step, we analysed a predictive model that (by intention) does not contain the NDRTs themselves but rather stable driver characteristics or measurable criteria of driver behaviour. To this end, a cluster analysis was conducted for the analysis of gaze behaviour, indicating four groups of gaze patterns from “long off-road” to “short and seldom off-road”. As opposed to Zeeb et al. (2015), the results here revealed a forth group of “short and seldom off-road glances” probably due to the audio listening task we used as a non-visual demanding NDRT. Further analysis showed that the found gaze patterns are also determined by the NDRT the driver performed during automated drive. “Short and seldom off-road” glances primarily occurred in the reference and audio listening task as those two tasks require little visual attention.

A multiple regression model was calculated based on the found gaze patterns and further driver characteristics in order to examine their influence on TOT. Results showed three significant predictors: “Medium off-road” gaze behaviour, individual reaction speed and previous experiences with ACC systems. Drivers showing this type of gaze behaviour and drivers with lower individual reaction speeds needed longer to take-over manual vehicle control. Previous experience with ACC systems, by contrast, resulted in decreased take-over times. All in all, the goodness-of-fit of
The regression model (adjusted $R^2 = .182$) is rather low, making it difficult to use the model for prediction purposes.

The data did not reveal significant effects of attitude to highly automated driving, sleepiness, age or gender of the driver. Two results are particularly interesting and call for further research. First, although the group of medium off-road gaze behaviour leads to increased TOTs, the group with long off-road gazes does not. According to the data of this study the link between off-road gazes and TOTs does not seem to be linear as opposed to the findings of Zeeb et al. (2015). This observation may result from behavioural compensation: Drivers who do not look at the road for prolonged periods of time may try to compensate this gaze behaviour by showing faster responses to the RtI. At the same time, drivers with medium off-road gaze behaviour may not yet perceive the need to compensate. The second observation result refers to the sleepiness of the drivers. The variance of the observer ratings was rather low which makes it difficult to find significant effects. Generally, the study design does not seem to be appropriate to provoke sleepiness. So, more research is needed to examine the influence of sleepiness and to further verify the influence of other driver characteristics on TOTs.

The results of this study refer to low demand take-over scenarios. Drivers were not required to act under time pressure or in safety-critical scenarios. Although this type of test scenario may be representative for most take-over scenarios in real life it may not be suitable to investigate the human performance limits to safely take-over. Still, the current study gives important insight in individual driver aspects and, hence, individual TOTs. Furthermore, the naturalistic design used here extends previous research and help to bridge simulator studies and real-world driving.

**Acknowledgements**

This work results from the joint project Ko-HAF - Cooperative Highly Automated Driving and has been funded by the Federal Ministry for Economic Affairs and Energy based on a resolution of the German Bundestag.

**References**


prediction of take-over time demand in conditionally automated driving  79


Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (in press). Non-critical state transitions during conditionally automated driving on German freeways: Effects of non-driving related tasks on take-over time and take-over quality. Human Factors.


### How drowsy is the driver?

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Not Drowsy</strong></td>
<td>- Appearance of alertness</td>
<td>- E.g., prolonged glances at a fixed position, slower eyelid closures</td>
<td>- Behaviour against drowsiness:</td>
<td>- E.g., eye lid closures of 2-3 s, rolling movement of the eyes, no focusing, decreased face tone</td>
<td>- E.g., falling asleep, very long eye lid closures (at least 4 s)</td>
</tr>
<tr>
<td></td>
<td>- E.g., normal face tone, normal fast aye blinks, short ordinary glances</td>
<td>- Still, alert enough to be fit to drive</td>
<td>- E.g., Rubbing the face/eyes, scratching, facial contortions, moving restlessly in the seat</td>
<td>- Lack of apparent activity or large isolated movements, e.g., large correction to steering or reorienting the head from a leaning or tilting position.</td>
<td>- Prolonged periods of lack of activity, large punctuated movements as a transition in and out of intervals of dozing.</td>
</tr>
<tr>
<td></td>
<td>- Occasional body movements or gestures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Adapted from Höfling, 2017; Wierwille & Ellsworth, 1994)
B) Items of the questionnaire to assess the attitude towards automated driving.

<table>
<thead>
<tr>
<th></th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Slightly Disagree</th>
<th>Neutral</th>
<th>Slightly Agree</th>
<th>Agree</th>
<th>Strongly Agree</th>
<th>Don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>I am generally positive about such automated vehicles described above</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated driving will increase driving safety of the driver.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated driving will increase traffic safety.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated driving is a meaningful and useful development.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would hand over control to the automated vehicle.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I think, automated vehicles that actively intervene in driving are prone to errors and are dangerous.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I would not like to drive such an automated vehicle.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Drowsiness and fatigue in conditionally automated driving – Towards an integrative framework

Jonas Radlmayr1, Anna Feldhütter1, Alexander Frey1, Oliver Jarosch1, Claus Marberger2, Frederik Naujoks5, Veronika Weinbeer6, & Klaus Bengler1
1Technical University of Munich, Chair of Ergonomics, 2Federal Highway Research Institute (BASt), 3BMW Group, 4Robert-Bosch-GmbH, 5Center for Traffic Sciences (IZVW), Universität Würzburg, 6Audi AG

Germany

Abstract

The introduction of conditionally automated driving (CAD) will change the role of the driver fundamentally. Drivers can be engrossed in non-driving related tasks (NDRTs) or fulfill a new role of a passive passenger while the automation executes the dynamic driving task (DDT). In both cases, fatigue and/or drowsiness could impair drivers’ availability for an upcoming take-over if the automation reaches a functional limit. In that case, CAD relies on drivers to act as a fallback ready user to take-over control manually. We propose an addition to the framework proposed by Marberger et al. (2017) with a more detailed look on the constructs drowsiness and fatigue and their interdependencies within the framework and additional aspects of driver availability. This work also provides a comprehensive overview on relevant literature regarding these constructs in general and the current state of research concerning CAD. Furthermore, it includes selected findings and recommendations from the German joint project Cooperative Highly Automated Driving – Ko-HAF. We conducted 14 experiments of prolonged automated driving periods, to reveal effects of drowsiness and fatigue on drivers’ availability. This work integrates the lessons-learned and the development of advanced methods and procedures derived from the extensive empirical expert knowledge.

Introduction

Increasing vehicle automation (SAE International, 2016), raises new demands on drivers’ abilities depending on the particular level of automation. In SAE level 2 (partial driving automation), 3 (conditional driving automation) and level 4 (high driving automation), long periods are imaginable, in which the automation executes dynamic control of the vehicle. Nonetheless, drivers will still have the need or possibility to drive manually. Especially in SAE-levels 2 and 3, humans represent the fallback level, and thus must be able to take over control in case the automation reaches a system limit. Therefore, current research projects focus on questions concerning the driver’s state and availability for such a take-over (e.g. on SAE-level: Cooperative, highly automated driving – Ko-HAF, 2015 – 2018). Estimating the
driver’s availability is necessary in SAE-level 3 to decide if the human driver can act as a fallback ready user (SAE International, 2016) in a defined time span. For level 2, drivers must remain in the driving loop and monitor the automation and the surroundings. In case a system limit is met, drivers are expected to continue driving manually right away. Consequently, problems regarding the driver’s attention (driver’s vigilance to execute the monitoring task in SAE-level 2), driver’s fatigue, or driver’s drowsiness in SAE-level 3 are prominently discussed in current research (Radlmayr et al., 2014; Körber et al., 2015; Miller et al., 2015; Neubauer et al., 2012).

In addition, drowsiness and fatigue are well known and relevant causes for road accidents in manual driving. Referring to the report on drowsy driving from the National Center for Statistics and Analysis (2017), 3662 fatal crashes occurred between 2011 – 2015 due to drowsy driving or an inadequate driver’s state. In another report published by the National Sleep Foundation in 2011, an alarming proportion of 52% of the participants indicated that they had driven drowsy. 6% of the participants reported to do so three times a week.

This highlights the challenges of designing driving automation and the necessity to focus on drowsiness and fatigue especially in SAE-level 3, conditionally automated driving (CAD): the automation level itself and the resulting paradigm shift (no need to control or monitor the system/automation/surroundings) potentially motivate higher levels of drowsiness and fatigue even though the human performance in take-overs is highly safety-critical.

Drowsiness and fatigue: literature, definitions, understanding

Starting with SAE-level 2, psychological constructs such as vigilance, fatigue, drowsiness, sleepiness and how they are interlinked with the overall driver state, play an important role for all efforts in improving road safety. In SAE-level 2, the driver needs to monitor the vehicle’s guidance as well as the surrounding environment, whereby the automation requires the human to intervene immediately, if a system limit is met. However, basic psychological research indicates that these vigilance tasks are very stressful and accompanied with human failures (Parasuraman & Riley, 1997; Warm, Parasuraman, & Matthews, 2008). Laboratory experiments show that humans miss stimuli after about 15 minutes using standardized vigilance tasks (Teichner, 1974). One possible explanation could be an increasing fatigue level, caused by monotonous tasks transmitting so called passive task related fatigue (May & Baldwin, 2009).

As mentioned above, in SAE-level 3, drivers must still be able to take over control safely in case of a request to intervene (RtI) (Marberger et al., 2017). Hence, this paper focuses on the transitional state between wakefulness and sleep, which might impair take-over performance. Johns (1998) defined this state as “drowsiness”, May and Baldwin (2009) defined this state as “fatigue”, including active (increased workload/overload, intensive secondary/non-driving related task (NDRT) engagement), passive (underload conditions, monotonous/extended driving, automation effects) and sleep-related fatigue. In addition, the terms drowsiness, sleepiness, fatigue, and tiredness are often used synonymous (for example some
researches investigate the influence of fatigue by using a sleepiness scale). As the sources and reasons of these constructs often overlap considerably, we use these terms drowsiness and fatigue interchangeably in the following and focus on the influence of different causes of their development, on the possibility to detect changes in them and the general driver state respectively and on the consequences on take-over performance. Nonetheless, we provide a short overview on the most prominently used definitions, which distinguish these constructs to allow a more comprehensive understanding.

Johns (2007) defines the constructs drowsiness and fatigue according to the Shorter Oxford English Dictionary. Drowsiness is a state of being "inclined to sleep, heavy with sleepiness, half asleep, dozing", being synonymously with the adjective "sleepy" (Johns, 2007). Consequently, the state of being drowsy describes the interval between wakefulness and sleep, being the "fluctuating state that shares some of the characteristics of alert wakefulness and some of sleep" (Johns, 2007). In addition, fatigue is being understood as "weariness resulting from bodily or mental exertion", which is synonymous with "tiredness" (Johns, 2007). Consequently, the two constructs drowsiness/sleepiness and fatigue/tiredness are clearly separated from one another, allowing a clearer attribution of effects, measures and consequences. Fatigue is a "subjective state of weariness" (Johns, 2007) and can be relieved by rest, whereas drowsiness is understood to be only relieved by sleep. This is in contrast to e.g. the understanding of May and Baldwin (2009), where fatigue is understood to contain both task-related fatigue and sleep-related fatigue where the latter can be understood to represent drowsiness/sleepiness in Johns (2007) and is defined very similar in both definitions.

In addition to drowsiness and fatigue, attention in general and vigilance need to be considered in the context of automated driving. While vigilance is of higher interest for SAE-level 2, since the driver needs to monitor the automation and the surroundings, the construct of attention and attention selection of drivers needs to be considered in SAE-level 3 as well. Drivers can engage in non-driving related tasks in SAE-level 3, potentially managing their level of arousal, subsequently influencing their individual levels of drowsiness and fatigue. Markkula, Victor, and Engström (2017) provide a conceptual model of attention. The concept includes a detailed look on attentional impairments, differentiating between impairments originating from being asleep versus being drowsy but only a more general view from attentional impairments from fatigue-related causes. In addition, the model focusses on manual driving, subsequently including distraction. In the context of CAD, this distraction becomes engagement into e.g. NDRTs and rather results in automation effects on a potential take-over in contrast to a degraded performance resulting from distraction.

Concerning the relation of vigilance and drowsiness and fatigue we refer to Vogelpohl et al. (2016). Shortly summarized, vigilance is defined as a condition of general alertness, which is additionally characterized by mental arousal, affected by physical as well as mental fatigue (Vogelpohl et al., 2016). Combining both constructs (vigilance and fatigue) – and not focusing on sleep deprivation – Vogelpohl et al. (2016) explain that factors like task difficulty or the (driving) environment cause a high workload condition (immediate effect of fatigue; active
mechanism) or a low workload condition (effect of vigilance, resulting in fatigue; passive mechanism).

Focusing more on the effects of drowsiness and fatigue on human performance in the driving context, previous studies like Jewett et al. (1999) reported deteriorated reaction times in a psychomotor vigilance task in reference to hours of sleep. When the amount of sleep increased, the reaction times of the participants improved. Other aspects like diminished capacity for work, disinclination to apply effort to a task, impaired personal efficiency and subjective discomfort are also linked to fatigue (Matthews & Desmond, 2002). Impaired driving performance and deteriorated reaction times due to drowsiness and fatigue could be demonstrated in driving simulator studies as well (Philip et al., 2005). Neubauer et al. (2012) suppose that factors provoking active task-related fatigue can be minimized in conditionally automated driving as the system is performing the driving task. For example, the system will also be able to drive in high traffic-density conditions and a previous, in manual driving, distracting secondary task will turn to a non-driving related task in automated driving. In manual driving such factors provoked task-related fatigue. Contrary, different effects arise when considering passive task-related fatigue. Referring to the fatigue model of May and Baldwin (2009), factors that can lead to this form of fatigue are directly connected to automated driving. In their model they name increasing monotony and automation as examples for origins for passive task-related fatigue. Increasing automation means a reduction of tasks for the human driver which will consequently lead to a more monotonous situation when driving automated.

In conclusion, this chapter illustrates the abundance of theoretical classifications of different concepts such as attention, vigilance, drowsiness, sleepiness, fatigue, arousal, etc. While all these models add to a better understanding of the interdependencies between these constructs, only very few offer empirical backup or are specifically focussing on conditionally automated driving. Consequently, we will use drowsiness and fatigue in this work as describing various constructs relating to conditionally automated driving. The combination of drowsiness and fatigue is not “unified” to one new or existing term or definition in order to convey the difference between “sleep/drowsiness-related effects” and “fatigue-effects”, in line with Johns (2007) and May and Baldwin (2009). The framework by Marberger et al. (2017) offers the theoretical understanding important to this work: in conditionally automated driving, the transition from automated driving to manual driving and consequently the transition of a current driver state to a target driver state (able to take-over and continue driving manually) is of utmost interest. Thus, we will focus on effects from e.g. prolonged automated driving or monotonous or engaging non-driving related tasks on “drowsiness and fatigue” of drivers with respect to how this changes or effects can be measured and how they are affecting an upcoming take-over and drivers’ performances.

**Methods and metrics to measure drowsiness and fatigue**

Robust and valid measures to record humans’ drowsiness and fatigue level are required. For the assessment of drowsiness and fatigue in previous studies different
methods were used. Frequently used methods can be distinguished in subjective, physiological and eye-lid based measures. A common questionnaire for the assessment of subjective drowsiness and fatigue is the Karolinska Sleepiness Scale (KSS; Åkerstedt & Gillberg, 1990). With the help of a nine-point Likert scale participants have to state their level of drowsiness and fatigue. If researchers ask participants to specify their level with self-reported scales like the KSS, participants possibly are reactivated at this moment (Schmidt et al., 2011), preventing a realistic progression. In addition, humans are often not able to estimate their true drowsiness and fatigue level (Schmidt et al., 2009). A possible solution is the usage of psychophysiological data. Typical parameters are heart-rate, heart-rate variability (Manzey, 1998) and electroencephalography (EEG; Simon et al., 2011). Another method for the assessment of drowsiness and fatigue are eye-lid based metrics. Percentage of eye-lid closure over time (PERCLOS; Wierwille & Ellsworth, 1994), mean blink duration (Schleicher et al., 2008) and blink-frequency (Dinges & Grace, 1998) are described to be good indicators for drowsiness and fatigue. In general, eye-tracking methods are a useful tool to estimate humans’ availability. In contrast to subjective rating tools (e.g. KSS, etc.) and intrusive physiological methods such as EEG, researchers can display the overall fatigue evolution process without disturbing the participant in case the eye-tracking is realized without a head-based eye-tracker.

The so called alpha spindles (EEG-based) are assumed as neuronal correlates of human fatigue levels (Simon et al., 2011). It is recommended to validate this method by analyzing participants’ video data for example. Wierwille and Ellsworth (1994) propose an opportunity to rate the level of drowsiness and fatigue by video data using specific cues. While individual cues are not sufficient to detect drowsiness and fatigue, a higher occurrence probability is highly correlated with higher levels of drowsiness and fatigue. For example, a higher frequency of head movements such as rolling or nodding is a cue for an onset of drowsiness and fatigue (Senaratne et al., 2007). In addition, hand movements towards the face and posture adjustments have been found to be valid cues (Anund et al., 2013; Senaratne et al., 2007).

Concluding, the various methods and metrics of detecting drowsiness and fatigue all present individual strengths and weaknesses but our experience in Ko-HAF has shown, that eye-tracking methods present the most promising tool. While head-based solutions typically offer a higher quality of tracking, these solutions are not qualified for future use in commercial vehicles due to obvious reasons. Concerning specific metrics, we found eye-lid-based metrics to be most valid for detecting drowsiness and fatigue.

**Recommendations for the empirical setup**

Driving simulators are a prominent tool to display any situation with high experimental control and any automated driving function. The simulation can induce a situation which prompts participants to become tired without representing a hazard. However, driving simulators typically work with projectors in dark environments. This effect of lighting could affect valid measurements of driver’s fatigue (Åkerstedt et al., 1979). Contrary, studies in real-driving environments deal with low internal validity, because reproducibility is very weak even though the
surroundings are real. For more details on the validation of selected driving simulator scenarios in comparison to a real driving scenario, please refer to Frey (2016). A solution balances advantages and disadvantages concerning internal and external validity problems. Test track conditions can perform very satisfactorily, if experimenters take the monotony, daytime and participant’s instruction into account.

To realize a realistic impression of vehicle automation on a test track, a so-called Wizard of Oz test vehicle can be used (Dahlbäck, Jönsson & Ahrenberg, 1993). There are some differences in vehicle concepts, but they all look to create an illusion of a pre-defined automated driving setting. In complex and highly immersive implementations, participants sit on the driver’s seat and are able to drive manually. Additionally, the wizard is able to take over longitudinal and lateral control of the vehicle by using a second set of pedals and a steering wheel in a darkened compartment behind the participant (invisible for him/her). This adds the possibility of automated driving for participants. All levels of vehicle automation can be reproduced by fitting the compartment with the necessary hardware, cameras and displays for the wizard. This allows for a very broad and powerful empirical setup, since different levels of automation can be realized by instructed and trained “wizard-drivers” without the need to use a dedicated technical system. In less immersive Wizard of Oz vehicles, an additional steering function and pedals from the passenger’s seat is sufficient for the research question and setup. In this case, participants are informed about the wizard but can still experience effects of automated driving, e.g. engaging into non-driving related tasks (Marberger et al., 2007). The usage of Wizard of Oz vehicles on test tracks is very advantageous: Real vehicle dynamics with arbitrary reliability and functional scope of an automated driving system are combined with moderate experimental control and realistic environmental influences (e.g. weather conditions, lighting, noises, etc.). This led to several experiments using Wizard-of-Oz-vehicles of different maturity concerning the displayed automation. The experiments focused mainly on the development of drowsiness and fatigue in real driving environments to allow a proof of concept of corresponding experiments in simulators. In many experiments on drowsiness and fatigue in simulators, the simulator environment is portrayed as limiting the validity of the methodical approach: participants might be more prone to allowing drowsiness and fatigue due to the different risk setting. Even though the experiments in real traffic were unable to compare safety critical take-overs, results indicate that the development of drowsiness and fatigue does not show differences between simulator and real traffic surroundings in the Wizard-of-Oz approach. This adds to the validity of results from simulator experiments on the influence of drowsy and fatigued drivers in take-overs. Different driving scenarios can lead to a development of different fatigue patterns. In the German, nationally funded project Cooperative, highly automated driving – Ko-HAF, several partners conducted experiments to induce drowsiness and fatigue by using monotonous setups, e.g. a prolonged duration of automation without NDRTs or highly monotonous NDRTs like the Ppd-task (Jarosch & Bengler, 2019). Regarding a similar, monotonous setup in a wizard-of-Oz vehicle, results from simulator studies in comparison with EEG data from a Wizard-of-Oz setup indicate that the onset and development of drowsiness and fatigue can be compared regardless of the authenticity of the surrounding environment (Wizard-of-Oz vs. simulator).
Empirical work on drowsiness and fatigue in Ko-HAF

In Ko-HAF L3 automation features were of high interest. The human factors work package focused on driver state in general and drowsiness and fatigue specifically.

During the project duration from June 2015 to November 2018, a total of 33 empirical studies, with 1723 participants, in over 1750 hours, resulting in 30 publications were conducted, focusing on the topics automation effects (including effects of NDRTs and drowsiness and fatigue) and optimizing the transition from automated to manual driving. This work provides a comprehensive overview of the studies on drowsiness and fatigue that were already published. Papers in press or in planning are in line with the presented findings and conclusions here.

Weinbeer et al. (2017) used a right-hand-drive vehicle (participants were never able to intervene in the real driving process) acting as a Wizard-of-Oz automated vehicle to investigate the suitability of different measures to induce drowsiness and the influence of drowsiness on take-over time aspects. The experiment was conducted in real traffic on a stretch of German Autobahn. Two experts evaluated participants’ drowsiness level (DL) during the test drive. Depending on the individual level of drowsiness, a request to intervene (RtI) was triggered. There was no statistically significant influence of drowsiness on take-over time aspects. Results suggest that drowsy drivers are still able to perceive and react to a request to intervene. However, the take-over scenario as a known and prominent factor influencing take-over performance was of overall very little complexity and criticality. In addition, any quality metrics such as accelerations could not be measured using this specific Wizard-of-Oz approach. The experiment showed that the methodological approach can induce and enhance drowsiness during simulated automated driving.

Further, Weinbeer et al. (2018) looked at different options to counteract drowsiness. Participants assessed various options of a driver-state related strategy and of a system-based strategy before and after a tiring simulated automated drive. Results showed that reducing the maximum speed was the best rated system-based option and that a specific use of NDRTs was the driver-state related option that was most widely supported. The work provides initial insights into the acceptance of various strategies for managing drowsiness during automated driving from a user perspective.

Jarosch et al. (2017) investigated the difference between activating and monotonous tasks during CAD in a driving simulator study. In addition, the development of drowsiness and fatigue was measured using the metric percentage of eye-lid closure (PERCLOS) and the self-report Karolinska Sleepiness Scale (KSS). Results suggest that fatigue can be caused through a monitoring task in CAD. PERCLOS could be confirmed as a valid parameter for detecting fatigue in CAD. Furthermore, passive task related fatigue caused by a 25 min monotonous monitoring task does not affect the drivers’ take over capability negatively in this setting. The data from this experiment was compared with a follow-up experiment.

In this driving simulator study Jarosch and Bengler (2019) compared the effects from prolonged automated driving with the engagement into NDRTs. In case
participants were driving 25 minutes and were engaging in an activating NDRT, results suggest that participants executed a lane change necessary to pass the system limit in the take-over situation. In case the NDRT was monotonous, less participants would execute the lane change but rather come to a full stop in their ego lane. Contrary, after a 50 min automated drive and engaging in the same two NDRTs, more participants reacted by stopping on their ego lane. Again, more participants executed a lane change maneuver when engaging in the activating task compared to the monotonous task. In the 50 min condition six participants lost control of their vehicle during the take-over compared to one participant in the 25 min condition. These differences in the take-over performance are significant, suggesting that drivers can show major decrements the take-over performance if they are driving conditionally automated for longer periods of time and cannot engage in any activating activity.

Weinbeer, Muhr, and Bengler (2019) investigated the effects of different NDRTs and their effects on drowsiness and fatigue. After a relaxation phase, the sample was divided into three groups that were given different non-driving-related tasks (a dictation, a sport activity and a relaxation task). No participant of the dictation- or sport-activity-group exceeded a high level of self-reported drowsiness after the reactivation phase. These results show that the use of NDRTs has the potential to represent a suitable option for managing driver drowsiness. Concluding, activating NDRTs have the potential to counteract high levels of drowsiness.

Feldhütter, Kroll, and Bengler (2018) also looked at the development of drowsiness and fatigue and their potential impact on driving performance in take-overs during the transition from conditionally automated driving to manual driving. The effect of fatigue on the take-over performance was examined in a driving simulator study with 47 participants assigned to two conditions: fatigued or alert. In the alert condition, the desired driver state was promoted by specific measures (e.g. daytime, caffineinated beverages, physical exercise). In the fatigued condition, the take-over situation was triggered once participants reached a certain level of drowsiness and fatigue. Two trained, independent observers assessed the drivers’ state with the support of a technical fatigue assessment system based on objective eyelid-closure metrics. In the alert condition, participants drove conditionally automated for a fixed 5-minute period. Since the request to intervene was issued only when a specific level of drowsiness and fatigue was reached by both the raters and the objective technical systems, actual periods of CAD differed in their duration. Results showed no significant difference between participants’ take-over times in the two conditions. However, drivers in the fatigued condition showed more signs of stress and burden during the take-over situation. Additionally, their take-over behaviour was assessed to be less confident behavior when coping with the situation. This behavior may negatively affect the transition from conditionally automated driving to manual driving in more complex situations.

Concluding, the experiments demonstrated that it was feasible to induce drowsiness and fatigue during CAD and measure the change in driver state by using the appropriate methods and metrics. The development of drowsiness and fatigue showed high individual differences with little to no effects on the take-over
performance. Nonetheless, extreme levels of drowsiness and fatigue might be seldom during the CAD but should still be avoided to avoid critical take-over behaviour in more complex situations or even the possibility of unsuccessful take-overs. Activating NDRTs showed a high potential to counter drowsiness and fatigue and potentially offer a way of prolonging the periods of conditionally automated driving.

Towards an integrative framework

While the theoretical and empirical work on drowsiness and fatigue in manual driving provides a very well-researched basis for CAD as well, the overview in this work underlines the necessity for a more integrative view. While various definitions and theoretical considerations allow a conceptual understanding of drowsiness and fatigue and connected constructs, in CAD the focus is shifting to the effects of drowsiness and fatigue on the driver state and the take-over performance. In case NDRTs or prolonged automated driving leads to a change in driver state in general and the effects can be attributed to an onset of drowsiness and fatigue, more important than a clear assignment to an individual construct is the effect on the current driver state. Regarding a target driver state necessary for a successful transition and succeeding manual driving performance in and after a take-over, the connection between these effects and what happens during the automated drive is of utmost importance. In addition, empirical results from Ko-HAF underline the complexity of quantifying this connection, since results indicate great inter- and intraindividual differences between participants. Nonetheless, results show that it was possible to induce drowsiness and fatigue in test situations without sleep deprivation but in the context of CAD. Driver state changes could be detected by using several metrics and methods under these experimental conditions and allow valuable recommendations for future research. In addition, results indicate while driving with conditional automation, extreme levels of drowsiness and fatigue (drivers close to falling asleep) must be avoided. Concerning higher levels of drowsiness and fatigue, clear and consistent effects on take-over behavior and performance could not be found. More importantly, based on the detection of high levels of drowsiness and fatigue, countermeasures (e.g. a specific offer of NDRTs) can be initiated to avoid or to postpone such extreme driver states. Concluding, most work on drowsiness and fatigue in the context of conditionally automated driving focusses on the safety-critical take-over process and potential problems for a successful take-over during automated driving. Concerning the general acceptance of automated vehicles in addition with the user experience necessary for the technology to be used by future customers, we recommend additional research being conducted focussing not only on safety-critical topics but comfort-related questions as well.

Acknowledgments

This work results from the joint project Ko-HAF - Cooperative Highly Automated Driving and has been funded by the Federal Ministry for Economic Affairs and Energy based on a resolution of the German Bundestag.
References


Should I stop or should I go? Evaluating Human Capabilities of Cooperative Manoeuvre Strategies on Automated Driving

Kristin Mühl, Sandrina Büsch, Franziska Babel, Philipp Hock, & Martin Baumann
Ulm University
Germany

Abstract
The cooperative driving framework in combination with adaptive automation provides a suitable solution for overcoming legal and system limitations by sharing responsibilities and tasks between driver and automation. The evaluation of human capabilities of partial take-overs in traffic is required as foundation for this approach. A basic cooperative manoeuvre strategy was tested in a driving simulator using a traffic light scenario. Participants drove on a rural road either with assisted or partial automation while occasionally approaching red or green traffic lights. For half of the trials they received a cue, indicating an upcoming traffic light. Drivers (N=24) had to decide for the appropriate oncoming behaviour of the automation while not or only partly taking operational control of the automated car. Results of behavioural and gaze data showed that attentive drivers were capable to perform cooperative manoeuvres. Due to transparent and distinct system behaviour, drivers were successful in acting cooperatively in low levels of automation. Implications for conditional or high automation, more complex driving tasks and distracted driving are discussed.

Introduction
The dissemination of automated driving increases continuously (Lu, Happee, Cabrall, Kyriakidis, & de Winter, 2016) even though different challenges exist (Casner, Hutchins, & Norman, 2016). The current highest level of automation available on the market is partial automation (level 2 by SAE, 2016) as the human driver still has to monitor the driving task. The human driver is responsible for the observation of the surroundings and serves as fall-back solution as the driver has to take-over control as soon as the automation is not able to handle the current situation. This kind of task sharing between driver and partial vehicle automation requires that the driver is able to construct an adequate situation representation including the state of the automation at all times. Research has shown that driving with assistance systems such as adaptive cruise control (ACC) or ACC and lane keeping assist, increased drivers situation awareness and decreased workload compared to manual driving (de Winter, Happee, Martens, & Stanton, 2014). However, in case of take-over situations several studies showed the detrimental
impact of driving with partial or conditional automation (SAE level 2 and 3) on subsequent manual driving behaviour and situation awareness (e.g. Brandenburg & Skottke, 2014; Merat, Jamson, Lai, Daly, & Carsten, 2014). Consequently, take-over situations are “one of the most crucial aspects of vehicle automation in the future” (Gold, Damböck, Bengler, & Lorenz, 2013).

To avoid such negative effects of automation resulting from the idea to substitute the human driver by automation, an alternative might be to design the automation as an effective team player of the human driver (e.g. Christoffersen & Woods, 2002; Hoc, 2000; Walch et al., 2017). Within this framework human and automation support and complement each other like team players. The dynamic allocation and shift of subtasks of the driving task between the human driver and the automation is necessary. This partial task-sharing process should be based on their current capabilities to guarantee an optimal performance of the driver-automation team.

Hundreds of subtasks are controlled while driving, which requires multitasking skills (Walker, Stanton, & Young, 2001). In general, the manual driving task can be structured by three levels of skills and control (Michon, 1985). Drivers have to keep manual lateral and longitudinal control (control level). They have to execute manoeuvres like overtaking (manoeuvring level), generate and pursue plans like the navigation destination (strategical level, Michon, 1985). Successful implementations of the integration of these tasks with different levels of automations already exist (see overview of Banks, Stanton, & Harvey, 2014). For instance, navigation systems support drivers by generating and selecting the correct way. This leads to reduced workload and increased attention to safety critical behaviour on a strategical level (Srinivasan & Jovanis, 1997).

Endsley and Kaber (1999) propose several possibilities of adapting automation by task-sharing. Based on the ten levels of automation with regard to decision making and action execution ranging from manual control (Level 1) to full automation (Level 10) from Sheridan and Verplank (1978), they developed a hierarchy of levels of automation applicable to dynamic cognitive and psychomotor control task performance.

So far, only some investigations on adaptive automation in highly automated driving exist. Banks, Stanton and Harvey (2014) proposed different propositions of subtask responsibilities for a pedestrian warning and detection system based on the taxonomy of Endsley and Kaber (1999) but missed to evaluate their concept within a user study.

In order to prevent a full take-over, Gold and colleagues (2013) introduced a monitoring request. The automated car was still able to execute the whole driving task but the driver was prompted to attentively monitor this process and intervene if necessary. In general, the monitoring request was rated positively with regard to comfort and usefulness, but monitoring reduced the probability to intervene in case of critical situations. In contrast, interventions in manual driving appeared already before the situation became critical. Consequently, this implementation of partial take-over seemed not successful in high-risk situations (Gold et al., 2013). However, take-overs in critical situations were observed to elicit better performance, if the
system offered partial deceleration support compared to a complete system fail (Strand, Nilsson, Karlsson, & Nilsson, 2014).

Due to technical limitations of the automation, the human driver needs to resume responsibility from time to time. In terms of traffic safety, it would be an advantage if lateral and longitudinal control was kept by the automation (Golias, Yannis, & Antoniou, 2002) and decision making was controlled by the human. As this reduces the task complexity for the driver, it should avoid a gaze focus on the road centre as this is associated with higher driver workload (Victor, Harbluk, & Engström, 2005) and lead to a broader scanning of the traffic environment. This fosters enhanced situation comprehension and as a result improved decision performance (Baumann & Krems, 2007, 2009) in an uncritical driving situation.

Therefore, in this experiment we tested a basic task sharing approach where the control of the operative level was still assigned to the automation in a non-safety critical situation. The automation was not able to make decisions on the manoeuvre level but initiated safe mode manoeuvres in case the human did not make a decision. Lateral and longitudinal control were either completely performed by the automation which we named as AUTOpilot (level 2) or only longitudinal control was controlled by the automation representing a traditional adaptive cruise control (ACC, level 1). Prior findings showed that ACC compared to AUTOpilot enabled faster and more appropriate reactions in critical situations with automation failure (Strand et al., 2014). The authors explained the results by reduced situation awareness and the more passive role of the driver. To increase engagement of the driver, an explicit clarification of strength and limitations of system behaviour was shown to improve monitoring performance (Dogan et al., 2017). Furthermore, the ability to regain control was ameliorated if system boundaries or take-over could be anticipated (Dogan et al., 2017; Merat et al., 2014). Providing cues and appropriate knowledge about the automated driving behaviour contributes to an enhanced shared situation representation and mutual predictability as a basis of cooperative driving.

Consequently, we deduced the following hypotheses:

I. Increased situation awareness with ACC compared to AUTOpilot is assumed, leading to faster recognition and activation of an oncoming partial take-over.

II. Additional cues enabling the anticipation of upcoming events result in faster recognition and activation of an oncoming partial take-over.

These assumptions were tested in a driving simulation using a traffic light scenario for manoeuvre control. It served as a foundation for investigating human capabilities of handling cooperative situations.

Method

Participants

Twenty-four (33% female) highly-educated drivers (83% got a general qualification for university entrance or even a university degree) participated in the driving simulator study. On average they were 27 years old ($SD = 6.64$) and held a valid driving license for about 9.79 years ($SD = 7.71$). The sample covered a wide range of driving experience including 37.5% daily drivers, 41.6% regular drivers (driving
once to four times a week) and 20.9% occasional drivers (driving three days or less per month). 71% of participants reported to have experience with driver assistance systems like (adaptive) cruise control. The participants received course credits or money for their participation.

Apparatus and material

The study was conducted in a fixed-base driving simulator consisting of a cockpit mock-up in front of three large projection walls providing a view of 190° (see Figure 1). The simulation software SILAB (5.1, WIVW GmbH) was used. Gaze data were gathered by three built-in eye-tracking cameras (sampling rate 60Hz, Smart Eye AB). Manual input and feedback was provided by the touch display in the centre console (16 x 21 cm, 1920 × 1080 pixels).

Furthermore, the mental workload was measured after driving with ACC and AUTOpilot respectively using the NASA-Task Load Index (NASA-TLX, Hart & Staveland, 1988). Additionally, two scales of Driving Activity Load Index (DALI) (Pauzie, 2008) which focused on visual and acoustic demand were assessed.

Figure 1. The driving simulation lab with the fixed-base mock-up and 190° view of rural driving environment.

Design and procedure

The effect of automation level (ACC and AUTOpilot), availability of cues (warning) and scope of action (traffic light status) on the human capability of cooperative manoeuvre strategies was evaluated using a within-subjects design to guarantee a high comparability and reduced intra-individual influences. The automation condition was presented as blocked-design and participants were randomly assigned to one of the two conditions. The order of automation conditions was counterbalanced. In each condition participants had to pass a rural, curvy road with a target speed of 80 km/h without oncoming traffic. They encountered four traffic lights in ACC and in AUTOpilot condition. The intervals between traffic lights varied between one and six minutes to reduce predictability. Beforehand, they were informed about the system limitation at traffic lights: due to low sensor reliability in bright lightening conditions and legal regulations, the car would always stop (safe mode) in front of the traffic light independently of the traffic light status. Therefore, participants had to support the automation by deciding on the oncoming action of the car using the green GO or the red STOP button on the touchscreen. To guarantee
the observation and understanding of the situation we varied the scope of action by showing either red or green traffic light in an unpredictable sequence in all trials (half of trials had red light).

By passing a right hand bend, the traffic light became visible 200m before reaching it. Half of the traffic lights included a warning cue at that point in time indicating an oncoming traffic light. This was combined with either a red or green traffic light status (see Figure 2). Participants were encouraged to indicate the upcoming action as soon as they observed the traffic light status in order to guarantee a smooth and comfortable ride.

The study lasted about 50 min. Starting with a demographic questionnaire and a manual driving condition in order to get used to the driving simulator, participants experienced a practice trial (3 min) followed by the experimental condition (each 15 min) for ACC and AUTOpilot. Latencies of activation, time to first gaze to traffic light and subjective workload ratings were gathered as dependent variables. Participants were instructed to leave the right hand on the rack in front of the centre console in both automation conditions to enable the comparison of the input latencies.

**Results**

*Cooperative action – activation of partial take-over*

First of all, every driver could handle the cooperative partial take-over by executing the correct action in all situations. The latency of activation was determined as the time between visibility of traffic light (equivalent to warning) and activation of STOP or GO, which was considered as cooperative action. Figure 3A shows averaged latencies with regard to automation level, warning availability and traffic light status. A repeated measures ANOVA was calculated for inferential analyses. Participants always reacted faster when they were driven by the AUTOpilot ($M = 4241$ ms) compared to ACC ($M = 5016$ ms), $F(1,23) = 14.25, p < .001$. As expected, warning cues reduced the reaction time (reduction of $M = 1147$ ms, $F(1,23) = 46.48, p < .001$). Moreover, traffic light status influenced the latencies: Participants were faster when they saw a green ($M = 4308$ ms) compared to a red traffic light ($M = 4989$ ms), $F(1,23) = 18.90, p < .001$. No interactions were observed.
Gaze behaviour – recognition of partial take-over

Figure 3. A - averaged latencies with regard to automation level (ACC and AUTOpilot), traffic light status (red or green) and the availability of warning cue. B – averaged latencies of first gaze towards the traffic light with regard to automation level (ACC and AUTOpilot), traffic light status (red and green) and the availability of warning cue. C – effect of automation level (ACC and AUTOpilot) on averaged results of subjective workload scales ranging from 1 – “very low” to 20 – “very high”.

Gaze data could only be analysed for twelve participants due to technical inconveniences. Accordingly, non-parametric Wilcoxon signed-rank tests were calculated and Bonferroni corrections were applied. Figure 3B provides an overview of the latencies to first fixation to the traffic light. Warning cues decreased the duration of recognizing the traffic light (reduction overall of \(M = 703\) ms), \(Z = 2.51, p < .05\). This difference was mainly achieved due to the difference of warning in the ACC condition (\(Z = 2.75, p < .05\)) but not in the AUTOpilot condition (\(Z = 1.33, p > .05\)). Furthermore, participants saw the traffic light earlier when there was no warning in the AUTOpilot compared to ACC condition (\(Z = 2.51, p < .05\)).
means that a warning accelerated the recognition of the traffic light only for ACC but not for AUTOpilot. The status of the traffic light had no influence on the time to first fixation.

**Subjective workload**

The different subscales of NASA TLX and DALI were considered for every automation condition. Overall, increased workload while driving with ACC ($M = 7.08$) compared to AUTOpilot ($M = 5.18$) was observed ($t(23) = 5.12, p < .001$). However, the subjective workload was rather small (see Figure 3C). The largest difference of automation level on workload was observed for physical demand, but also mental, visual demand and effort were larger for ACC compared to AUTOpilot ($p < .001$).

**Discussion**

The present research investigated the effect of automation level, cues and scope of action on the human capability of partial take-overs in a simple strategic manoeuvre. Individuals were able to manage this cooperative action with driver assistance and partial automation. The task was predominantly visually and mentally demanding, because drivers had to recognize the traffic light and decide for the correct action. Automated system behaviour was activated by pressing the STOP or GO button. Additional cues were observed to increase the recognition and activation of partial take-overs (in line with hypotheses II). Contrary to hypotheses I, results showed longer latencies of activation in the ACC compared to the AUTOpilot condition which can be explained by the additional task of steering and lane keeping while driving with ACC. Gaze data support this assumption: If participants were warned about an oncoming traffic light they were as fast in ACC as in AUTOpilot condition in recognizing the target. But they still showed slower reactions due to manual steering activity. If there were no warning cues drivers kept their attention on the street and recognized the target later in ACC condition.

These results showed that higher level of automation does not necessarily decrease driving performance in case of a partial take-over as it was observed in full take-over situations (e.g. Strand et al., 2014). However, the participants of this study were not confronted with a critical situation and the intervention was not on a control but rather manoeuvring level.

The scope of action was varied by the status of the traffic light to ensure that participants observed the target. Surprisingly, participants reacted faster when the traffic light was green, but there were no differences in gaze behaviour. It allows for drawing the conclusion that participants were aware that the car would always stop at the traffic light. Consequently, there was an increased requirement to react when the light was green in order to prevent braking and ensure smooth continuation. The time and space for reaction was smaller for green compared to red traffic light. Therefore, it is of prime importance to provide the relevant information about the operating principles of the automated car and consequences of the own action to ensure appropriate behaviour within the scope of driver-vehicle interaction.
Overall it can be stated, that people are able to manage partial take-overs or can deal with external triggered adoption of level 1 and 2 automation in a simple scenario on the decision stage. Even with a rather small sample size, the effect of warning cue and scope of action could be observed on a behavioural level. Restrictively it must be considered that the traffic light scenario is a rather simple task with regard to information acquisition and analysis. Complexity by increased traffic density was shown to have a negative effect on driving performance in take-over situations (Gold, Körber, Lechner, & Bengler, 2016). The evaluation of more complex and varying scenarios with increased difficulty is necessary to further evaluate the concept of cooperative partial take-overs.

Furthermore, participants were not distracted while driving. Thus, they observed the automation and the external environment attentively all the time. It can therefore be assumed, that the drivers were in the loop while being temporarily chauffeured, as it is supposed to be. Further research should focus on longer driving durations or interleaved non-target-tasks (i.e. navigation task) to gain insights in partial take-over performance with reduced situational awareness. In relation to that, insights in level 3 of automated driving should be provided when drivers are allowed to turn attention to other tasks. Within that case it is of utmost importance to develop appropriate partial take-over situations in line with the cooperative driving framework. The situation specific dynamic adaptive automation might be the foundation of capable, safe and comfortable handling of future mobility. Further research should in addition focus on acceptance and trust of cooperative adaptive automation, which is pivotal for the dissemination of automated vehicles.

References


SAE. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: J3016_201609*.


Abstract

Current haptic control systems provide feedback torques based on a lateral deviation with respect to a reference trajectory (i.e., centre of the lane), which do not capture the satisficing behaviour human beings typically adopt during a lane keeping task. As such, a novel time-to-lane-crossing-based controller is proposed, which is expected to provide more human-like guidance. The aim of this study is to describe a novel time-to-lane-crossing-based controller and investigate its potential as an alternative to previous reference-trajectory-based guidance. In a simulator study twenty-four participants drove three trials through a single-lane, 10.8 km long road (width: 3 m), receiving three types of guidance, namely 1) none (manual), 2) reference based controller, 3) TLC-based controller. Results showed that both the reference-based, as well as the TLC-based guidance provided significant safety benefits, in terms of more centred and less varying lane position, and higher safety margins. Moreover, no significant differences were revealed between the two guidance approaches. In conclusion, the TLC-based guidance is a potential alternative to reference trajectory-based guidance. Nevertheless, a more detailed analysis is warranted to investigate the two approaches in different driving conditions, like road width, straights, and curves.

Introduction

Haptic shared control has been proposed as a viable alternative to complete automation for multiple applications, such as surgery (Li and Okamura, 2003; Nudehi et al., 2005), teleoperation (Sheik-Nainar et al., 2005) and vehicle operation (Griffiths & Gillespie, 2005; Forsyth & Maclean, 2006; Mars et al., 2014). It has been argued that haptic shared control provides performance benefits as well as keeps the operator engaged in the perception-action cycle (Abbink et al., 2018), consequently mitigating many well-known automation issues, like vigilance loss.

For driving, a haptic shared control is usually provided through an actuated steering wheel to assist lateral control (Mulder et al., 2012) or an actuated gas pedal for longitudinal control (Mulder et al., 2011). In lane keeping and curve negotiation tasks, studies have shown haptic shared control yielded benefits in terms of decreased lateral deviation with respect to the lane centre line (Forsyth & Maclean,
2006; Mohellebi et al., 2009; Saleh et al., 2013; Mulder, et al., 2008) a reduction in workload (Mars et al., 2014; Van Der Horst, 2004) and safety margins (i.e., time-to-lane-crossing; Melman et al., 2017) compared to manual driving. However, haptic shared control has also been reported to increase physical effort (i.e., steering torques).

Increased driver torques have been hypothesized to be the result of conflicts between the driver and the haptic guidance system (Mars, et al., 2014). Conflicts arise when a mismatch occurs between the driver and automation intention. For example, when the driver wants to cut into a curve, but the automation is programmed to drive in the centre of the lane. Indeed, it has been reported in a simulator study that participants had to ‘fight’ the controller (Abbink et al, 2011), which can worsen the overall performance (Griffiths & Gillespie, 2005; Mars, et al., 2014) and decrease acceptance (Petermeijer, et al., 2015). To reduce conflicts, the intention of the automation should match that of the driver closely.

Most of the current haptic steering guidance systems for lane keeping minimize either the lateral deviation and/or the vehicle heading deviation with respect to a reference trajectory (e.g., the centre of the lane, Griffiths & Gillespie, 2005; Mohellebi et al., 2009; Saleh et al., 2013; Abbink & Mulder, 2000). The resulting guidance system operates as an optimizing controller. However, it has been argued that drivers adopt satisficing instead of optimizing behaviour, i.e. is they aim to stay within certain safety thresholds (Goodrich, 2000). Accordingly, Boer (2016) argued that haptic shared control should provide guidance based on safety margins, more specifically based on time-to-lane-crossing (TLC, Van Winsum and Godthelp, 1996).

TLC serves as a measure of situational criticality (Saleh et al., 2013; Van Winsum et al., 2000). It is a metric for lateral control of the vehicle, analogous to time-to-collision in longitudinal control, which has successfully been used to provide guidance forces on a haptic gas pedal (Mulder et al, 2011) and in obstacle avoidance systems (Della Penna et al., 2010). Yet, to the best of our knowledge, no previous efforts have been made to assist drivers in a lane keeping task by a time-to-lane-crossing-based haptic steering system.

TLC is dependent on road curvature, lane width, velocity, vehicle orientation and steering wheel angle, which causes a TLC-based controller to adapt inherently to changing driving conditions, like road width or velocity. Such behavioural changes are similar to those of human drivers as driving conditions change (Van Winsum and Godthelp, 1996).

Although the conceived benefits of TLC-based steering guidance are promising, its use in a control structure, rather than as an evaluation metric, is subject to limitations. TLC is a non-linear parameter, with discontinuities when the lane boundary to be crossed switches (Van Winsum et al., 2000). As the vehicle approaches the lane boundary, small steering corrections can inflate (or deflate) the measured criticality. Moreover, TLC by itself does not have a direction; whilst it offers quantification of criticality, it does not inherently provide a direction towards safety.
In this paper, a control structure is proposed to alleviate the aforementioned limitations of TLC, by means of incorporating human-like uncertainty around the current trajectory, generating a field of safe travel (Boer, 2016). A trigonometric approach for TLC computation is elaborated (Van Winsum et al., 2000), and a driving-simulator study is conducted to evaluate the benefits and limitations of the developed criticality-based haptic steering guidance, in comparison to a previously developed reference trajectory-based guidance system (Mulder et al., 2012).

Figure 1. TLC calculation for straight (top) and curved road sections (bottom). Vehicle trajectories are indicated by DTLC and are assumed to have a constant velocity and steering wheel angle. (1) Straight driving on a straight road; (2) Straight road with steering input; (3) Straight driving on a curved road; (4) Driving on a curved road with steering input. Symbols: $\alpha$: heading deviation, $y$: lane margin, $R_v$: vehicle curve radius, $R_i$: inner road curve radius, $L$: lane width.

A trigonometric approach to compute TLC

Trigonometric TLC computation

In this paper we present an extension of the trigonometric approach of TLC-computation, based on derivation from Boer (2016) and Van Winsum et al. (2000), which offers an accurate, reliable and robust estimation of the TLC. The trigonometric approach requires consideration of four different scenarios (see Figure 1), namely driving a straight (steering wheel angle $= 0$) or curved vehicle trajectory (steering wheel angle $\neq 0$), on either a straight or curved road section. The vehicle drives with a velocity $v$ (m·s$^{-1}$); furthermore, steering wheel angle and velocity are
assumed constant. A kinematic bicycle model as in (Rajamani, 2006) was used to describe lateral vehicle motion.

**Straight road, straight trajectory**

The vehicle has a margin to the lane boundary \( y \) (m) from front left or front right wheel, with heading deviation \( \alpha \) (deg) between road and vehicle heading.

\[
TLC = \frac{y \cdot \sin(\alpha)}{v}
\]  

(1)

**Straight road, curved trajectory**

When steering input is not equal to zero, the vehicle follows a curved trajectory, with yaw-rate \( \dot{\psi} \) (rad·s\(^{-1}\)), which determines the vehicle curve radius \( R_v \) (m), as follows.

\[
t_{\text{circ}} = \frac{2\pi}{\dot{\psi}}
\]  

(2)

\[
d_{\text{circ}} = t_{\text{circ}} \cdot v
\]  

(3)

\[
R_v = \frac{d_{\text{circ}}}{2\pi} = \frac{v}{\dot{\psi}}
\]  

(4)

Referring to figure 1b:

\[
A = \frac{y}{\cos(\alpha)}
\]  

(5)

\[
B = R_v - A
\]  

(6)

\[
\beta = 90 + \alpha
\]  

(7)

In order to compute \( \phi \) (deg), length \( C \) (m) needs to be determined, by applying the cosine rule for side \( R_v \) (m), as follows.

\[
C = \frac{2B \cdot \cos(\beta) \pm \sqrt{(2B \cdot \cos(\beta))^2 - 4(B^2 - R_v^2)}}{2}
\]  

(8)
Applying the cosine rule to solve for $\varphi$ (deg) and calculate the corresponding arc length $D_{TLC}$ (m) to determine TLC (s).

$$\varphi = \arccos\left(\frac{B^2 + R^2 - C^2}{2B \cdot R_v}\right)$$  \hspace{1cm} (9)

$$TLC = \frac{D_{TLC}}{v}$$  \hspace{1cm} (10)

**Curved road, straight trajectory**

Driving on a curved road with a straight vehicle trajectory is depicted in Figure 1c. The law of cosines is applied to calculate $D_{TLC}$ (m), similar to Equation 8. Included in the calculation are heading deviation $\alpha$ (deg), lateral lane margin $y$ (m), lane width $L$ (m) and inner curve radius $R_r$ (m).

$$A = R_r + y$$  \hspace{1cm} (11)

$$B = R_v + L$$  \hspace{1cm} (12)

$$D_{TLC} = \frac{2A \cos(\beta) \pm \sqrt{(2A \cos(\beta))^2 - 4(A^2 - B^2)}}{2}$$  \hspace{1cm} (13)

**Curved road, curved trajectory**

The fourth TLC-calculation is a curved trajectory on a curved road, as visualized in Figure 1d. It requires the calculation of $\varphi_1$ (deg), through means of computing $\varphi_2$ (deg) and combined angle $\varphi_{12}$ (deg).

$$\varphi_{12} = \arccos\left(\frac{D^2 + R^2_v - (R_r + y)^2}{2R_v \cdot D}\right)$$  \hspace{1cm} (14)

$$\varphi_2 = \arccos\left(\frac{D^2 + R^2_v - R^2_r}{2R_v \cdot D}\right)$$  \hspace{1cm} (15)

With $D$ (m) the distance between the center of $R_r$ (m) and $R_v$ (m).

$$\varphi_1 = \varphi_{12} - \varphi_2$$  \hspace{1cm} (16)

$$D_{TLC} = \varphi_1 \cdot R_v$$  \hspace{1cm} (17)

The considerations of both previous sections are relevant here. Calculation is altered when vehicle trajectory will cross the outer lane boundary: lane width $L$ (m) is added to road curve radius $R_r$ (m). If vehicle heading is also oriented towards the outer lane boundary, the following equations are used for $\varphi_{12}$ (deg) and $\varphi_2$ (deg).
Guidance approaches

Reference-based guidance

Earlier, Mulder et al., (2012) developed a controller, that used a predicted lateral deviation with respect to a reference trajectory (i.e., the center of the lane). In this paper we will refer to this approach as the reference-based guidance.

The referenced-based guidance (REF) controls for two parameters, predicted lateral deviation $e_{\text{future, lat}}$ and predicted heading deviation $e_{\text{future, heading}}$ at lookahead time $t_{\text{lha}} = 0.7s$, assuming a constant vehicle speed and steering wheel angle. Guidance torques $T_{\text{guidance}}$ (Nm) were calculated using a three gains, namely P, D, and $K_{pbg}$, see equation 20.

$$T_{\text{guidance}} = (e_{\text{future, lat}} \cdot P + e_{\text{future, heading}} \cdot D) \cdot K_{pbg}$$

Here, $e_{\text{future, lat}}$ is defined as positive leftwards of lane centerline, $e_{\text{future, heading}}$ as positive leftwards of zero heading deviation and $T_{\text{guidance}}$ as positive in rightwards steering corrections (clockwise). Feedback gains were set to $P = 0.9$, $D = 0.08$ and $K_{pbg} = 2$.

TLC-based guidance

Time-to-line crossing approaches 0 s at increasingly risky driving situations. Equation 21 is used to generate a usable deviation signal $\Delta e$.

$$\Delta e = \frac{TLC \cdot \gamma + \theta}{TLC \cdot \frac{\Phi}{\Theta} + 1}$$

Such that,

$$\lim_{TLC \to \infty} \Delta e = \phi, \quad \lim_{TLC \to 0} \Delta e = \theta$$

As such, $\phi$ and $\theta$ determine lower and upper bounds of criticality, respectively. Finally, is related to relative weighing between these two bounds. The presence of noise (e.g., motor, sensory, or external noise) influences driving behaviour (Kolekar et al., 2016) to illustrate, drivers usually stay away from the edge of the road, regardless of their accuracy in following the road heading. To account for this noise, the impact from potential steering disturbances on safety margins is taken into
account. Similar to Boer (2016), the current vehicle curve radius $R_v$ is disturbed with a factor $\lambda$ (m$^{-1}$).

\[
R_{\lambda} = \frac{1}{R_v^{1 \pm \lambda}}
\]  

(23)

As shown in Figure 2, driving straight (effectively with $R_v = \infty$) yields two uncertainty trajectories with $R = \pm \lambda^{-1}$ (m). Conversely, on curved trajectories $\lambda$ is linearly related vehicle curvature (Boer, 2016). For both uncertainty trajectories, with vehicle curve radius $R_{-\lambda}$ and $R_{+\lambda}$ corresponding TLCs are computed. Combining equations 21 and 23 yields equation 24, which is the control algorithm to determine guidance torques $T_{\text{guidance}}$ (Nm), with $\phi = 0.01$ for lower limit control activity, $\theta = 10$ for upper limit control activity, $\gamma = 0.1$ for the relative weighting, $\lambda = 0.004$ for the driver uncertainty, and $K_{\text{cbg}} = 0.3$ as the deviation-to-torque gain.

\[
T_{\text{guidance}} = K_{\text{cbg}} \left( \frac{TLC_{-\lambda} \cdot \gamma + \theta}{TLC_{-\lambda} \cdot \frac{\gamma}{\phi} + 1} \right) \frac{TLC_{+\lambda} \cdot \gamma + \theta}{TLC_{+\lambda} \cdot \frac{\gamma}{\phi} + 1}
\]  

(24)

Using this algorithm, driving conditions with equal TLC$_{-\lambda}$ and TLC$_{+\lambda}$ will provide zero control input. Figure 3 clearly shows how the TLC-based guidance adapts the feedback torques in relation to road width, compared to the reference-based guidance. Reference-based guidance increases the maximum guidance torque, whereas the TLC-based guidance widens the range of guidance torque; only close to the lane boundary the feedback torques rapidly increase.
Figure 3. Magnitude of guidance torques, plotted as function of lateral deviation, on a straight road, at velocity $v = 130$ km/h, heading deviation $\theta = 0$, and yaw rate $\dot{\gamma} = 0$. Performance based guidance (PBG, blue) and criticality-based guidance (red, CBG) are determined for road width $= 3$ m (left) and road width $= 5$ m (right).

**Experimental method**

**Participants**

Twenty-four participants, recruited from the TU Delft student population, took part in the experiment (mean age 24.1, SD 1.9, 16 male). All participants had normal or corrected to normal eyesight, and were in possession of a valid driver’s licence for at least one year.

**Apparatus**

This study was conducted on a fixed-base simulator; the setup has previously been used (Mulder et al., 2012; Melman et al., 2017; Petermeijer et al., 2015). A dedicated computer controlled a Moog-FCS ECol8000S motor, to provide actuation on the steering wheel at 2500 Hz. The visual environment was updated at 60 Hz; three projectors were used to provide 180 horizontal and 40 vertical field of view.

A single-track model was used for vehicle dynamics to mimic the driving dynamics of a Nissan luxury car (i.e., heavy sedan). An automated gearbox was used, and velocity was fixed at 130 km/h. A light centering stiffness, as a function of the steering wheel angle, was applied in all conditions to emulate wheel-ground interaction forces. Car kinematics, guidance torques, as well as the driver input on the steering wheel were recorded at 100 Hz.

**Road environment**

The participants drove the vehicle over a 10.8 km long, single lane road without other traffic, for approximately 5 minutes. The road was composed of straights (length 220 metres), left and right single curves (length 218 metres, inner curve radius 500 metres) and winding sections (four alternations, inner curve radii 500 metres). Curves were interspersed with straight sections (length 150 metres) to prevent crossover effects. Moreover, a long straight section was included to
investigate steadystate behaviour. At the start the vehicle would accelerate to a fixed speed of 130 km/h, until the end of the trajectory, where it would decelerate to zero.

**Experimental design and instructions**

Three guidance conditions, namely 1) manual, 2) reference-based guidance (REF), and 3) TLC-based guidance (TLC) were each driven in three trials, namely a training trial, a trial on a normal road (width: 3 m) and a trial on a wide road (width: 5 m). Road width and guidance conditions were counterbalanced over all participants.

All participants read the experiment instructions, signed the informed consent form. Participants were verbally reminded that they are free to pause or stop the experiment at any time (if nausea arises), and to drive as they normally would, and were informed that no other road users would be encountered during the trials. No questions regarding specific controller functionality were answered during the experiment.

After each trial, NASA Task Load Index (NASA-TLX) forms were filled out to assess subjective workload (Hart and Staveland, 1988). Subsequently, participants were inquired for their nausea with a six item question, ranging from not experiencing any nausea (1) to vomiting (6). The experiment would be stopped if any participant responded with a nausea level of 4 or higher; which did not occur throughout the experiment. After a trial in which steering guidance was presented, participants were interrogated about their acceptance of the assistance system, by means of a five point scale containing nine items (five related to usefulness, four to satisfaction) (Van Der Laan et al., 1997). After each guidance condition, consisting of three experimental runs of approximately 20 minutes, a five minute break was taken.

**Dependent Measures**

The following dependent measures were analysed for the trials using the normal road width (3 m). Analysis was done on the data recorded between 10 s and 2.5 min of the track.

- Mean absolute lane position (m): a measure of choosing lane position.
- Standard deviation of the lane position (m): a measure that describes the driver’s variability in lane keeping performance (i.e., swerving behaviour).
- Median time-to-line crossing (s): a measure of the safety margin throughout the driving task.
- Standard deviation of the steering wheel angle: a measure of the driver’s variability of the steering wheel input, which reflects the lane control activity of the driver.
Statistical analysis

The independent measures were transformed to ranks in order to deal with any non-normal distributions, according to Conover and Iman (1981), before they were subjected to a one-way analysis of variance (ANOVA). A post-hoc analysis was conducted by performing a pairwise comparison using a Tukey honest significance criterion. The significance level was set to 0.05.

Results

Figure 4. Distribution of the lateral position (left) and time-to-lane crossing (right) per condition. Area underneath the distribution equals one.

Table 1 shows the mean and standard deviations of the independent measures across participants. ANOVAs showed an effect for mean absolute lateral position (F(2,69) = 14.24, p < .001) and standard deviation of lateral position (F(2,69) = 15.11, p < .001).

Table 1. Means and standard deviation of vehicle state measures - Straights

<table>
<thead>
<tr>
<th>Variable</th>
<th>MAN (1) M(SD)</th>
<th>REF (2) M(SD)</th>
<th>TLC (3) M(SD)</th>
<th>ANOVA</th>
<th>Pairwise comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>M abs lateral position (m)</td>
<td>0.282 (0.079)</td>
<td>0.197 (0.066)</td>
<td>0.208 (0.060)</td>
<td>F(2,69) = 14.24</td>
<td>X</td>
</tr>
<tr>
<td>SD lateral position (m)</td>
<td>0.315 (0.076)</td>
<td>0.222 (0.068)</td>
<td>0.246 (0.064)</td>
<td>F(2,69) = 15.11</td>
<td>X</td>
</tr>
<tr>
<td>Median TLC (s)</td>
<td>1.909 (0.103)</td>
<td>2.020 (0.101)</td>
<td>1.991 (0.089)</td>
<td>F(2,69) = 7.89,</td>
<td>X</td>
</tr>
<tr>
<td>SD steering wheel angle (deg)</td>
<td>15.830 (1.188)</td>
<td>15.282 (1.033)</td>
<td>15.430 (1.019)</td>
<td>F(2,69) = 3.41,</td>
<td>X</td>
</tr>
<tr>
<td>M abs guidance torque (Nm)</td>
<td>- (0.147)</td>
<td>0.684 (0.137)</td>
<td>0.685 (0.137)</td>
<td>F(1,46) = 0.04,</td>
<td>X</td>
</tr>
</tbody>
</table>

Note: The M abs guidance torques for the manual condition (MAN) are missing, because there are no guidance torques exerted on the steering wheel during manual driving.
time-to-lane crossing steering guidance

115

= 14.24, p < .001), standard deviation of the lateral position F(2,69) = 15.11, p < .001), the median of the TLC (F(2,69) = 7.89, p = .001), and the standard deviation of the steering angle (F(2,69) = 3.41, p = .039). The pairwise comparison revealed that for all metrics except the SD steering wheel angle and the mean guidance torques, the both guidance approaches yielded better performance (i.e., lower mean absolute lateral position, lower standard deviation of the lateral position, and higher median TLCs) from manual driving. For the standard deviation of steering wheel angle only the reference-based guidance was significantly lower than the manual condition. The mean absolute guidance torque did not differ between the reference-based and TLC-based approach.

Figure 4 (left) shows the distribution of the lateral position condition. It can be seen that the manual has a slightly wider distribution compared to the two guidance systems. On the other hand, the two guidance systems seem have the same narrow distribution of lateral position. Meaning, participants drove more in the centre of the lane when they used haptic support.

The right plot in Figure 4 shows the distribution of the time-to-lane crossing per condition. Similar to the left plot it can be seen that the two support systems yield slightly safer behaviour (i.e., higher time-to-lane-crossings) compared to manual driving.

Figure 5 illustrates the mean guidance torques as a function of time-to-lane-crossing. It can be seen that the TLC-based guidance provided lower mean feedback torques compared to reference-based guidance. Note, however, that the TLC-based guidance plot lies more to the left, meaning the TLC-based controller recorded lower time-to-lane-crossings than the reference-based controller.

Figure 5. Force feedback as a function of TLC for the two guidance approaches. X-axis: TLC-bins are 0.4 (s) large; Y-axis: The mean guidance torques within the TLC-bin. No TLCs < 0.8 (s) occurred.
Discussion

Drivers using TLC-based guidance yielded similar behaviour compared to the reference-based guidance, in terms of lateral position and time-to-lane crossing. Note, however, that these metrics were analysed over straight and curved sections of the track with a road width of 3 metres. It was hypothesized that drivers using the TLC-guidance on a wider road will drive more akin to manual drivers (cf. Figure 3). Moreover, it is expected that the two guidance approaches yield different behaviour in the curves, since the TLC-based guidance should allow curve cutting behaviour, whereas the reference-based guidance does not.

In line with the results, the mean guidance torques did not significantly differ between approaches. However, Figure 5 suggests that the guidance torques are quite different for situations with lower values of TLC (i.e., TLC < 1.8). It can be seen that for similar values of TLC the reference-based guidance provides higher torques than the TLC-based ones. These differences are expected to be more distinct for a wider road, since the TLC-based guidance would intrinsically adapt, whereas the reference-based guidance does not.

In the current TLC-based approach, the trigonometric TLC-calculation used four different procedures based on four separate situations (i.e., straight/curved trajectory and straight/curved road). Yet, in future research these calculations can be generalized to one situation, namely the curved trajectory and curved road. The only constraint to this procedure is that a steering angle of exactly zero (i.e., driving straight forward) may never mathematically occur. Though, one can assume a very low steering angle near zero, resulting in a near straight trajectory and similar TLC-values.

Conclusion

In this study we proposed, developed, and evaluated a novel lane keeping guidance, based on time-to-line-crossing (TLC) in order to mimic a more human-like driver style. In a simulator experiment the novel guidance was compared to manual driving and a previous reference deviation based guidance. Results showed that the two guidance approaches are equally effective in improving lane keeping performance and safety margins compared to manual driving, in terms of absolute lane position and time-to-lane crossing, respectively. Hence, the TLC-based guidance is a viable alternative to the reference-based guidance. Subsequent analysis should be performed to evaluate the approaches separately in curves and straights, and wide roads and narrow ones, since those are sections where more distinct differences are expected to emerge.

Acknowledgements

D.A. Abbink, S.M. Petermeijer, and S. Kolekar are supported by the Netherlands Organization for Scientific Research (NWO); VIDI project 14127. This paper is a dissemination of the master thesis of Sjors Oudshoorn at the Delft University of Technology. The authors would like to thank Timo Melman for his efforts in supervising Sjors Oudshoorn during this process.
time-to-lane crossing steering guidance

References


Perceived Safety: a necessary precondition for successful autonomous mobility services

Jan Grippenkoven¹, Zoë Fassina², Alexandra König¹, & Annika Dreßler¹
¹German Aerospace Center - Institute of Transportation Systems
²Technische Universität Berlin

Abstract

In order to reach the goal of reducing traffic volume in cities, new smart, sustainable and user-centred mobility alternatives to motorized individual traffic are needed. The progress in vehicle automation offers a promising potential for increasing the attractiveness of public transport by improving the flexibility and availability of service. Besides the expected benefits, autonomous mobility services will open up a range of new challenges. Their ultimate success will crucially depend on user acceptance, while perceived safety issues of autonomous driving are a recurring theme in the media. Therefore, besides the technological challenges in any effort to build up autonomous mobility services, the perceived safety of prospective users needs to be taken into account as a necessary precondition for success from the beginning. Based on the results of a workshop with prospective users of an autonomous public transport system, an online survey was developed to analyse fears with regard to using the system. The identified fears can be clustered in three groups: (1) Fears of other persons as co-travellers, (2) Fears resulting from lack of transparency of the system and (3) Fears of technical malfunctions. Countermeasures to cope with those fears were defined and rated by the respondents.

Introduction

Autonomous driving, shared transportation and the goal to create a shift from individual transport towards public transport play a major role in the discussion about relevant trends for the future of mobility (Zukunftsinstitut, 2018). Driverless public bus shuttles are a vision that combines those mobility trends and could become a central element within public transport (Fraedrich, Beiker, & Lenz, 2015; OECD, 2015). These shuttles could contribute to the reduction of traffic volume and emissions in cities by decreasing the number of rides in private cars while increasing the average number of passengers per vehicle. However, when introducing a radical change like autonomous bus shuttles to the public mobility market, it is important to take travellers’ needs, requirements and fears into account at an early stage of development (Brandies, König, Viergutz, Fraedrich, Gebhardt, Ulmer, Sippel & Dotzauer, 2017). Any innovative service should be based on actual users’ needs. This is a necessary precondition in order to attain acceptance and willingness to use

it (Rogers, 2003). Therefore, psychological research is an essential building block in system development throughout all stages of implementation.

With a focus on regular fixed schedule bus transport, various authors outlined that reliability, fare, distance of bus stop, frequency and travel time are amongst the relevant factors that affect the appraisal from a passengers’ perspective (Beirão & Cabral, 2007; Bourgeat, 2015; Jianrong, Wei & Bing, 2011; De Oña, De Oña, Eboli & Mazzulla, 2013; Hensher, 1994). A reliable service is a service that is controllable as well as predictable for the user, for example with regard to schedule and route. Therefore, the reliability of a transport service can be assumed to affect the passengers’ feeling of certainty. Passengers should be able to understand and predict the status and position of the vehicle as well as their time of arrival. As an extension of certainty, it is important that a transport service does not give the impression that it might cause harm to the users. It is therefore assumed that the perceived safety is a factor that has to be ensured for any new mobility service to be successful. Thus, in the development of new autonomous bus services, a primary goal should be to create a risk-free environment that gives users a feeling of certainty as well as safety. The use of a public autonomous transport system that operates small shuttle vehicles to be shared with persons that the user is not familiar with might cause concerns or even fears. The crucial characteristic of autonomous public transport in this context is the absence of a driver as a contact person and supervising authority who can offer protection. Since transport with autonomous bus services does not exist on a regular basis yet, perceived safety of users is a topic that to date has not been addressed sufficiently in transportation research.

Therefore, the understanding of factors that affect perceived safety is an important goal in the development of innovative transport services. As a first step towards this goal, a workshop with potential users was conducted to identify and rate major fears that exist with regard to the characteristics of driverless bus shuttles. Ideas for measures and systems that may support the coping with those fears were derived within the workshop. As a second step, the identified aversive scenarios were translated into an online questionnaire, together with the measures derived for their mitigation. The questionnaire served to achieve a quantification of the measures’ potential to reduce fears related to autonomous public bus transport. In line with future mobility visions, the target transportation concept that the workshop as well as the online questionnaire referred to, was a transport with electrified and shared driverless bus-shuttles that are booked via an app, operate on demand and are flexible with regard to route and stops.

Workshops

Methods

Participants

Two workshops were conducted independently of each other, with nine participants each. Of the 18 participants in total, 11 were women and 7 were men. They were recruited in the German capital Berlin. The average age of the sample was 39.6 years (SD = 11.9 years). 17 out of the 18 participants were regular users of public transport. Thirteen of the participants reported using public transport daily, four
stated to use it up to six times a week and one person stated to use public transport less than once a month.

**Material**
Since autonomous transport systems do not regularly exist in public transport, it was considered beneficial to introduce participants to this mobility concept before their participation. Therefore, it was recommended to participants to voluntarily try out a driverless shuttle service on one of the test fields that exist in Berlin before the dates of the workshops. On the dates of the workshops all participants confirmed that they had tried one of the existing services. The workshop itself took place in a meeting room. Thoughts and ideas were noted or sketched with markers on post-its and attached to movable workshop whiteboards.

**Design and procedure**
In the *introduction phase* of the workshop, all participants were welcomed and briefly introduced to the topic and the objectives. The experiences that participants had made before with bus shuttles were shared by the participants in the group. In addition, participants were introduced to the concept of mobility on demand and the idea of public transport without fixed routes and stops in the beginning of the workshop. It was clarified that the target system to be discussed was a transport service based on electrified and shared driverless bus-shuttles that can be booked via an app, operates on demand and is flexible with regard to the route and stops. This concept including the potential booking system via app was additionally illustrated with explanatory videos to offer a more vivid impression. It was pointed out that the shuttle picks up passengers and lets them exit at virtual smartHubs, stops that are temporarily established along the route at spots that depend on the travel requests and catchment area of the prospective passengers. It was furthermore explained, that the optimal route is determined by an algorithm and that the distance and travel time could therefore vary depending on the starting points and destinations of individual passengers as well as mobility requests of other passengers that want to access the vehicle.

After the detailed introduction, a *brainstorming phase* was introduced. The first task of the participants was to individually elaborate on the question: *What would determine my personal safety experience in flexible and autonomous bus shuttles in the public transport of the future?* Participants documented their individual thoughts on post-its. After the individual work, participants were instructed to share their ideas with each other. The pursued goal was to collect as many aspects that influence the personal safety experience (PSE) as possible. After participants had introduced their thoughts and ideas to each other, they were invited to cluster them in superordinate groups and label them with appropriate headlines.

Based on the identified clusters of ideas an *idea phase* was started. The goal in this workshop phase was to identify measures that can enhance the PSE in the context of the introduced public transportation concept. Participants were asked to discuss how it might be possible to counteract each of the fears. All ideas that emerged during the workshop were documented on post-its and collected on movable workshop whiteboards. The entire workshop took three hours. Each participant received 30€ for the participation.
Results

The central result of the workshops was a list of fears identified with regard to a public transportation system with flexible autonomous bus shuttles and a list of measures that could be part of the system to counteract the fears. The fears that were rated most relevant by the participants of the workshops are listed in table 1. A five-point Likert scale was used for the ratings. Each fear that was on average rated with a value of three or higher was considered further in the process of the workshop.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fears related to other passengers</td>
<td>A frightening group dynamic arises (e.g. drunken hooligans)</td>
<td>group dynamics</td>
</tr>
<tr>
<td></td>
<td>Unpleasant / strangely behaving passenger is entering the bus</td>
<td>unpleasant passenger</td>
</tr>
<tr>
<td></td>
<td>Being subjected to physical violence or sexual assault</td>
<td>physical violence</td>
</tr>
<tr>
<td></td>
<td>Criminal assault / theft</td>
<td>theft</td>
</tr>
<tr>
<td></td>
<td>No help of passengers in emergency situations</td>
<td>no help</td>
</tr>
<tr>
<td></td>
<td>Overcrowded bus</td>
<td>overcrowded</td>
</tr>
<tr>
<td></td>
<td>Signs of vandalism and destruction in the bus</td>
<td>vandalism</td>
</tr>
<tr>
<td>Uncertainty - Fears related to a not transparent system</td>
<td>Unexpected stop of bus - not knowing what to do</td>
<td>unexpected stop</td>
</tr>
<tr>
<td></td>
<td>Unpunctual arrival, planning uncertainty</td>
<td>unpunctual</td>
</tr>
<tr>
<td></td>
<td>Bus takes a route through a bad area</td>
<td>bad area</td>
</tr>
<tr>
<td></td>
<td>Bus takes a route that is unfamiliar - not knowing where you are</td>
<td>unfamiliar route</td>
</tr>
<tr>
<td></td>
<td>Not knowing whether the sensors recognized an obstacle in the direction of driving</td>
<td>sensor uncertainty</td>
</tr>
<tr>
<td>System safety – Fears related to technical malfunctions and the design of the system</td>
<td>Fear of a puncture</td>
<td>puncture</td>
</tr>
<tr>
<td></td>
<td>Fear of inability to intervene in the driving course</td>
<td>no intervention</td>
</tr>
<tr>
<td></td>
<td>Fear of insufficient maintenance of the vehicle</td>
<td>no maintenance</td>
</tr>
<tr>
<td></td>
<td>No emergency exit / Fear of inability to exit when it is very crowded and there is only a single door</td>
<td>blocked exit</td>
</tr>
<tr>
<td></td>
<td>Fear with regard to protection of personal data (booking via app)</td>
<td>personal data</td>
</tr>
</tbody>
</table>

The identified fears could be clustered in three major groups: fears related to other passengers, fears related to a potential lack of transparency of the service (leading to a state of uncertainty) and fears related to technical malfunctions and the design of the system. Fears related to other passengers contained for example the fear of being in a bus with a group of aggressively behaving persons, the fear of becoming the victim of a criminal assault or the fear of being alone in an autonomous bus.
perceived safety: a necessary precondition

shuttle with no other passenger or driver that could help in case of an emergency. Identified fears labelled as uncertainty, refer to situations in which an individual is not able to understand or predict the status, position or time of arrival of the system. Planning uncertainties due to the flexible nature of the service play a role in this group as well as uncertainties regarding the detection performance of sensors in case of potential obstacles. Fears related to technical malfunctions or system design contain for example the fear of not being able to exit the vehicle through the only door provided when it is crowded or the fear that the vehicle could have a puncture and no supervisor is present to monitor the status.

Based on the list of fears, a variety of measures were identified that could counteract the fears and make users feel safer in the envisaged bus shuttle system. Some exemplary ideas that were sketched based on the results of the workshop are illustrated in figure 1.

Figure 6. Illustrations of exemplary measures to cope with fears in flexible autonomous bus shuttles, identified in the workshop. (a) mobile security and puncture service; (b) digital information panel inside the bus; (c) panic button with voice communication to safety service or police; (d) Video surveillance and assistance by traffic management central; (all illustrations designed by Zoë Fassina)

These illustrations are examples of a variety of ideas that were identified by the participants during the workshop. Out of the variety of ideas, thirteen ideas were rated as relevant by participants. A five-point Likert scale was used for the ratings.
All ideas that received an average relevance rating of three or higher served as an input for the questionnaire study that is described in the next section. The following list introduces these thirteen ideas. It is not meant as a recommendation, but to give an overview of the results collected in the workshop:

**Seat monitoring system:** Passengers can use an app of the transport company to display free seats, track them and plan their trips based on available seats.

**Individual cabin:** In addition to the ordinary seating and standing places in the bus shuttle, a special room could be offered that is separated by transparent walls and can be locked from the inside.

**Locking system:** The door of the bus shuttle can be locked from the inside if it passes through areas that passengers consider unsafe and hinder persons from the outside to enter the bus.

**Emergency brake:** An emergency brake can be operated by passengers inside the bus shuttle. It can be used to elicit a full brake application.

**Mobile security service:** The cabin of each bus shuttle is monitored by a video surveillance centre. If critical situations arise or passengers make an emergency call from the shuttle, a mobile security service company is contacted. The security service is on duty during the entire service time and able to intervene quickly on site in case of technical issues or passenger conflicts.

**Video assistance system:** The cabin of each bus shuttle is monitored by a video assistance centre. The video assistant can be contacted by the passengers via a button in the bus. Video communication with sound and image is provided between passengers and the assistance centre. The assistant can help passengers in various situations.

**Mobile application:** Via a mobile application of the transport company that can be installed on smartphones, passengers can report for example defective parts, dirt or signs of vandalism in the bus shuttles. In case of emergencies, an emergency call can be made with the application.

**Navigation interface:** Monitors in the bus show a map overview and indicate the current location of the bus shuttle as well as the position of other public transport vehicles in the vicinity. Information about the route, upcoming arrivals and departures are provided.

**Passenger training:** A training is offered for passengers who (want to) use the flexible bus shuttle and would like to know in detail how it works. The trainings are offered regularly by the transport company.

**Information system:** Each bus contains an information system based on a digital information panel. The system provides information about the bus, a general user manual, a question-and-answer menu and explanatory videos.
Head-up display: An augmented display is projected on the windshield of the vehicle. It displays velocity information and announces the forthcoming direction of vehicle movement. In addition, it offers information about detected road users, obstacles and traffic signs.

Panic button: When the panic button is operated by a passenger, a speech connection to an emergency call centre is established. In contrast to the video assistance system, no video communication is provided between passengers and the assistance centre.

Telephone: A telephone inside the bus shuttles enables contact making within the German telephone network for persons who do not own a mobile phone. In case of emergency or changes to the route and/or the ride duration, it is possible to contact other persons.

Online questionnaire

While fears related to flexible autonomous bus shuttles and measures to cope with those fears were identified during the workshop, the online questionnaire was designed to quantify the PSE and the assumed effectiveness of the measures described in the previous section to increase the safety experience of passengers in flexible autonomous bus shuttles.

Methods

Participants
Three questionnaires (version A, B and C, see table 2) were designed to combine all aversive situations identified in the workshop with all of the measures envisioned to cope with those fears. Participants were randomly assigned to one of the three versions of the questionnaire. Out of 221 respondents who started to fill in one of the questionnaires, 106 completed it. Within the sample of these 106 respondents, 56 were female, 46 were male and four categorized their gender as diverse. The average age of the respondents in the sample was 41.3 years (SD = 17.7 years). In total 42 respondents completed questionnaire version A, 31 respondents completed version B and 33 respondents completed version C.

Material
The three versions of the online questionnaire (table 2) were designed with the software Limesurvey Version 2.06. In total, the influence of thirteen measures on the PSE was tested in 17 fear scenarios derived from the workshop (listed in table 1). To rate the PSE with regard to a scenario, a continuous horizontal scale with a virtual sliding controller was used. The scale ranged from 1 to 100. The minimum was labelled with 1 – low personal safety experience and the maximum with 100 – high personal safety experience. The participants’ descriptions from the workshop were used to formulate short introductions to the measures. In addition, sketches comparable to the ones depicted in figure 1 were provided to illustrate the measures.
Table 2. Distribution of the 17 aversive scenarios across three online questionnaire versions A, B and C. Each of the 13 mitigation measures was presented in combination with each scenario.

<table>
<thead>
<tr>
<th>Questionnaire A</th>
<th>Questionnaire B</th>
<th>Questionnaire C</th>
</tr>
</thead>
<tbody>
<tr>
<td>unfamiliar route</td>
<td>puncture</td>
<td>no maintenance</td>
</tr>
<tr>
<td>no help</td>
<td>vandalism</td>
<td>overcrowded</td>
</tr>
<tr>
<td>blocked exit</td>
<td>personal data</td>
<td>bad area</td>
</tr>
<tr>
<td>sensor uncertainty</td>
<td>no intervention</td>
<td>physical violence</td>
</tr>
<tr>
<td>group dynamics</td>
<td>theft</td>
<td>unexpected stop</td>
</tr>
<tr>
<td></td>
<td>unpunctual</td>
<td>unpleasant passenger</td>
</tr>
</tbody>
</table>

**Design and procedure**
To quantify the PSE, participants were first introduced to an aversive scenario based on the fears identified in the workshop (table 1) and were asked to imagine themselves in the scenario and give a rating of their PSE (e.g.: “Imagine that you are a passenger of the previously introduced flexible and autonomous bus shuttle. How would you rate your personal safety experience in this vehicle given the possibility to be subject of a criminal assault or theft?”). To rate the PSE with regard to this scenario, the value of a sliding controller had to be adjusted between 1 and 100 (1 – low personal safety experience, 100 – high personal safety experience). These PSE values were used as a personal baseline with regard to each specific scenario. To evaluate the potential effect of measures on the PSE in a given aversive scenario, the scenario was presented to the respondents in combination with each one of the thirteen measures in succession. In every combination, participants rated their PSE with regard to the scenario under the assumption that the traffic company would introduce the specific measure in the bus-shuttle (e.g.: “Assuming that there was a panic button in each vehicle that you could use to immediately establish a connection to a safety service, how would you rate your personal safety experience in this vehicle given the possibility to be subject of a criminal assault or theft?”). To avoid order effects, the scenarios as well as the measures described for each scenario were presented to each participant in random order.

Since the effect of each of the thirteen measures on each of the 17 identified aversive scenarios was to be tested, the resulting number of combinations was high. In addition, the flexible autonomous bus service, as well as each of the measures and each of the scenarios were introduced in detail. In case of a complete combination of scenarios and measures for each participant, this would have made the online questionnaire overly time consuming to fill in. Therefore, the overall item set was split into three questionnaires. Each contained only a subset of the scenarios, but all countermeasures (table 2). It took between 25 and 30 minutes to fill in one version of the online questionnaire.
Results

The first analysis served to analyse the impact of the 17 aversive scenarios on the PSE ratings. In each of the three questionnaire groups (sample A, sample B and sample C), a PSE baseline was determined. The baseline indicates the PSE that participants reported with regard to the idea of using the flexible autonomous bus shuttle in general. This baseline was then compared to the PSE reported for the 17 aversive scenarios. For each of the three questionnaire samples, paired samples t-tests were conducted to compare the PSE baseline to the PSE in each of the scenarios (Bonferroni correction applied to significance level). The descriptive results are illustrated in Figure 7.

In sample A, the PSE difference of all aversive scenario to the baseline PSE value reached statistical significance, except for the scenario unfamiliar route, *t*(41) = 2.194, *p* = .034, *d* = 0.34, for which the *p*-value did not fall below the Bonferroni-adjusted significance level. As shown in table 3, the scenarios no help, *t*(41) = 6.065, *p* < .001, *d* = 0.94, blocked exit *t*(41) = 7.287, *p* < .001, *d* = 1.12, sensor uncertainty, *t*(41) = 7.533, *p* < .001, *d* = 1.16, and negative group dynamics, *t*(41) = 7.440, *p* < .001, *d* = 1.15, all led to significant differences in PSE compared to the baseline. In addition, these significant changes illustrate large effects of the scenarios on the PSE rating, as indicated by the high values of Cohen’s *d*.

In sample B, for the respondents of the second version of the questionnaire, only the difference in PSE values between the baseline and the scenario vandalism was statistically significant, *t*(30) = 4.438, *p* < .001, *d* = 0.80. Signs of vandalism in a bus shuttle appear to have a large effect on the PSE. PSE differences in other scenarios did not reach statistical significance. However, the differences in PSE between the baseline and the scenario puncture, *t*(30) = 2.276, *p* = .030, *d* = 0.41, as well as the scenario theft, *t*(30) = 2.220, *p* = .034, *d* = 0.40, were close to statistical significance with small to medium effect sizes, but classified as not significant due to the Bonferroni-adjusted alpha. The difference in PSE with regard to personal data, *t*(30) = 1.395, *p* = .173, *d* = 0.25, that are given to the traffic company, the inability to intervene, *t*(30) = 1.647, *p* = .110, *d* = 0.30 in the process of driving and the route, and a potential unpunctuality, *t*(30) = 0.436, *p* = .666, *d* = 0.08 of the bus shuttle did...
not reach statistical significance compared to the baseline PSE. Still, small effects on the PSE were found for fears related to personal data that are given to the traffic company and the fear of the inability to intervene.

**Table 3. Comparison of the PSE of the scenarios to each sample’s baseline based on paired samples t-tests.**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Scenario</th>
<th>t</th>
<th>df</th>
<th>d</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>unfamiliar route</td>
<td>2.194</td>
<td>41</td>
<td>-</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>no help</td>
<td>6.065</td>
<td>41</td>
<td>0.94</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>blocked exit</td>
<td>7.287</td>
<td>41</td>
<td>1.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>sensor uncertainty</td>
<td>7.533</td>
<td>41</td>
<td>1.16</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>negative group dynamics</td>
<td>7.440</td>
<td>41</td>
<td>1.15</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>B</td>
<td>vandalism</td>
<td>4.438</td>
<td>30</td>
<td>0.80</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>puncture</td>
<td>2.276</td>
<td>30</td>
<td>0.41</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>theft</td>
<td>2.220</td>
<td>30</td>
<td>0.40</td>
<td>.034</td>
</tr>
<tr>
<td></td>
<td>personal data</td>
<td>1.395</td>
<td>30</td>
<td>0.25</td>
<td>.173</td>
</tr>
<tr>
<td></td>
<td>intervene</td>
<td>1.647</td>
<td>30</td>
<td>0.30</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td>unpunctuality</td>
<td>0.436</td>
<td>30</td>
<td>0.08</td>
<td>.666</td>
</tr>
<tr>
<td>C</td>
<td>physical violence</td>
<td>4.301</td>
<td>32</td>
<td>0.75</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>unpleasant passenger</td>
<td>6.865</td>
<td>32</td>
<td>1.20</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td>bad area</td>
<td>3.827</td>
<td>32</td>
<td>0.67</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td>unexpected stop</td>
<td>2.247</td>
<td>32</td>
<td>0.39</td>
<td>.032</td>
</tr>
<tr>
<td></td>
<td>maintenance</td>
<td>-1.178</td>
<td>32</td>
<td>-0.03</td>
<td>.860</td>
</tr>
<tr>
<td></td>
<td>overcrowded</td>
<td>-1.923</td>
<td>32</td>
<td>-0.34</td>
<td>.063</td>
</tr>
</tbody>
</table>

For sample C, the PSE value of six scenarios was compared to the baseline. Three of the six scenarios led to a significant decrease in the PSE. Large effects on the PSE could be found for the scenarios that describe experiences of physical violence or sexual assaults and an unpleasant passenger who is entering the bus shuttle. The PSE value differences for the scenario physical violence, \( t(32) = 4.301, p < .001, d = 0.75 \) as well as for the scenario unpleasant passenger, \( t(32) = 6.865, p < .001, d = 1.20 \), reached statistical significance. Moreover, the scenario of a bus shuttle driving through a bad area led to a statistically significant decline in PSE, \( t(32) = 3.827, p = .001, d = 0.67 \), with a medium to large effect. The PSE differences for an unexpected stop, \( t(32) = 2.247, p = .032, d = 0.39 \), the scenario overcrowded, \( t(32) = -1.923, p = .063, d = -0.34 \) and the scenario that describes doubts about the maintenance state of the vehicle, \( t(32) = -1.178, p = .860, d = -0.03 \) did not reach significance.
perceived safety: a necessary precondition

statistical significance. However, the scenario of an unexpected stop revealed a small to medium effect on the PSE rating.

Next, the impact of the identified measures was analysed for each scenario. Pairwise comparisons of the PSE values were calculated between the general PSE ratings for each aversive situation and the PSE ratings under the assumption that one of the thirteen additional measures would be included in the particular situation. Therefore, thirteen paired-samples t-tests were conducted for each aversive situation. The critical p-value of .05 was corrected for thirteen pairwise comparisons per scenario, using the Bonferroni correction. An overview of the impact of the measures on the PSE in the particular scenarios is provided in table 4.

Table 4. Overview of the paired samples t-tests results. Bonferroni correction was applied for the thirteen pairwise comparisons per line of the table. The * in the green fields indicate a pairwise comparison that reached statistical significance after Bonferroni correction with a value of p < .0038

Since a lot of comparisons of the PSE values resulted from the combinations of all identified scenarios with all proposed measures, only a part of the analyses can be presented in detail in the following: Pairwise comparisons between scenario baseline PSE and PSE with a measure applied will be reported for five scenarios that revealed particularly low PSE values and the three measures for each of the scenarios that had the largest effect in enhancing the PSE value. The five scenarios are: no help of other passengers (\(M = 30.67, SD = 21.25\)), blocked exit of the bus shuttle (\(M = 27.93, SD = 23.96\)), sensor uncertainty (\(M = 27.45, SD = 22.09\)), negative group dynamics (\(M = 29.05, SD = 18.88\)) and physical violence / sexual violence (\(M = 37.15, SD = 30.48\)).

For the scenario no help of other passengers in emergency situations, the measures that led to the largest increase in PSE were the mobile security service (\(M = 57.71, SD = 23.47\)), the video assistance system (\(M = 58.17, SD = 27.15\)) and the panic...
The increase in PSE reached statistical significance with \( p < .001 \) for each of the measures compared to the scenario baseline. The implementation of a panic button had the strongest effect on the increase of PSE, \( t(41) = -7.570, p < .001, d = -1.49 \), followed by the video assistance system, \( t(41) = -8.310, p < .001, d = -1.28 \) and the mobile security service, \( t(41) = -7.670, p < .001, d = -1.18 \).

The PSE rating for the scenario blocked exit was increased most by the same three measures, mobile security service \( (M = 50.93, SD = 29.60) \), the video assistance system \( (M = 45.12, SD = 27.94) \) and the panic button \( (M = 51.52, SD = 31.72) \). All measures significantly increased the PSE ratings, compared to the scenario baseline. The emergency brake was rated as the most effective solution in the scenario, \( t(41) = -6.672, p < .001, d = -1.04 \). Still, the video assistance system, \( t(41) = -4.947, p < .001, d = -0.76 \), and the mobile security service, \( t(41) = -4.762, p < .001, d = -0.76 \), revealed a medium to large effect on the PSE ratings.

To cope with the scenario sensor uncertainty the most effective measures appeared to be the emergency brake \( (M = 52.52, SD = 28.01) \), the mobile security service \( (M = 41.95, SD = 27.32) \), and the video assistance system \( (M = 43.81, SD = 27.52) \). All measures significantly increased the PSE ratings, compared to the scenario baseline. For the scenario of a negative group dynamic that could arise in the bus shuttle, again the trio of the mobile security service \( (M = 60.02, SD = 23.76) \), the video assistance system \( (M = 50.24, SD = 23.54) \) and the panic button \( (M = 59.76, SD = 26.22) \) received the highest PSE ratings from the participants of the questionnaire. All PSE differences reached statistical significance compared to the scenario baseline. The panic button, \( t(41) = -8.616, p < .001, d = -1.33 \), was rated as most effective to increase the PSE value, followed by the mobile security service, \( t(41) = -5.641, p < .001, d = -1.28 \) and the video assistance system, \( t(41) = -5.641, p < .001, d = -1.87 \).

In the scenario physical violence / sexual violence, again the mobile security service \( (M = 61.30, SD = 30.21) \), the video assistance system \( (M = 60.76, SD = 28.35) \) and the panic button \( (M = 64.91, SD = 26.07) \) led to the largest PSE ratings. All measures significantly increased the PSE ratings compared to the scenario baseline. The effect size of the panic button turned out to be the largest \( t(32) = -6.077, p < .001, d = -1.06 \). However also the mobile security service, \( t(32) = -4.873, p < .001, d = -1.33 \) and the video assistance system, \( t(32) = -4.839, p < .001, d = -0.84 \) revealed large effect sizes with regard to their potential to increase PSE ratings in the sample.

**Discussion**

Three groups of fears were identified: fears related to other passengers, fears related to a potential lack of transparency of the service leading to a state of uncertainty and
fears related to technical malfunctions and the design of the system. Especially fears related to potentially dangerous third persons that could threaten one’s own physical integrity turned out to be very relevant in the discussed mobility service. The fact that there is no driver in the vehicle who also serves as a supervisor appears to play a major role in the PSE ratings of persons who envisaged a flexible and autonomous bus shuttle service of the future. Low PSE values linked to fears like having no individual in the vehicle that is able to help in case of an emergency, being attacked or encountering negative group dynamics underline the particular relevance of fears of other persons. However, other fears like being unable to exit a crowded bus shuttle or being uncertain whether the sensors of the bus recognize obstacles reliably appear to be important for the PSE of potential passengers, too.

Those fears underline the relevance of taking into account the perceived safety of potential passengers in the design process of future bus shuttle systems. Among the measures presented, it is noteworthy that measures like the mobile security service, the video assistance system or the panic button turned out to have a strong positive effect on the PSE in the majority of aversive scenarios. The similarity of these three measures is that they ensure contact to an authority who is supposed to serve as a protection from various harms. It appears that the physical absence of a person that is a defining characteristic of autonomous bus shuttles might be a key challenge that needs to be taken into account in systems design. It is noteworthy, however, that neither in existing public transport systems like subways nor in trains a driver can be approached by passengers. A crucial difference to bus shuttles might be the size of the vehicle. When passengers encounter an unpleasant situation, like a crowd of drunken persons, they are able to create a greater distance to the crowd when they are in a subway car, compared to a bus shuttle. The lack of interpersonal distance could be a factor that leads to a decrease in PSE. As described, systems that enable direct contact to an authority might be a possible solution for this problem, a good interior design of the vehicle might be another.

It is noteworthy that this work has some shortcomings that limit the generalizability of the results. First of all, bus shuttle systems that operate without a driver do not exist in regular traffic to date. Thus, it is questionable how well participants of the questionnaire were able to envisage what it is like to use such a shuttle. A detailed introduction was provided in the questionnaire to clarify the concept as well as possible. Still, this introduction should be assumed to be the only reference point that participants had to the target transportation system. In addition, the described transportation concept had many characterising features, with the autonomy of the vehicle and the flexibility with regard to routes and schedules being the most prominent. Each of the characteristics might have its own particular set of fears that are connected exclusively to it. Therefore, it might be worthwhile to conduct additional research about characteristics of a transportation concept like autonomy and flexibility isolated from each other.

Furthermore, especially during the workshop, all of the participants were inhabitants of Berlin. This implies that the fears they expressed with regards to autonomous and flexible shuttle systems as well as the potential solutions, represent perspectives of urban citizens. Those fears and concepts could differ from the associations that
people from rural areas could have with the same transportation concept. Research on future flexible and / or autonomous bus shuttle systems should focus on rural areas as well, since these concepts appear to be promising solutions especially for these areas (König, Meyer & Grippenkoven, 2017).

As a general conclusion, the results of the workshop and the questionnaire underline, that besides the technological challenges that still exist in the field of driverless and flexible public transportation systems, psychological factors have to be taken into account as well. Regardless of the fact that PSE ratings of participants were made without previous experiences with such a system, the results reveal indicators that should be taken into consideration in the process of system development. Considering potential fears of prospective passengers in early stages of the development could help to avoid a public rejection and expensive failures when products are commercially launched in the mobility market.

Acknowledgment

This publication is based on excerpts from the master thesis of Zoë Fassina, who was responsible for the preparation, execution and evaluation of results of the described workshops and the online questionnaire. Illustrations in this publication are based on originals of the master's thesis by Zoë Fassina with the German original title Betrachtung des persönlichen Sicherheitsempfindens in Bezug auf geteilte, autonome Fahrzeuge im öffentlichen Personennahverkehr (Eng.: Evaluation of the personal experience of safety in shared, autonomous vehicles in public transport). The described studies of the master thesis were supervised by German Aerospace Center e.V. (DLR) and are embedded in the research activities of the Institute of Transportation System of DLR. The thesis will be handed in at the Department of Psychology and Ergonomics, Chair of Human-Machine Systems of TU Berlin.

References


Designing for Automated Vehicle and Pedestrian Communication: Perspectives on eHMIs from Older and Younger Persons

Ina Othersen¹, Antonia S. Conti-Küffer², André Dietrich², Philipp Maruhn², & Klaus Bengler²
¹Volkswagen Aktiengesellschaft, Wolfsburg, Germany
²Technical University of Munich, Garching, Germany

Abstract

The automation of automobiles requires much theoretical, legal and empirical work in order to define and ultimately resolve the complexities associated with it. One of the many challenges that automotive manufacturers face is how driverless, automated vehicles will communicate to other traffic participants. This issue is especially crucial for vulnerable road users such as pedestrians, where the consequences of miscommunication are potentially critical. One possible way for automated vehicles to communicate to pedestrians besides their movement behaviour is through an external human-machine interface (eHMI), where an explicit, informative message in the form on an icon (static) or dynamic animation, for example, is presented outside the vehicle. In this paper, we report the subjective results of an experiment investigating four different eHMI concepts in two age groups (20-30 and 60-70 years old). The participants were exposed to the concepts in an immersive, virtual environment, consisting of three different traffic scenarios, presented on a head-mounted display. The aim of the subjective inquiry was to understand the participants’ preferences regarding the investigated eHMIs experienced as well as their suggestions regarding improvements. Overall, participants preferred dynamic messages over static ones.

Introduction

With the increasing automation of vehicle control, the progressive complexity of traffic scenarios and interaction between road users, new possibilities are being considered to facilitate the communication between road users (Färber, 2015; Fekete, Vollrath, Huemer & Salchow, 2015; Zimmermann, Fahrmeier, Bengler, 2015). However, this issue has not yet been thoroughly, empirically investigated. Current research mainly focuses on, for example, CarXCar or environmental information acquisition (Federal Ministry of Transport and Digital Infrastructure, 2015) or the interaction between automation and vehicle occupants (Bendewald, Stephan, Petermann-Stock & Glaser, 2015; Othersen, 2016). Higher levels of automation also raise many unanswered questions, including: how could nearby human road users communicate and interact with automated vehicles or how may
current communication strategies be affected by the use of automated and driverless vehicles. Mostly, in conventional road traffic, an informal communication takes place in the form of non-verbal communication, e.g., gestures, facial expressions, and eye contact (Färber, 2015). These ways of communicating will no longer be available to pedestrians when autonomous vehicles are used due to the lack of a human driver. However, it can be assumed that the other road users’ need for information will not change in the first steps of the introduction of such a system (Färber, 2015).

Communication between a vehicle and a vulnerable road user (VRU, e.g., a pedestrian, cyclist, etc.) is of particular interest and must be examined more closely as a potential area of conflict with regard to higher levels of automation. The improvement of traffic flow and traffic efficiency has been specified as a potential benefit of automated and driverless vehicles (Friedrich, 2015). Studies have shown that pedestrians do not simply step in front of a vehicle but rather closely observe vehicle behaviour and safely adapt their own actions to that of the vehicle (Schweizer et al., 2009). The most important influencing factors are speed (Schneemann & Gohl, 2016; Habibovic et al., 2018), distance (Schneemann & Gohl, 2016; Habibovic et al., 2018), clear deceleration of the vehicle (Schneemann & Gohl, 2016), as well as vehicle behaviour and motion (Cramer, Siedersberger & Bengler, 2017; Habibovic et al., 2018). A study by Šucha (2014) revealed that 46% of participants waited for a vehicle to come to a complete stop and 84% sought eye contact with the driver before crossing the road. In crossing scenarios, the potential for an increase in traffic flow due to automated vehicles would therefore be obsolete or limited.

The objective of external communication is to facilitate symbiotic actions between road users and the recognition of intentions by displaying a current HMI on the exterior of the vehicle (external Human Machine Interface – eHMI). This could increase the cooperation between road users in manual and assisted driving and represent new communication channels for automated and driverless driving. In this context, new HMI components that make it easier for drivers or a VRU to communicate with other road users are being designed. Initial research looked at visual indicators on the exterior of a vehicle. These indicators comprised either displays or lighting elements on the front area of the vehicle (Lagström et al., 2015; Clamann et al., 2016; Fridman, Mehler, Xia, Yang, Facusse & Reimer, 2017).

In the reported experiment, the interaction between informal communication between pedestrians and vehicles, pedestrians’ adjustment of their road-crossing behaviour, and the clarity of the messages are investigated. The underlying research question in this paper is: What influence does a display-based eHMI have on pedestrians’ road-crossing behaviour? The content of the information, which can be displayed as either a lighting element or symbol, plays a particularly important role. Therefore, the aim is to identify road-crossing behaviour and to subjectively evaluate different variants of a visual indicator in a study conducted in a pedestrian simulator. In addition, the potential effect of age is considered in this study.
Methodology

Study setting and study design

The study was carried out in the pedestrian simulator at the Technical University of Munich (for more information on the simulator, see Dietrich, Willrodt, Wagner, & Bengler (2018)). The participants carried an HTC VIVE, which has two OLED displays with a resolution of 1080 x 1200 pixels, as a head-mounted display. A workstation (CELSIUS M740B Power) with an Intel Xeon E5-1620v4 (3.5 GHz) processor, 16 GB DDR4 RAM (2400 MHz), and NVIDIA Quadro M5000 graphic cards with 8 GB VRAM was used. To allow participants to feel as immersed as possible, real environmental sounds stemming from the laboratory were suppressed and sounds from the virtual world were played to the participants using BOSE noise-cancelling headphones. Unity 3d was used as the simulation software. The virtual environment resembled an inner-city situation from Munich. The roads were based on German standards with a width of 3.25 metres. The participants were able to move freely within an area of approximately 3 x 5 metres, and a blue grid was displayed to the participants to mark the end of this accessible area.

A 2x2x3x2 mixed design was used for the trial, with the following factors:

- eHMI type (lighting element vs. symbol)
- eHMI dynamics (static vs. dynamic)
- Scenarios (scenario A: 30 km/h on a straight road; B: 50 km/h on a straight road; and C: 30 km/h with a right turn)
- age groups (older vs. younger)

The last factor was the only “between” factor, so the participants were naturally divided into two age groups (older: 60-70 years, and younger: 20-30 years). The eHMI type, eHMI dynamics, and scenarios were implemented as a “within” design. The eHMI differed with regard to the type of display (eHMI type: lighting element vs. symbol) as well as the level of animation (eHMI dynamics: static vs. dynamic). Each display variant was shown in three different situations. In addition, each scenario was repeated without eHMI in order to collect baseline data for each scenario (baseline condition). The participants were given the task of crossing the road after the first vehicle. Following traffic showed variable gap distances in between the vehicles. Participants were introduced to cross whenever they felt safe in doing that. However, a certain vehicle in the line of traffic came to a complete stop and, where appropriate, communicated via the eHMI. The deceleration strategy of the yielding vehicle was identical within each scenario but the number of vehicles the participants could not cross in front of was pseudorandomised (a precise description of the study scenarios can be found in Dietrich, Othersen, Maruhn, Conti-Kufner & Bengler, 2018).

eHMI

For a better evaluation of the complexity of the information and the communication strategies, two different designs were used in the study (factor: type). Each design was displayed as a static and as a dynamic display (factor: dynamics). The static
display was primarily used to express the information “I have seen you,” and the dynamic display was more an illustration of the information “you may now cross the road in front of me.” Thus, these two types (see figure 1) had an effect on the communication strategy and complexity.

<table>
<thead>
<tr>
<th>Communication Strategies</th>
<th>Static</th>
<th>Dynamic / animated</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lighting element</strong></td>
<td>Light Bar</td>
<td>Knight Rider</td>
</tr>
<tr>
<td><strong>Symbols</strong></td>
<td>EyeCon</td>
<td>Walking Man</td>
</tr>
</tbody>
</table>

*Figure 1. Types of communication strategy with the two “between” factors: type and dynamics*

In the light-based design, a virtual LED bar was integrated into the grille. This variant approximated the various lighting designs (see Ritter, Adikari & Nagel, 2017). The static variant lights up once the vehicle is 50m away from the pedestrian and is deactivated again after the interaction has come to an end. In the dynamic variant of the light-based design, the LED bar was shown as a chaser light with a repeated sequence from left to right, indicating the crossing direction to the addressed pedestrian. A symbol-based design by Volkswagen Group Research was developed as a second design. In the static variant, an eye was shown, indicating that the pedestrian was perceived by the yielding vehicle. In the dynamic variant, a pedestrian was shown in a walking animation augmented by four animated arrows. It also included directional instructions to cross from left to right.

**Procedure**

Participants first filled out a demographic questionnaire and consent form. The participants were then introduced to the experimental procedure and VR pedestrian simulator. Prior to the experiment, height measurements of the participant were taken and their vision was assessed (colour vision and virtual reality visual acuity). Participants were instructed to cross the street whenever they felt safe to after the first vehicle has passed them. Additionally, participants were informed that some vehicles have a display, which enables them to communicate with their environment. Participants trained the crossing task in the VR environment until they reported to
feel comfortable with the task/environment and the experimenter judged their performance as satisfactory. The experiment was divided into three different blocks, representing the different tested scenarios: A, B, and C. Blocks and eHMIs were permuted for each age group. Between each block, participants were given a break and filled out a short questionnaire of whether they had any positive or negative feedback regarding the scenario or eHMIs experienced in the past block. After all blocks were completed, final questionnaires were administered to participants. The experimental session lasted approximately one hour.

**Data and analysis**

The current paper presents the results from different questionnaires given to participants during and at the end of the experiment as well as a short insight in the objective variable of crossing initiation time to vehicle stop (CIT\textsubscript{VS}) (for a detailed analysis see Dietrich, Othersen, Maruhn, Conti-Kufner & Bengler, 2018).

Various metrics to measure traffic efficiency on a macroscopic level or the severity of potential conflicts are widely utilized today to understand urban traffic. However, to quantify a pedestrian’s reaction to a yielding vehicle requires a metric for unhazardous situations. The crossing initiation time to vehicle stop is included in this paper as a metric to enable an objective evaluation of the participants’ subjective statements. The CIT\textsubscript{VS} is the difference between the time, at which the vehicle comes to a full stop and the time at which the pedestrian initiates his/her crossing (measured when the participant’s head moved 70 cm in crossing direction). If this value is negative, it means that the participant started to cross the road before the vehicle came to a standstill.

Several open questions, the User Experience Questionnaire (UEQ) created by Laugwitz, Schrepp & Held (2008) and an eHMI variant ranking, which were presented to the participants in the final interview, were used as a subjective data.

The aim of the open questions was to understand the participants’ preferences regarding the eHMIs presented as well as their point of view in terms of improvements. To this end, participants were asked to respond to 5 open-ended questions (the following has been translated from German):

- What did you like about the eHMI? Please explain your answer.
- What did you not like about the eHMI? Please explain your answer.
- If you could change something about the eHMI, what would you change (e.g., so that it would be clearer or more attractive)? Please explain your answer.
- Do you have suggestions for improvements for the eHMI (e.g., should it be presented somewhere else or should it look different?)? Please explain your answer.
- Which modality would you prefer for an eHMI (e.g., visual, acoustic, multimodal, other)? Please explain your answer.

The participants’ responses to the open questions were digitally transcribed, encoded and analysed using MAXQDA. In the qualitative text analysis, word and category
frequencies were analysed in order to draw conclusions on user attitudes (Mayring, 2000; Mey & Mruck, 2010).

The quantitative data (CTvS, UEQ, and ranking) were analysed descriptively and using statistical tests, i.e. t-test and 3 (scenario) x 5 (eHMI) x 2 (age group) mixed ANOVA with scenario and eHMI as within subject factors (overall four eHMI variants depicted in fig. 1 plus one baseline condition) and age group as between subject factor.

Participants

A total of 43 participants (N) participated in this study. The average age of the older group (60-70 years, n = 21, 43% female) was 64.7 years, and the average age of the younger age group (20-30 years, n = 22, 45% female) was 24.3 years.

Both groups were also asked about their usual mobility behaviour, e.g. the type of means of transportation used and the number of daily road crossings, and their communication in traffic, e.g. the number of daily interactions via glances, gestures and facial expressions. With regard to the mobility behaviour (see figure 2), the younger group used public transport most frequently, followed by car, and then travelling on foot or by bicycle. In contrast, the older group travelled most frequently on foot, followed closely by public transport, and then by car. Therefore, members of the older group are more likely to be involved in traffic as vulnerable road users.

Figure 2. Modes of transport, separated for the two age groups: older (60-70 years) and younger (20-30 years).
However, both groups indicated that they cross a road in traffic more or less equally often. In the two groups, approximately 46% do so often, approximately 46% sometimes, and 8% rarely. This indicates that younger road users either participate less in road traffic as a pedestrians, but do more road crossings on average than older road participants or that the older participants were more restrained when answering the questionnaire. In addition, the participants were asked about how often they communicate in traffic using glances, gestures, and facial expressions. The older group report using such non-verbal communication more often (43% often, 48% occasionally, and 9% rarely) than the younger group (23% often, 41% occasionally, and 36% rarely).

**Results and subjective insights**

*Crossing initiation time to vehicle stop (CIT\_VS)*

The road crossings of the 43 participants, of which 34 completed all conditions\(^1\), were compared in order to identify the effect of the different eHMI variants compared to one another and to identify a baseline for their road-crossing behaviour. There were no interaction effects between Scenario, Age Groups and eHMI variant. A main effect analysis shows that there is a significant effect of the eHMI variant on the CIT\_VS \((F(4, 29) = 25.58, p = .000, \eta^2_p = .78)\). A Bonferroni post hoc test revealed significant differences between the crossing initiation times in the baseline condition (mean value of 1.07s, SD=0.99s) compared to the eHMIs (EyeCon 0.39s ± 1.14s, Knight Rider 0.23s ± 1.27s, Light Bar 0.52s ± 1.15s & Walking Man 0.15s ± 1.10s). These results indicate that participants started crossing the road earlier if there was an eHMI presented. Furthermore, the CIT\_VS is smaller with dynamic eHMI displays, i.e. Knight Rider and Walking Man, than with static ones (see figure 3). The data also revealed that 28% of the participants crossed the street before the vehicle came to a full stop when the approaching vehicle had an eHMI equipped in comparison to 9% without an eHMI.

There was a significant effect of the scenario where participants encountered the traffic conditions on their CIT\_VS \((F(2, 31) = 7.21, p = .003, \eta^2_p = .32)\). A Bonferroni post hoc test showed that the crossing initiation time in scenario A (30km/h, straight road: 0.22s ± 1.24s) was significantly lower compared to scenario B and C (50km/h, straight road: 0.59s ± 0.99s; 30km/h, right turn: 0.71s ± 1.19s).

No difference in CIT\_VS between the two age groups was found \((F(1, 32) = 0.05, p = .820)\). Older participants had a CIT\_VS of 0.54 seconds \((SD=1.16)\) compared to 0.47 seconds \((SD=1.16)\) of the younger participants.

\(^1\) The yielding vehicle fully stopped for 2 seconds before accelerating again. Nine participants did not initiate a road crossing in time. Therefore 34 from 43 participants were included in the statistical analysis.
Open questions

The final evaluation open questions (questions 1 and 2), which asked about positive impressions and suggestions for improvement, resulted in more positive than negative comments. The responses to these questions were assigned to one of the following categories: Understandability, Perceptibility, Appeal, Safety, and Other response (see tables 1 & 2).

The “Walking Man” eHMI was mentioned in 23 responses and achieved the most positive responses for the categories Understandability, Perceptibility, and Appeal in both age groups. In particular, it was positively commented that the dynamic display clearly signalled the vehicle’s intention and enabled the next action to be inferred. Likewise, the dynamic “Knight Rider” eHMI was positively evaluated and received a total 17 mentions in the same categories.

- Example regarding “Walking Man”: “It provides security and confirms to pedestrians/me that I can safely cross the road. I found the dynamic variants intuitively easier to understand. I found the size of the display to be just right. Pictograms are good.” (younger participant; translated from German)
- Example regarding “Knight Rider”: “Dynamic displays could be perceived early, and you understood in which direction pedestrians were recognised/could go. Over time, you become confident that cars with displays will really stop for you, and you gain confidence. It is good that it is always consistently in blue tone.” (older participant; translated from German)
Table 1. Frequency of positive and negative statements for questions Q1 & Q2 separated by age group – Categories “Understandability,” “Perceptibility,” “Appeal” (y: younger group; o: older group)

<table>
<thead>
<tr>
<th></th>
<th>Understandability</th>
<th>Perceptibility</th>
<th>Appeal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>positive</td>
<td>negative</td>
<td>positive</td>
</tr>
<tr>
<td>Knight Rider</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Walking Man</td>
<td>6</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>EyeCon</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Light Bar</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Other response</td>
<td>5</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Overall</td>
<td>15</td>
<td>13</td>
<td>13</td>
</tr>
</tbody>
</table>

Table 2. Frequency of mentions in positive and negative statements for questions Q1 & Q2 separated by age group – Categories “Safety” and “Other” (y: younger group; o: older group)

<table>
<thead>
<tr>
<th></th>
<th>Safety</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>positive</td>
<td>negative</td>
</tr>
<tr>
<td></td>
<td>y o y o</td>
<td>y o y o</td>
</tr>
<tr>
<td>Knight Rider</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walking Man</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EyeCon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Light Bar</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other response</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Overall</td>
<td>8</td>
<td>6</td>
</tr>
</tbody>
</table>

In contrast, the “EyeCon” and “Light Bar” eHMIs were only given suggestions for improvement, as the content was not considered informative, or immediately and clearly perceptible. In addition, the participants stated that it was difficult for them to see the static displays relating to the vehicle status or the future behaviour of the vehicle.

- Example regarding “EyeCon”: “The static displays attract your attention less, are easier to overlook…” (older participant; translated from German)
- Example regarding “Light Bar”: “The static displays without arrows are hard to understand, you have to think about them a lot. Possible distraction for oncoming traffic.” (younger participant; translated from German)

Overall, the participants were able to give additional free comments that did not explicitly relate to one of the eHMI variants (question 3 & 4). In particular, the eHMIs’ potential ability to contribute to the pedestrians’ subjective feeling of safety and trust by clearly displaying a vehicle’s status and facilitating the anticipation of vehicle behaviour was mentioned as positive in the Safety category. For the above reasons, large (22% of younger participants and 14% of older participants) and
multi-coloured symbols (19% of younger participants and 23% of older participants) with a dynamic display were mentioned and evaluated as quickly learnable. The negative comments mostly referred to a possible familiarisation effect, which could lead to a certain insecurity with vehicles without eHMI. The position on the vehicle and the risk of confusion with the regular lighting were also mentioned. Younger participants recommended using the complete radiator grille exclusively for the eHMI. In contrast, the older participants suggested a completely new location, e.g., in the area of the windshield based on the visual axis to today’s drivers in this area.

In response to the question on the type of eHMI (question 5, multiple responses possible), the participants stated that a visual display was their main preference (60% of younger participants’ responses and 50% of older participants’ responses). A multimodal design was positively evaluated by 36% of younger participants and 44% of older participants. On the one hand, environmental conditions were given as reasons for this (“Several modalities are imperative, as depending on the time of day and how much sun there is, purely visual displays are not always easy to see.”) and, on the other hand, impairment or distraction of people was mentioned (“With a visual/acoustic design, you can reach everybody – including blind and deaf people.”, “If you are wearing headphones or are looking at your mobile phone, you can still be contacted.”). A purely acoustic design only received one vote per age group.

**UEQ and ranking**

The UEQ showed a positive evaluation for the general use of an eHMI with an evaluation over +0.8 points (see figure 4). The evaluation did not differ significantly between the age groups (attractiveness: t(37) = -1.65, p = 0.11; efficiency: t(37) = 0.63, p = 0.63; perspicuity: t(37) = -1.36, p = 0.18; dependability: t(37) = 0.15, p = 0.86; stimulation: t(37) = -0.17, p = 0.87; novelty: t(37) = -0.15, p = 0.88).

![Figure 4 UEQ scores, separated for the two age groups: older (60-70 years) and younger (20-30 years).](image)

Finally, the test persons were able to give an all-over evaluation by a ranking the four eHMIs. The ranking likewise showed a preference for the dynamic displays followed by the static displays. The eHMI with the dynamic symbol was ranked
first, followed by the dynamic lightning element in second place. There was very little difference between third and fourth places, with third place being taken by the static icon and fourth place by the static lightning element.

**Discussion and summary**

The aim of this experiment was to investigate the effect of eHMIs on pedestrian experience and road-crossing behaviour. To this end, road-crossing behaviour and the subjective evaluation of the different variants of a visual display for two different age groups were assessed.

The results for road-crossing behaviour, measured using crossing initiation time to vehicle stop (CIT\(_{VS}\)), showed that participants crossed the road sooner when an eHMI was used. In addition, data revealed that with an available eHMI 28% of the participants crossed the street before the vehicle came to a full stop in comparison to 9% without an eHMI. Dynamic displays seem to potentially enhance this effect, where no clear difference was observed between the dynamic lightning element and the dynamic symbol. This study in the pedestrian simulator showed first tendencies that the safety behaviour of pedestrians can potentially change with using an eHMI and that the willingness to cross a road potentially increases before the vehicle has come to a standstill. However, as the study was conducted in a VR simulator and only covered somewhat simple scenarios, external validity of the results and the applicability of eHMIs in real world urban traffic needs to be reassessed in future studies.

Positive implications for the use of an eHMI can also be taken from the subjective data, which comprises open questions, the UEQ, and the ranking as an overall evaluation. The eHMIs as a whole were positively evaluated in terms of attractiveness and perspicuity above all in the UEQ. According to participant statements, the eHMI can be understood as a design element, but can also make an important contribution to the understanding of a vehicle's future status. This picture also becomes clear from the categorised responses to the open questions and the ranking. The dynamic displays received predominantly positive responses for the categories “Understandability,” “Perceptibility,” and “Appeal” in both age groups. In contrast, the static displays were not evaluated as immediately and clearly perceptible, and the participants found it difficult to draw conclusions on the vehicle's future status from the static displays. The static displays occupied the last two places in the ranking. As potential improvements, the participants mentioned the use of animations for an eHMI or the use of existing colours or road signs. With regard to a preferred modality, the majority of participants were also in favour of a visual design. However, the participants also saw the advantages of a multimodal design, comprising a visual and acoustic content. In this regard, potential environmental constraints (e.g., weather conditions or interference from the environment) or constraints relating to the recipient themselves (e.g., impairment or distraction of people) play a decisive role.

This experiment was able to show that the use of an eHMI could generate potential benefits for interactions with an automated or driverless vehicle, e.g., an increased feeling of safety and quicker road-crossing behaviour. This enables optimal...
adaptation of the human-machine interaction to the new form of communication. However, further studies are required to investigate additional display options and other scenarios. It is also necessary to look in more detail at road-crossing and hedging behaviour depending on speed, distance, and vehicle movements, as these factors appear to have a significant effect (Cramer, Siedersberger & Bengler, 2017; Habibovic et al., 2018). Future experiments should also consider to be conducted in real driving environments to ensure validity.

References


Färber, B. (2015). Kommunikationsprobleme zwischen autonomen Fahrzeugen und menschlichen Fahrern [communication problems between autonomous vehicles and human drivers], In M. Maurer, J.C. Gerdes, B. Lenz, and H. Winner (Eds.), Autonomes Fahren (p. 127-146), Berlin: Springer Vieweg.


Fridman, L., Mehler, B., Xia, L., Yang, Y., Facusse, L.Y., & Reimer, B. (2017). To Walk or Not to Walk: Crowdsourced Assessment of External Vehicle-to-
designing for automated vehicle and pedestrian communication


Othersen, I. (2016). Vom Fahrer zum Denker und Teilzeitenker - Einflussfaktoren und Gestaltungsmerkmale nutzerorientierter Interaktionskonzepte für die Überwachungsaufgabe des Fahrers im teilautomatisierten Modus [From driver to thinker and part-time driver - influencing factors and design features of user-centered interaction concepts for the monitoring task of the driver in the partly automated mode]. Wiesbaden: Springer Fachmedien.


Cooperation through communication – Using headlight technologies to improve traffic climate

Matthias Powelleit, Susann Winkler, & Mark Vollrath
Technische Universität Braunschweig
Germany

Abstract

In order to promote cooperative interactions and positive encounters in daily road traffic and to improve traffic situations which lack communication, a cooperative laser beam based on new light technologies is developed. Via light projections, drivers can get in touch with other road users. In a first step, relevant traffic situations and the situation-specific needs of diverse road users were identified by means of an online survey and a one-week diary study including car drivers, cyclists, and pedestrians. From the results, scenarios were derived, which may benefit from the use of new communication technologies. A selection of these scenarios was implemented and evaluated in a driving simulator using a small-scale prototyping study (n = 7) which is described in this paper. Generally, drivers reacted positively to the new communication possibilities, which differed in their understandability and emotional effects. Furthermore, a visualization of the predicted driving trajectory showed the potential to distract drivers and alter their self-reported gaze and driving behaviour. Overall, a cooperative laser beam offers great opportunities to improve the traffic climate. However, further applications and safety-relevant consequences should be discussed.

Introduction

Motivation

Traffic relies heavily on the smooth interaction between road users. A mutual understanding of intentions and the reasons for certain driving manoeuvres is of vital importance. Drivers are taught in driving school to drive vigilantly, cooperatively, and communicate their intentions clearly, especially when interacting directly with other road users.

Several forms of communication can be found in the traffic environment. De Ceunynck et al. (2013) differentiate implicit and explicit communication, using the example of vehicle interactions at intersections. While implicit communication is essentially conveyed via the vehicle’s motion (e.g., accelerating, decelerating, steering), explicit communication comprises specific signals which are not essential for the driving task itself but are used for interacting more directly with other road users (e.g., gestures, turn indicators, horn honking, flashing headlights). However, communication (implicit or explicit) is often ambiguous and has to be understood in
the current driving context. A driver may brake, for example, to turn at the next intersection. A pedestrian at the road side might understand this action as (implicit) invitation to cross the road, even though the driver might not have recognized the pedestrian. This scenario can easily turn into a critical situation. Enhanced communication between road users can serve to clarify such situations and contribute to traffic safety.

Additionally, enhanced communication might also contribute to an improved traffic climate, as missing or misleading information can cause frustration and anger (Dollard, 1939; Merten, 1977). Research has shown that driving anger is related to aggressive driving (Risser, 1985; review given by Zhang & Chan, 2016) and impairs attention allocation, reasoning, judgement, and decision making (Blanchette & Richards, 2010), which are crucial to the driving task and traffic safety.

In contrast to well investigated negative emotions in traffic, there might also be the chance to elicit positive emotions. For example, sharing clearer information about current and future behaviour as well as expressing appreciation for cooperative gestures might influence the traffic climate positively. However, the use of available communication channels and signals is not standardized, even though every vehicle has the same signalling devices. A specific signal can be used to convey different and even opposing messages. Flashing the headlights can mean “Please, go ahead” or “Caution, I am moving ahead” (Merten, 1977). As current vehicle signals are often ambiguous, the perceived message depends highly on the context and can only be understood correctly if the communicators share the same context. In addition, more personal forms of communication like gestures might not be applicable in every situation. For example, at large distances or at night the vehicle driver might not be visible so that such explicit communication forms cannot be utilized.

A new communication channel, designed to transmit specific messages, could address the described problems and support clearer communication in traffic. New signals can be designed with specific messages in mind and as such be more personal than, for instance, flashing headlights. A lot of research is starting to focus on communication between automated vehicles and other traffic participants, like pedestrians. Many of them are using external human machine interactions (HMI; see Deb, 2018; Dey et al., 2018; Dey & Terken, 2017; Hudson et al., 2018; Schieben et al., 2018; Song et al., 2018; Witzlack et al., 2016). The HMI concepts vary from displays in the windshield or on the vehicle roof to audio messages. This paper follows a different approach. Technological advances in the field of vehicle lights allow new ways of communicating by projecting information onto the road surface, for example, by using laser beams integrated into the vehicle’s headlights. Further information on the research aims and the development of scenarios will be given in the following.

Objectives

Within the research project “KoLa” (cooperative laser beam, Fraunhofer ISIT, Volkswagen AG, Universität Siegen, Technische Universität Braunschweig), new forms of communication are developed, which enable drivers to express unambiguous messages and show more cooperative behaviour. These enhanced
communication possibilities might improve traffic climate and in turn support traffic safety.

To identify possible fields of application for the cooperative laser beam, two exploratory studies were conducted in collaboration with colleagues of the Universität Siegen and the Volkswagen AG. An audio diary study (N = 27) focused on situations in which road users could benefit from (better) communication and cooperation. For a time period of one week, participants documented everyday traffic situations that were characterized by the notable presence or absence of communication or cooperation. Using a voice recorder, data was collected as soon as possible after the occurrence of a situation (but not while driving). All modes of transport of the participants and interaction partners were regarded (by foot, bicycle or motorized vehicles). In addition, an online survey (N = 165) focused on factors that let a traffic situation be experienced as either positive or negative. Participants described a positive or negative situation (random assignment) they had encountered within the last four weeks. Again, all modes of transport were included as long as one of the interaction partners was a vehicle driver.

Using qualitative data analysis methods, profiles of the situations were generated to identify cases in which communication could have clarified or deescalated situations, motivated or facilitated cooperation and in turn created a more positive situation experience. This way, traffic scenarios were derived that could potentially be enhanced through improved communication and cooperation. In addition, the reported intentions, expectations, feelings, needs and wishes of the participants in the situations were used to derive cooperative messages that might have affected the situation experience positively. Finally, ideas and concepts were developed on how to articulate and present such cooperative messages in the according situations. More detailed information on the design of both studies and their results will be presented and discussed in a different publication.

Building upon this work, selected communication-relevant scenarios were implemented in a driving simulator environment to investigate different communication concepts, two of which will be presented in this paper (“Thanking” and “Trajectory display”). For all concepts, different design choices (graphics and animations) were compared regarding drivers’ ratings (e.g., comprehensibility, appeal, and distraction potential).

Method

Design

Since the aim of this study was to investigate multiple design parameters of the light communication concepts, a small-scale prototyping approach was used to gain first insights and then take the feedback into consideration for further design evaluations. Each participant experienced and rated different versions of two communication concepts (“Thanking” and “Trajectory display”). The sequence of the two concepts and the sequence of the different versions were randomized for each participant.
Participants

Seven test persons participated in this study. They were recruited through personal invitations. Six participants were experts in the fields of traffic psychology, safety, and human machine interaction, while one had no study-related professional background. The participants were aged between 25-54 years (M = 34.1 years; SD = 10.4 years). Five participants were female, two were male.

Driving simulator and implementation of light concepts

The study was conducted in the Department of Engineering & Traffic Psychology at the Technische Universität Braunschweig. A fixed base driving simulator was used, running SILAB 5.0 simulation software (Krüger et al., 2005; see www.wivw.de). The mock-up consisted of a driver seat, a steering wheel and pedals. Three TV screens (each 1920 x 1080 pixels, 55” diagonal) presented the simulated environment, covering approximately a 180° field of view. Sound was presented via stereo speakers and a subwoofer. The experimenters sat next to the participant, giving instructions and leading through the interview. No driving data was recorded during this study as the setup served a mere visualization purpose. The simulated environment was set to night time (see Figures 2 and 3).

The simulation software had no native support for user defined light projections. Pre-tests were done to find ways to a) implement individual light sources, b) customize the light sources in order to project custom graphics undistorted onto the road surface, and c) animate the projections. As a result, grayscale images were used as an alpha mask to define the shape and shade of the light projection, very similar to a slide projector (Figure 1). Simple animations of the projected graphic objects (translation, rotation, uniform scaling) were done within the simulation environment by moving and rotating the light source.

Figure 1. Illustration of how the light projections were implemented in the driving simulator.
For more complex animations (unidimensional scaling and bending of texture graphics) the grayscale image was replaced by a video source that acts as a dynamic texture (available in SILAB software since version 5.0). To trigger and control the video animations, an application was written in Processing 3 (www.processing.org, Processing Foundation), which realized custom vertices animation, based on driving parameters from the driving simulation software. The animations were then handed to the simulation software via video stream, integrated as a dynamic texture for the light source, and finally rendered as light projection onto the street, similar to a film projector.

Scenarios and communication concepts

To simplify the interaction and at the same time be more realistic with regard to the penetration rate of new technologies, always only one vehicle was equipped with a projection-based communication system.

Scenario “Thanking”

In the scenario “Thanking”, the participant had the opportunity to give way to another driver (waiting at the driveway of a property) by flashing the headlights. In the previous studies, participants reported that the recognition of their gesture by the other road user (nodding the head, raising the hand, or smiling back) was an important part for the interaction to be perceived as positive. Hence, for this scenario, different variants of a concept were developed to acknowledge the participant’s gesture. After the participant let the waiting vehicle turn into the lane, the now leading vehicle expressed gratitude via a backwards light projection.

Symbols served as light-based “Thank you!” messages as written language comes with several disadvantages (language barriers, takes longer to be understood, required resolution). To vary the tone of the message, three different symbols were presented to the drivers: A) a check mark, B) a smiley, and C) a thumbs up (Figure 2, left). Furthermore, three animation styles were varied to check if the participants feel differently addressed by the messages. In version A, the check mark symbol appeared at a fixed distance behind the leading vehicle. In version B, the smiley symbol seemed to be laid down onto the road by the leading vehicle as it stays at the same location (fixed to the road), while the leading vehicle moved on. Just before the participants reached the location of the projected smiley and were about to drive over the projection, the symbol popped up slightly and then shrank until it vanished completely. This animation should give the impression that the symbol responded to the participant’s vehicle. In version C, the thumbs up symbol appeared at the rear end of the leading vehicle, then moved towards the participant’s vehicle, and finally faded out before “touching” the participant’s vehicle.
In contrast to the “Thank you!” scenario, the participant is the sender of a message in the “Trajectory” scenario. However, in this case no specific situation was realized. Instead, the participant’s car was continuously projecting the own driving path in front of the vehicle. This was done passively without any actions being required by the driver. Such visualization could help other road users to recognize the driver’s actions and intentions even from a distance. For example, participants of the previous studies reported that, as pedestrians, they have difficulties to recognize speed changes when a driver slows down to give them way. Furthermore, due to reflections or dim lighting, gestures of the driver may not be visible through the wind shield. In such a situation, the projection of the vehicle’s driving path could visualize the speed reduction, e.g. by retracting the size of the display and even further explicate the driver’s intention by projection gestures onto the street.

Independently of the specific use case, the effects of such dynamic displays on the driver’s sensation of the driving environment were of interest to this part of the study. As it might be the case with all additional information displays, this visualization may cause irritation, especially as it dynamically feedbacks the driver’s behaviour.

In this scenario, a flexible light band was projected directly in front of the own vehicle that changed its appearance in order to represent the anticipated driving path of the vehicle. Thus, it extended with acceleration and retracted with deceleration.
and finally disappeared when the vehicle came to a full stop. When the participants used the steering wheel, the light band bent accordingly. The animation parameters for the length and curvature of the projected light band were adjusted to match the driving path within the next second, based on the current speed and steering wheel angle. Ten different texture styles (designed by project partners of the University of Siegen) of the light band were implemented, with varying contrasts, shapes, and line widths (Figure 3, left).

Procedure

Participants were introduced to the procedure and general aim of the study, then gave informed consent to audio and video recording. They sat down on the driver seat and were told to drive through a short urban track, adhering to traffic rules and a speed limit of 50 km/h.

Both scenarios were first experienced in a baseline drive without any additional communication concepts. This way, participants could compare their driving experience without and with light projections. Afterwards, participants were shown each version of the light concepts. Participants were instructed to think aloud while driving, expressing their spontaneous thoughts, emotions, and reactions to the presented concepts openly. In addition, a structured interview and rating scales (Heller, 1982) were used to evaluate factors like appearance and understandability of the concepts. Finally, the participants were asked to pick their favourite version for both scenarios, make improvement suggestions for future designs. The evaluation of each scenario took around 20-40 minutes, resulting in a total of 40-80 minutes per participant.

Results

As the aim of this study was to gain first insights into the design of light-based communication concepts, data analysis focussed on the audio recordings (thinking aloud and open interview questions).

Scenario “Thanking”

Animations
With regard to the animation of the symbols, participants felt less addressed by version A (locked distance to leading vehicle) compared to version B and C. The locked distance gave the impression that the symbol (check mark) belongs to the leading vehicle and represents some kind of vehicle status rather than a message directed to the participant. Furthermore, with increasing distance to the leading vehicle, the projected symbol became smaller and thus more difficult to recognize for the participant in the following vehicle.

Version B and C differed strongly in their presence. The smiley symbol (B) “rested” steadily on the road until the participant reached its position, whereas the thumbs up symbol (C) was visible shorter as it moved towards the participant’s vehicle. As the participants reported, the resting smiley was perceived more clearly but might bind the driver’s attention more. The moving thumbs up symbol was perceived as more
subtle, but drivers might oversee it more easily. Furthermore, if the symbol moves too fast towards the following vehicle, the message might be perceived as hasty and less sincere, as one participant indicated. Regarding the pop-up animation of the smiley symbol, some participants expressed having an enjoyable experience comparable to collecting tokens that they can take home like “karma points”.

**Symbols**

Participants perceived very different messages from the presented symbols. The smiley symbol conveyed a message which can be summarized as “I am happy (with your actions)”. The thumbs up symbol was understood as saying something like “Your actions were good”. In contrast, the check mark symbol was understood as “Your actions were correct”.

Participants perceived the smiley symbol as very friendly and the other driver as open and communicative. Participants reported that the other driver felt very happy and wanted to share his or her feelings with them. The driver, who projected the thumbs up symbol, was perceived as cool but more distanced than the driver with the smiley. Thus, the thumbs up symbol was rated as more neutral than the smiley but friendlier than the check mark. Participants considered the check mark symbol to be very impolite as they felt judged from a higher moral ground by the other driver, such as in a student-teacher setting. Thus, the driver with the check mark was perceived as know-it-all and arrogant.

The results from the rating scales were in line with the statements that were verbally expressed while driving. As Figure 4 shows, the smiley symbol was overall judged the most positively and was rated as very understandable, very distinct and very pleasant. None of the symbols was perceived as disturbing.

![Figure 4. Average ratings of the three “Thank you!” symbols.](image)

**Scenario “Trajectory”**

The projected light band was designed to represent the vehicle’s path one second ahead. When drivers enter an intersection or tight curve this bar could overlap other
lanes (Figure 5, left) or the side walk (Figure 5, right). This overlap reportedly influenced drivers’ steering behaviour. For example, drivers started steering earlier than necessary for the given road geometry. In general, the projected light band confused participants, as they tried both staying in the lane with their vehicle and avoiding overlaps of the light band with lane markings.

Figure 5. Light band of the “trajectory display” overlapping with lane markings (left) and the side walk (right).

In addition, the light band was perceived as rather intrusive and hard to ignore during driving. Participants reported to show less attention to peripheral objects (e.g., traffic lights and pedestrians) and the area near the horizon which is normally focused for foresighted driving. Instead, participants described a tendency to fixate the tip of the projected light band to avoid overlaps with the lane markings. Such distractive effects were especially high when the texture of the projected light band had strong contrasts, contours, and high opacity (see Figure 3, left, e.g. texture D and E). An additional problem arose when the light band had patterns or contours that were similar to road markings, such as zebra crossings or lane markings (see Figure 3, left, e.g., texture C to E).

Drivers rated the trajectory display as moderately understandable, distinct, helpful, and pleasant. The usefulness was rated as rather low. Participants wondered about the purpose and meaning of the light band. Most participants misinterpreted the light band as the stopping distance; some regarded the curved light band as an instruction to where they should steer. While participants reported that they might find the trajectory display useful for the external world like pedestrians or cyclists, they would not want to have it in their own vehicle as it was distracting and disturbing and added no additional information or value for themselves.

Discussion

This study investigated how a new communication channel (light-based projections onto the road) can be used to promote cooperative behaviour in road traffic and thus positively influence the overall traffic climate. Using the example of two scenarios “Thanking” and “Trajectory display” an explorative driving simulator study analysed the potential of light-based communication.
In the scenario “Thanking”, participants took on the role of a prosocial road user by giving up their right of way to let another vehicle turn into their lane. The turning vehicle expressed its gratitude by utilizing light projections that were directed towards the participant in the following vehicle. The influence of different symbols (check mark, smiley, and thumbs up) and animations (fixed to the leading vehicle, fixed on the road, and moving towards the following vehicle) on the content and tone of the messages were investigated.

Depending on the symbol, the driver of the leading vehicle was perceived differently, ranging from “friendly” and “open” (smiley) to “cool” and “reserved” (thumbs up) to “arrogant” and “know-it-all” (check mark). Participants preferred the smiley symbol as it conveyed a simple emotional response of the other driver (“I am happy”). This message seems to work similar to facial expressions in personal interaction: No explicit message is conveyed, but rather an emotional state that is implicitly related to the context. As the driver of the leading vehicle decides to send this state explicitly, he or she is perceived as very open and communicative, and as someone who cares about other road users.

Most of the participants started smiling themselves and some continued even long after the scenario had finished. This observation shows that emotions can be conveyed by light projections of a vehicle, even though the driver is not visible and enveloped by the car’s shell. Furthermore, these emotions had a positive effect on the participants’ own emotional state, which might have implications for the traffic climate more generally. Especially regarding the large effect driving anger can have on the driving behaviour and traffic climate (see Bliersbach et al., 2002; Levelt, 2003; Maag, 2004), there might also be potential of beneficial safety effects when drivers feel happy while driving.

In the scenario “Trajectory display”, the effects of a dynamic visualization of the vehicle’s driving path were investigated. The visualization was projected in front of the own vehicle in the form of a flexible light band with different versions of textures. The visualization was perceived as rather intrusive and distracting, especially when textures with high contrast and opacity were used. At the same time, the trajectory display offered no value for the drivers. Even more, participants reported negative influences of the display on their gaze and driving behaviour. This finding revealed conceptual problems that arise when predicted trajectories are to be displayed, based on current vehicle dynamics parameters. As the trajectories were predicted one second ahead, the visualization was incongruent with short term changes of the driver’s behavioural intentions. This was especially contradicting for lateral movements, like turning at intersections or short corrective actions.

While other road users may benefit from a trajectory display that enhances the predictability of a vehicle, the findings of this study strongly suggest looking for alternative ways of visualization. One solution may be to disregard the steering wheel action and only focus on longitudinal movements in order to avoid overlapping of the projection and lane markings. In this case, the light band may follow the shape of the road geometry. Future studies should also consider the perspective of other road users in order to verify their need for such a system.
Furthermore, varying levels of information detail will be evaluated to find out if a generic message can be used for multiple scenarios without compromising understandability of the message. In addition, further investigations may focus on situational factors (visibility, movement, and type of the sending vehicle) more closely, as the perceived message might depend on these factors.

Conclusion

In this study, light projections were tested for their applicability as a novel communication channel. In the context of expressing gratitude, it is recommended to use simple emotional messages as found in non-verbal communication in personal interactions to create positive experiences. Furthermore, a dynamic visualization of the driving path in front of the own vehicle was experienced as distracting and disturbing without offering any additional value for drivers. Future evaluations will investigate the influences of varying levels of information detail and situational parameters on the understandability of perceived messages.

References


Improving cooperation in traffic - Novel HMI approaches to enhance driver-driver interactions

Adrian Haar1,2, Andro Kleen1, Martin Schmettow2, & Willem B. Verwey2
1 Volkswagen Group Research, Wolfsburg (DE)
2 University of Twente, Enschede (NL)

Abstract

The strong increase of traffic in a limited infrastructure requires drivers to cooperate in order to arrive safely and on time at their destination. However, 30% of these interactions fail (Benmimoun, Neunzig & Maag, 2014). Recent technological developments provide the opportunity to support these interactions. In a simulator study (N = 50) participants drove on a two lane motorway and were confronted with lane-changes by other cars. During these lane changes, participants were supported by three novel support systems. In one condition, the indicator of the merging car distinguished between planning and executing the manoeuvre. In another condition, we added a head-up display (HUD) that signalled the upcoming lane-change of the partner car. In a third condition, the HUD also indicated the distinction between planning and execution of the manoeuvre. The results showed that the novel indicator concept in the first condition was not intuitively interpretable by the participants. However, providing information via the HUD did show a strong effect on cooperative behaviour of the participants. This effect was stronger as the participants could distinguish planning and execution of the manoeuvre.

Introduction

Traffic itself is one of the most complex interactive systems that individuals on earth deal with on a daily basis. In order to keep this system running a lot of formal traffic rules have been established to prescribe behaviour of everyone acting within this system. On the one hand, these rules have the benefit that every individual has a formal guide of what to do and what to avoid in traffic. This helps everyone to behave in a save way, in addition it also makes the behaviour of others predictable. Everyone in traffic will behave, in most cases, according to the rules of traffic. On the other hand, these rules help to juridically determine who was right and who was wrong after an accident has occurred. However, the formal and juridical character of those rules do not guarantee a positive and smooth driving experience. Especially in situations in which drivers interact with each other, acting strictly after the established rules sometimes hinder such a positive experience. Drivers who act
permissively, not insisting on their right of way, contribute to a smoother traffic flow, which would result into a more positive driving experience (Hidas, 2002).

Then again, permissive behaviour, or in more general context cooperative behaviour, of drivers does not follow the formal rules of traffic and is therefore not predictable to other traffic members involved in the situation. To make behaviour more predictable, drivers need to communicate their intentions in such driver-to-driver interactions more clearly. However, the only explicit means of communication for this task drivers possess today is the turn indicator. This yellow, flashing light has been constructed for a very specific purpose, namely communicating that a turn is to be engaged. However, it does not fit the wide variety of interactions that are possible in contemporary traffic. Other means of communication, for example facial expression, headlight flash or the horn, are sometimes used to solve this problem, however, they differ widely in their precise usage and, in the case of the headlight flash or the horn, are only allowed in case of imminent danger.

One solution to this problem is the development of modern vehicle-to-vehicle (V2V) communication technologies. Improving driver-driver interaction by sharing information and listening to other systems will be a central issue in future mobility. A better "getting along" with others and an effective accident prevention will also be reflected in a better driving experience and increased comfort of individual mobility (Fekete, Vollrath, Huemer, & Salchow, 2015). Another reason to improve the interaction between drivers is the future introduction of automated driving functions. These functions can provide information about planned manoeuvres to other traffic members, without the need for manual disclosure, explicit communication or model-based behavioural predictions. If an automatic system is planning a lane change, this intention can be made accessible by a general or local limited broadcast and received by the affected vehicles at the same time. This information can then be used to increase the clarity of the behaviour to other traffic members that will be affected by the actions of the automated vehicle.

One important example for interaction in traffic is the lane change manoeuvre (Ellinghaus, 1986). The lane change manoeuvre is a complex task and 9% of accidents are related to this manoeuvre (Sen, Smith, and Najm, 2003). According to German traffic regulations (StVO, §7), every lane change has to be signalled clearly and early enough, and the driver on the target lane has the right of way (Leue, 2017). During a lane change, both drivers need to communicate their intention in order to achieve a common goal. However, an interview study by Benmimoun, Neunzig, and Maag (2004) shows that 30% of all cooperative offerings are prone to misunderstandings. An example of such a misunderstanding during a lane change would be that the driver on the target lane wants to create a gap for the merging driver but this driver aims to merge on the target lane after the car passed.

Clear communication would improve the interaction in this manoeuvre. However, as mentioned earlier the only obvious means of communication in current vehicles is the turn signal. The use of this signal during a lane change is required by the StVO but the exact usage can differ between individuals. Salvucci and Liu (2002) found that drivers can be divided in two groups concerning the use of the turn indicator during lane change manoeuvre. The first group uses the indicator to communicate...
improving cooperation in traffic

their intention to change the lane and wait for a gap. The other group uses the indicator to inform the traffic about an immediate lane change. Haar, Kleen, Schmettow, and Verwey (2018) also found that participants in a driving simulator differed widely in their reaction to a car communicating a lane change by the turn indicator. These findings illustrate that the two-dimensionality of the turn indicator (on/off & right/left) does not always communicate sufficient information for an efficient and comfortable interaction between drivers.

We therefore propose to use the technological advances to enhance the clarity of intention in a lane change manoeuvre. In order to reduce the ambiguity of the turn signal during lane-changes we propose to question the underlying “semantics”. The term semantics describes the underlying meaning of varying concepts. Today the turn indicator signifies that a driver wants to change its lane, and other drivers cannot distinguish between the two usages Salvucci and Liu (2002) described. We propose to separate the intention and execution of the lane change manoeuvre. In the case of the turn indicator this could be realized by different patterns. A first pattern would signal that the driver plans to change its lane and under circumstances request that a gap is opened by other drivers. A second pattern would indicate the immediate start of the lane change manoeuvre. Those two signals do not necessarily follow each other. Some drivers tend to communicate their lane change in the same second as they start the manoeuvre. In this case, only the execution signal would be provided. In the following, we will call this separation of information enhanced semantics.

These enhanced semantics are not bound to the turn indicator. To further enhance the clarity of other driver’s intentions, these intentions can also be provided to the driver by using the internal head-up display (HUD). This would render the driver more independent of the technological equipment of others. Information about the planning and the execution of a manoeuvre can be provided by the own car, too, but this does require an automatic manoeuvre recognition or communication by V2V.

In order to evaluate the effectiveness of these enhanced semantics and the use of a HUD to communicate the intentions of others during a lane change manoeuvre we executed a simulator study. In this study, we simulated a prototypical turn indicator and a HUD system. These systems are illustrated in Table 1. The modified turn signal follows the earlier mentioned recommendations and is composed of two different patterns. The planning of a lane change manoeuvre is communicated by a swiping animation in the direction of the planned lane change. The start of the manoeuvre is communicated by a flashing animation of the whole indicator.

Analog to this the HUD system was designed. During the planning of a manoeuvre, a flashing arrow is projected on the street next to the car that plans the lane change manoeuvre. In the second phase of the manoeuvre, this arrow moves from right to left and creates a plane equivalent to the needed space of the car on the target lane. In both cases, the lane-changing car uses the normal turn indicator.

Providing this more clear and detailed information about the intention of others to drivers will influence the general perception of the situation itself. This changed perception will also influence the permissive behaviour of drivers. In the case that
drivers will have more information about the detailed intention of others they will be showing behaviour that is more permissive. In order to evaluate these effects, we executed a driving simulator study.

*Table 1 - Overview showing the prototypical HMI concepts used in the study*

<table>
<thead>
<tr>
<th>Mean of communication</th>
<th>Manoeuvre</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turn Indicator</td>
<td>Signal 1: Planning a manoeuvre</td>
<td><img src="signal1.png" alt="image" /></td>
</tr>
<tr>
<td></td>
<td>Signal 2: Starting a manoeuvre</td>
<td><img src="signal2.png" alt="image" /></td>
</tr>
<tr>
<td>Head-up Display</td>
<td>Signal 1: Planning a manoeuvre</td>
<td><img src="head-up1.png" alt="image" /></td>
</tr>
<tr>
<td></td>
<td>Signal 2: Starting a manoeuvre</td>
<td><img src="head-up2.png" alt="image" /></td>
</tr>
</tbody>
</table>

*Method*

**Participants**

In our study, 51 individuals form the Volkswagen Participant pool participated. All of them signed an informed consent and were informed about the study. One participant showed simulator sickness during the experiment and was therefore excluded from the analysis. Of those remaining 50 participants (32 male, 18 female) the average age was 33.4 years (σ = 10.2 years) and the average annual driven distance was 17,012 km (σ = 9730 km).

**Experimental Design**

We used a 2x2 within-subject design with semantics and HUD as independent variables (Table 2). All participants drove scenarios in a randomized order.
Table 2 - Overview over the four conditions that each driver experienced in randomized order

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Semantics</th>
<th>HUD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline (regular turn signal)</td>
<td>Old semantics</td>
<td>Disabled</td>
</tr>
<tr>
<td>Enhanced turn signal</td>
<td>Enhanced semantics</td>
<td>Disabled</td>
</tr>
<tr>
<td>HUD</td>
<td>Old semantics</td>
<td>Enabled</td>
</tr>
<tr>
<td>Enhanced HUD</td>
<td>Enhanced semantics</td>
<td>Enabled</td>
</tr>
</tbody>
</table>

In order to evaluate the intuitive comprehensibility of these systems the participants were not instructed about the detailed functionality of the HMIs. In each scenario, the participants experienced three lane change situations of another car. This gave them the possibility to learn the function and logic behind each system.

Scenario’s

In each scenario, the participant drove on a highway with two lanes with a target speed of 130 km/h. During the scenario, the participants would experience lane change attempts. They had to decide whether they would insist on their right of way or allow the other car to change its lane. In order to allow the lane change, the participants needed to decelerate and create a gap for the merging driver. This behaviour triggered the lane change of the other car and in both “enhanced” conditions, the executions signal. In total, the participant experienced five encounters in which the participant would see a slower car on the right lane. In three situations the other car would engage a lane change to change to the participant’s lane. In the other two situations the car did not attempt a lane change (Figure 1).

![Figure 1. Scenario used in the experiment. The dark arrows represent the target situations in which a car attempts a lane change manoeuvre. The grey block represents situations where a slower car was seen, but it did not attempt to change lanes.](image)

Materials

A fixed-base driving simulator at the research facilities of the Volkswagen Aktiengesellschaft was used. The simulator was equipped with three projectors for a 180° front view. The interior mock-up, which was used to simulate a realistic cockpit, was equipped with side-mirrors and a rear-mirror aiming on three LCD screens. The steering of the simulation was realized through original manufacture steering wheels and pedals, build into the interior mock-up (Figure 2).
Figure 2. Driving simulator of the Volkswagen Aktiengesellschaft used in the study. In the study three projectors were used to create a 180° front view.

Procedure

At start of the experiment, the participant was asked to fill in a demographic questionnaire. After this, the participants received the opportunity to familiarize themselves with the simulator in a training scenario. Afterwards, the participant completed all four scenarios in randomized order. Each scenario took approximately five minutes and additional five minutes for the participant to fill in a questionnaire regarding the experienced scenario and the experienced HMI. Afterwards, additional background information regarding the experiment and a small gift as compensation were provided to the participant.

Data analysis

In the present study the primary outcome variable was whether the participant did not insist on his/her right of way and allowed the other driver to merge by decelerating and creating a gap. The data were analysed by means of a Bayesian multi-level regression model (Gelman & Hill, 2006). The regression model includes the independent variables semantics and HUD as predictors. In order to explore the interplay between these factors, we also added the interaction effect. By virtue of the within-subjects design these effects could be estimated simultaneously on the participant and the population level. The population level effect represents the average response across participants. The participant-level effects were summarized by their degree of variation (standard deviation). A population level effect can be called homogenous, when the underlying participant level effects show little variation relative to the effect itself.

Due to the dichotomous nature of the behavioural response (permissive behaviour/no permissive behaviour), a logistic regression model was estimated with the predictors just described. Note that in logistic regression the coefficients are no longer on the scale of the outcome variable, but on a log-odds scale. As a consequence, these coefficients cannot be interpreted as additive changes per unit. Instead, when a coefficient takes the value $\beta$, then $\exp(\beta)$ is an odd and can be expressed as “the odds for permissive behaviour increased by $\exp(\beta)$”.

In addition to the dichotomous variable of cooperation we used the ratings resulting from a questionnaire developed by Zimmermann, Fahrmeier, and Bengler (2015). This questionnaire asks the participant to evaluate the experienced cooperative situations on different scales. Hereby the questionnaire differentiates between perception of the situation, the own cooperative behaviour and the cooperative behaviour of the partner. These items were answered by the participants on a 1 to 15 scale. We assume that the error will normally distributed around the mean of this scale and therefore modelled these ratings with the same model mentioned above but using linear regression.

The authors decided to refrain from classical *null hypothesis significance testing* due to the serious concerns that are connected to this statistical procedure (Gelman & Loken, 2014; Wasserstein & Lazar, 2016). Therefore, we decided to use Bayesian parameter estimation approach rather than hypothesis testing (Kruschke & Liddell, 2018). Consequently, no p-values will be reported for any treatment effects, but certainty statements are being derived from the marginal posterior distribution as 95% credibility intervals. When this interval does not include zero, this is interpreted as sufficient certainty that there is an impact. The model was estimated using the package *brms* (Bürckner, 2016, v1.5.1) for the R statistical computing environment (R Core Team, 2017, v3.4.2).

**Results**

**Permissive behaviour**

Our first research question regarded the change in permissive behaviour of the participant. Figure 3 illustrates the proportion of manoeuvres in which the merging car successfully changed its lane. Visual inspection of this plot shows a clear increase over the different encounters the participant experienced in one scenario. In the first encounter, around 40% of lane changes were permitted. In the third encounter, around 80% of the lane changes were permitted, without taking differences between the manipulations of the HMI into account. However, using a HUD to enhance the clarity of the situation, the proportion of successful lane changes, are 10% higher in comparison to not using the HUD (illustrated by the upward trend of each line). Using the enhanced semantics does not have an effect on the behaviour of the participants (both lines, red and blue, seem to not differ from each other).
These visual inspections are backed by the coefficients estimated by the logistic regression (Table 3). The odds of a successful lane change in the first encounter only using the classical turn indicator is \( \exp(-1.62 + 1.15) = .63 \). This translates to for every three unsuccessful lane changes, there were roughly two successful lane changes. The positive and rather large coefficients related to encounter, illustrates the clear upward trend observed in Figure 3. Inspecting the credibility effects of each estimate we can conclude that the effects are estimated with considerable high certainty (zero is not included in the intervals). However, the intercept, the effect of the classical turn indicator, varies strongly between individual participants (illustrated by the large standard deviation \( \sigma \)).

The positive effect of the HUD found in Figure 3 can also be found in the coefficients estimated by the model. The usage of a HUD increases the proportion of permissive behaviour by the factor of \( \exp(.76) = 2.14 \) and is estimated with rather high certainty. In contrast, the usage of enhanced semantics did practically have no effect (\( \exp (-.02) = .98 \)) on the proportion of successful lane changes. Despite the variation between participants observed with the classical turn indicator, the effects related to the HUD, the enhanced semantics and the encounter are rather homogenous between participants, illustrated by the smaller standard deviation.

Next to the adjustment of the actual behaviour, the subjective perception of the situation was evaluated by a questionnaire after each scenario. The questionnaire used originates from the study executed by Zimmermann et al. (2015). This questionnaire measures the perceived own cooperativeness, the cooperativeness of the partner car, the cooperativeness of the situation in general and the impact of the situation on the general traffic.
Table 3. The coefficients table of the model predicting how often the participant allowed a lane change on a logistic scale. The values in brackets are the odds \( \exp(\log(\text{odds})) \).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>log(odds)</td>
<td>-1.62 (-.20)</td>
<td>.76 (2.14)</td>
<td>-.02 (.98)</td>
<td>1.15 (3.16)</td>
<td>.15 (1.16)</td>
</tr>
<tr>
<td>Lower 2.5%</td>
<td>-2.48 (.08)</td>
<td>.11 (1.12)</td>
<td>-.63 (.53)</td>
<td>.82 (2.27)</td>
<td>-.67 (.51)</td>
</tr>
<tr>
<td>Upper 2.5%</td>
<td>-0.82 (.44)</td>
<td>1.42 (4.14)</td>
<td>.59 (1.80)</td>
<td>1.53 (4.62)</td>
<td>1.02 (2.77)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>1.39</td>
<td>.52</td>
<td>.45</td>
<td>.41</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 4. Subjective ratings regarding the own cooperativeness, cooperatives of the other car, cooperativeness of the situation and the impact of the situation on the general traffic. Higher ratings represent more cooperativeness.
Visual inspection of Figure 4 suggests the same trend in the subjective ratings as observed in the permissive behaviour. However, no learning effect over the encounters could be measured because the subjective measurements were taken after the participant finished the complete driving scenario. The positive effect of the HUD can be observed in most of the ratings. Nevertheless, the effect of the HUD seems to be smallest regarding the own cooperativeness and largest regarding the cooperativeness of the partner car. The enhanced semantics do not seem to have an effect on the perceived cooperativeness on any scale.

All those ratings were also modelled using a multi-level regression. The results of this regression will be reported below regarding each scale of the questionnaire.

**Own cooperativeness**

First we will inspect the regression results regarding the own cooperativeness. The coefficients resulting from the multi-level regression (Table 4) support the visual inspection. Using a HUD in order to communicate the lane change manoeuvre does increase the subjective impression of the participants own cooperativeness \((\text{HUD} = .75 \ [\text{CI} = .22 : .90])\). However, the enhanced semantics did not influence this rating with confident certainty \((\text{Enhanced Semantics} = .24 \ [\text{CI} = -.10 : .55])\). In addition no credible interaction effect \((\text{HUD:Enhanced Semantics} = -.26 \ [\text{CI} = -.71 : .20])\) or noteworthy inter-participant variation could be observed.

<table>
<thead>
<tr>
<th></th>
<th>Population Effects</th>
<th>Participant-level Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Lower 2.5%</td>
</tr>
<tr>
<td>Intercept [No HUD / Current semantics]</td>
<td>4.93</td>
<td>4.65</td>
</tr>
<tr>
<td>HUD</td>
<td>.57</td>
<td>.22</td>
</tr>
<tr>
<td>Enhanced semantics</td>
<td>.24</td>
<td>-.10</td>
</tr>
<tr>
<td>HUD:Enhanced semantics</td>
<td>-.26</td>
<td>-.71</td>
</tr>
</tbody>
</table>

**Cooperativeness of merging driver**

In the following section, the regression results regarding the cooperativeness of the merging driver are described. The coefficients resulting from the multi-level regression (Table 5) also support the visual inspection and are comparable to the result regarding the assessment of the personal cooperativeness. The HUD communicating the lane change of the merging driver increases the subjective cooperativeness of this driver \((\text{HUD} = .66 \ [\text{CI} = .20 : 1.09])\). However, the enhanced semantics did not influence this rating with confident certainty or practically noticeable effect \((\text{Enhanced Semantics} = -.06 \ [\text{CI} = -.48 : .34])\). In
addition no credible interaction effect \( (HUD:Enhanced\ Semantics = -0.23 [CI = -0.32 : .80]) \) or noteworthy inter-participant variation could be observed.

Table 5. Regression coefficients resulting from the linear multi-level regression regarding the perceived cooperativeness of the merging driver.

<table>
<thead>
<tr>
<th>Population Effects</th>
<th>Participant-level Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate Lower 2.5% Upper 2.5%</td>
</tr>
<tr>
<td>Intercept [No HUD / Current semantics]</td>
<td>3.65 3.26 4.02</td>
</tr>
<tr>
<td>HUD</td>
<td>.66 .20 1.09</td>
</tr>
<tr>
<td>Enhanced semantics</td>
<td>-0.06 -0.48 .34</td>
</tr>
<tr>
<td>HUD:Enhanced semantics</td>
<td>-0.23 -0.32 .80</td>
</tr>
</tbody>
</table>

Cooperativeness of the situation

In addition to the personal cooperativeness, the participants rated the cooperativeness of the situation in general. The coefficients resulting from the multi-level regression modelling this rating (Table 6) also support the visual inspection made earlier. The HUD communicating the lane change of the merging driver increases the subjective cooperativeness of the situation \( (HUD = 0.96 [CI = 0.44 : 1.47]) \). As before, the enhanced semantics did not influence this rating with confident certainty \( (Enhanced\ Semantics = 0.42 [CI = -0.04 : 0.87]) \). However, the majority of the credibility interval regarding the enhanced semantics is positive. Therefore, we could conclude a tendency towards a positive effect. No credible interaction effect \( (HUD:Enhanced\ Semantics = -0.13 [CI = -0.78 : 0.53]) \) or noteworthy inter-participant variation could be observed.

Table 6. Regression coefficients resulting from the linear multi-level regression regarding the perceived cooperativeness of the lane change situation.

<table>
<thead>
<tr>
<th>Population Effects</th>
<th>Participant-level Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate Lower 2.5% Upper 2.5%</td>
</tr>
<tr>
<td>Intercept [No HUD / Current semantics]</td>
<td>4.16 3.76 4.57</td>
</tr>
<tr>
<td>HUD</td>
<td>.96 .44 1.47</td>
</tr>
<tr>
<td>Enhanced semantics</td>
<td>.42 -.04 .87</td>
</tr>
<tr>
<td>HUD:Enhanced semantics</td>
<td>-.13 -.78 .53</td>
</tr>
</tbody>
</table>
Impact of the situation on general traffic

As last scale, the cooperative questionnaire did include the subjective rating of the impact of the situation on the general traffic. This rating was also modelled using the same multi-level linear regression. The coefficients resulting from the model can be found in Table 7. The results paint the same picture as the visual inspection of Figure 4 and the results reported above. Using the HUD to communicate the lane change increases the impact rating, which indicates a more positive impact ($HUD = 1.00 \ [CI = .23 : 1.46]$). Again, the enhanced semantics did not influence the rating with confident certainty or noticeable effect size ($Enhanced\ Semantics = .14 \ [CI = -.27 : .53]$). There is also no credible interaction effect ($HUD:Enhanced\ Semantics = .03 \ [CI = -.55 : .61]$) or noteworthy inter-participant variation observable.

Summarizing all the results, we can say that providing additional information with a HUD system does have a credible and practical relevant effect. In contrast, the enhanced semantics did not have a credible and relevant effect on the depended variables. In addition to this, we could observe a large variation between participants regarding the basic turn indicator and a learning effect indented of HUD or enhanced semantics.

Table 7. Regression coefficients resulting from the linear multi-level regression regarding the perceived impact on the general traffic.

<table>
<thead>
<tr>
<th></th>
<th>Population Effects</th>
<th>Participant-level Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Lower 2.5%</td>
</tr>
<tr>
<td>Intercept [No HUD / Current semantics]</td>
<td>3.74</td>
<td>3.38</td>
</tr>
<tr>
<td>HUD</td>
<td>1.00</td>
<td>.56</td>
</tr>
<tr>
<td>Enhanced semantics</td>
<td>.14</td>
<td>-.27</td>
</tr>
<tr>
<td>HUD:Enhanced semantics</td>
<td>-.03</td>
<td>-.55</td>
</tr>
</tbody>
</table>

Discussion

The goal of this experiment was to investigate if better communication between drivers would lead to behaviour that is more permissive and increase the subjective cooperativeness of the situation. Therefore, we tried to enhance this communication by using a HUD system and applying additional semantics to the message communicated in a lane change manoeuvre.

Regarding the new technologies, we could also observe very clear results. The first would be that the enhanced semantics we used did not have any effect on the permissive behaviour observed or the subjective experience of the participants. One
improving cooperation in traffic

explanation could be that the implementations of these semantics did not have an intuitive meaning for the participants. On purpose we did not explain the different semantics prior to the experiment to investigate if participants would get the difference intuitively but this was obviously not the case. All participants were Volkswagen employees from different divisions, with wide ranging variety in backgrounds and interest in new technologies used in vehicles. However, even this sample close to automotive technologies did not have any advantages in interpreting the new semantics used in the study. Therefore, it would be of great interest if the enhanced semantics have a positive effect if they are explained to the participants prior to the experiment.

In contrast to this, we could observe clear positive effect of the HUD concepts. Both, the permissive behaviour and the subjective assessment of the situations, were positively influenced if the intention of the lane change was not only communicated by the turn indicator but additionally by the HUD system. In contrast to the enhanced semantics, this function could be understood intuitively, because this positive effect was observable from the very first encounter and stayed constant over the duration of the situation. This supports the assumption that adding information to the limited capacity of the turn indicator can enhance the communication between drivers and does also increase permissive behaviour of drivers. Even though the information in general stays the same, the more salient presentation and the location in the field of view, generated by the HUD, seem to address the driver more direct.

A possible explanation could be, that the driver feels more responsible, if the information is provided by the HUD of his own car and not only by the turn indicator of another car.

Interestingly we could also observe a large variation between participants in reaction to the traditional turn indicator, especially in their behaviour. This supports the theory of Salvucci and Liu (2002) that there are different types of usage of the turn indicator and reproduces the observation of Haar et al. (2018).

Another interesting observation was the clear learning effect in permissive behaviour with each encounter the participant had in every situation. This effect was independent of the technology used. However, the learning effect was independent from the order in which the participants experienced the different technologies and the technologies itself. One could think that participants could learn the positions of the lane changes in the different scenarios. In that case, we should be unable to observe a positive trend in the following scenarios. However, this was not the case. Therefore, the general upward trend in permissive behaviour needs to have a different explanation than learning the position of the behaviour or the different technologies.

However, there are also some limitations to the findings of this experiment. The experiment was executed in a driving simulator where real participants interacted with scripted behaviour of computer-controlled drivers. As mentioned before, interactions in traffic are complex situations with a lot of influencing. Even though participants reported that the car controlled by the simulation did behave realistically, natural traffic brings more variation in the behaviour of drivers and
therefore each situation is different. Therefore, the effect of the used technologies should be tested in a more realistic driving environment.

Conclusion

The study described here illustrates that cooperative lane change maneuverers can be enhanced by providing information to the driver from additional sources than only the classical turn indicator. However, we could not show that rephrasing the meaning of the traditional turn signal by using enhanced semantics can be interpreted intuitively. It could be indicated that the information provided by the traditional turn signal maybe not sufficient to communicate all information that a driver needs. Therefore, this study can be a starting point to rethink the traditional technologies used to communicate in traffic and develop solutions that aid the driver in future traffic.

References


Short-term effects of unnecessary collision alarms on driver behaviour

Christina Kaß1, Gerald Schmidt2, & Wilfried Kunde3
1Opel Automobile GmbH, 2Furtwangen University, 3Julius-Maximilians-University Würzburg
Germany

Abstract

Prior research has focused on the effects of classical false alarms on drivers’ compliance (e.g. cry wolf effect). The present research investigated short-term effects of unnecessary collision alarms on drivers’ braking responses in two driving simulator studies. In both experiments, a hard braking lead vehicle caused time to collision values that fell below usual thresholds for collision alarm activation. However, collision alarms were eventually unnecessary, providing that the intended driving manoeuvre of the ego driver or that of the lead vehicle were taken into account. In the first experiment, the braking lead vehicle was irrelevant for drivers’ own planned trajectory. In the second experiment, drivers were able to anticipate that the lead vehicle braked to turn out of the lane within a short period of time. Whereas half the participants received collision alarms, the other half did not. Overall, results revealed that drivers who received unnecessary alarms braked more often and decelerated more strongly than drivers who showed their natural driving behaviour without receiving alarms. These findings underline the need to reduce the rate of unnecessary alarms in order to avoid driver overreactions as potentially critical short-term effects.

Introduction

Forward collision alarms (FCA) signal drivers to take immediate evasive action to avoid an external hazard (International Organization for Standardization, 2016). The extent to which drivers act according the alarm signal and take evasive action is referred to as compliance (Meyer, 2004). Prior research has repeatedly shown that FCAs have a positive impact on driver performance in imminent collision situations (Kusano & Gabler, 2012; Lee, McGehee, Brown, & Reyes, 2002; Maltz & Shinar, 2004). However, it does not necessarily reflect safe behaviour to unconditionally respond to warnings (Green & Swets, 1966). The most appropriate alarm response is highly dependent on the nature of the warning, particularly whether it is correct, false, or unnecessary. Advanced Driver Assistance Systems (ADAS) frequently activate unnecessary alarms (Flannagan et al., 2016; General Motors Corporation, 2005). Unnecessary alarms are associated with situations judged as hazardous by the implemented algorithm, but not subjectively by the driver (Lees & Lee, 2007). In
contrast to false alarms that are activated by sensor malfunctions, unnecessary alarms are associated with an apparent trigger, such as another vehicle. However, the alarm trigger does not actually constitute a collision risk. Instead of taking an evasive action in response to false and unnecessary alarms (commission errors), drivers need to ignore the warning and refrain from executing strong braking responses (Meyer, 2004). Very little is currently known about the effects of unnecessary collision alarms on driver behaviour. Therefore, the two presented driving simulator studies aimed to elucidate how unnecessary alarms influence driver behaviour compared to drivers’ natural driving behaviour in the same traffic situations.

**Effects of false alarms on driving behaviour**

The cry wolf effect describes the detrimental impact of frequently experienced false alarms on drivers’ compliance with correct alarms (Breznitz, 1984). Low compliance is manifested through longer reaction times (e.g. Sullivan, Tsimhoni, & Bogard, 2008; Yamada & Kuchar, 2006), decreased reaction-rates (Bliss & Acton, 2003), and reduced reaction intensity to correct alarms (Lees & Lee, 2007). However, the present research focuses on drivers’ response selection when receiving unnecessary alarms. Prior research on operators’ responses to classical false alarms revealed their general ability to behaviourally discriminate between false alarms and correct alarms. Compared to correct alarms, responses to false alarms are slower (Abe, Itoh, & Tanaka, 2002; Maltz & Shinar, 2004), less frequent, and less intense (Lees & Lee, 2007). These effects were already present after short-term exposure to the warning systems. However, the number of commission errors in response to false alarms as a measure of decision-making accuracy is associated with the false alarm rate (Cummings, Kilgore, Wang, Tijerina, & Kochhar, 2007; Maltz & Shinar, 2004).

**Effects of unnecessary alarms on driving behaviour**

Previous research revealed contradictory findings concerning drivers’ responses to unnecessary alarms. Results of a driving simulator study on short-term effects of unnecessary alarms showed that these alarms caused drivers to brake more often and more intensively than drivers who did not receive alarms in the same traffic situations (Lees & Lee, 2007). In a study by Zarife (2014), drivers braked in 66% of the time in response to a (generic) unnecessary collision alarm. The study lacks a comparison to driver behaviour in the same situations without unnecessary alarm. In both studies, unnecessary alarms were caused by other road users whose behaviour was neither predictable for the system nor for the driver. For example, a cyclist was approaching the road from the side and finally turned and drove along the sidewalk before it could have crossed the driver’s way.

There are two field studies that provide insights into driver responses to unnecessary alarms for long-term system exposure (Flannagan et al., 2016; General Motors Corporation, 2005). These studies revealed that 30 to 50% of all issued FCAs could roughly be categorized as unnecessary (Lees & Lee, 2007). In these situations, a stopped or braking lead vehicle triggered an FCA according to the implemented
algorithm (grey vehicle in Figure 1). However, the ego and the lead vehicle did not remain in the same lane within four seconds after the FCA. Therefore, the ego driver (green vehicle in Figure 1) might have perceived the situation as not hazardous either because...

1. the ego driver intentionally approached the lead vehicle as a prelude to change lanes or to turn (green arrow in Figure 1) or
2. the ego driver approached the lead vehicle as he or she was able to predict the other’s intention to change lanes or to turn out of the lane in the near future (grey arrow in Figure 1).

Accordingly, the ego driver could predict a dissolving outcome of the inter-vehicle conflict and, thus, did not or only minimally respond to FCAs in these constellations. Drivers responded with braking in only 19 to 30% of the time with average decelerations of 0.5 m/s². Importantly, the absence of driver responses did not result in collisions.

![Diagram](image)

*Figure 1. Illustration of two possible reasons for activations of unnecessary collision alarms identified in the field studies. Ego vehicle = green; lead vehicle = grey.*

Findings of the cited driving simulator and field studies might diverge for two reasons. First, drivers’ ability to predict a dissolving outcome of a potential inter-vehicle conflict might influence braking responses to unnecessary alarms. In the simulator studies, drivers could not predict the behaviour of the other road users (Lees & Lee, 2007; Zarife, 2014). Presumably, drivers were able to predict their own subsequent actions and those of other road users in the field studies (see Figure 1, Flannagan et al., 2016; General Motors Corporation, 2005). Second, long-term system experience might change the way drivers respond to unnecessary alarms. In the short term, the process of validating an alarm might be more demanding for unnecessary than for false alarms. In contrast to false alarms, drivers are able to understand what triggered the unnecessary alarm (Lees & Lee, 2007). Therefore, they need to additionally comprehend and anticipate if the alarm trigger might constitute a hazard within the next few seconds or not. However, as the number of experienced unnecessary alarm situations increases, drivers might develop an understanding of typical traffic constellations that may result in unnecessary alarms. As a consequence, the process of cross-checking the validity of unnecessary alarms might then require less cognitive resources. In the field study by Flannagan et al. (2016), alarm rates decreased over time in the scenarios where the ego and the lead vehicle did not remain in the same lane after the FCA. The authors argued that drivers were able to anticipate these scenarios and adjusted their behaviour to avoid
setting of an alarm. This finding supports the assumption that system experience might influence alarm responses.

Research goals

The goal of this research was to gain insights into drivers’ responses to unnecessary alarms. Two driving simulator studies aimed to reveal whether the superfluous braking responses to unnecessary alarms reported by Lees and Lee (2007) and Zarife (2014) were primarily caused by limited experience with unnecessary collision alarms or by a lack of ability to predict a dissolving outcome of inter-vehicle conflicts. Therefore, it was focused on short-term driver responses to unnecessary alarms issued in traffic situations with predictable dissolving conflict.

In the present research, a hard braking lead vehicle triggered an FCA. In the first experiment, ego drivers had the manoeuvre intention to turn directly behind the braking lead vehicle (following green arrow in Figure 1). In the second experiment, ego drivers were able to predict that the lead vehicle brakes to turn out of their lane within a short period of time (following grey arrow in Figure 1). Therefore, the potential inter-vehicle conflict dissolved and collision alarms were unnecessary. Drivers’ braking responses (frequency of braking responses and magnitude of speed reductions) when receiving unnecessary alarms were compared to those of drivers who experienced the same situation without FCA.

Experiment 1

Method

Participants
A total of 42 participants (12 females) took part in the study. They were employees of a German car manufacturer (Opel Automobile GmbH) and were recruited from the company’s own participant’s test panel. The mean age was $M = 42.07$ ($SD = 11.23$). On average, participants held their driver’s licence for $M = 23.31$ years ($SD = 11.07$). Only eight participants had prior experience with FCA systems, $M = 3.5$, $SD = 1.07$ on a 5-point scale (1 = very little, 5 = very much experience).

Apparatus
The study was conducted in the fixed-base driving simulator of Opel Automobile GmbH (mock-up: Opel Insignia). The simulator provided a 130° frontal field of view with three 70” thin film transistor (TFT) screens and a full rear view offered by an additional 70” TFT screen visible through a conventional rear view mirror and by two 7” TFT screens as side mirrors. The driving simulation software was SILAB 6.

Experimental design and dependent measures
The experiment used a 2x2 mixed design with manoeuvre intention as within-subject factor and level of assistance as between-subject factor. The initial situation was the same for each test event. Participants followed the lead vehicle at 50 km/h. The lead vehicle suddenly braked to a standstill. Time to collision (TTC) values
between ego and lead vehicle always fell below 1.9 seconds. To manipulate manoeuvre intention, a navigation system provided auditory and visual announcements. Participants either intended (a) to stay or (b) to turn behind the braking lead vehicle (see Figure 2). In the latter case, the braking lead vehicle could not cause a collision and the conflict dissolved. Therefore, alarms in these situations were considered as unnecessary. The factor level of assistance was represented by either (a) receiving an FCA during each test event or (b) never receiving FCAs. FCAs were triggered by TTC values below 1.9 seconds and consisted of an urgent auditory warning tone (5 x 100 ms on/off, 2,000 Hz) and a flashing red LED segment at the bottom of the windshield (5 x flashing 100 ms). To obtain groups of equal size for both groups, participants were quasi randomly allocated to the two groups (with alarms n = 21; without alarms n = 21). This means that the number of participants randomly assigned to each level of assistance was limited.

![Figure 2](image)

Figure 2. Realization of manoeuvre intention in both test scenarios. Braking lead vehicle is marked by a frame. Solid line: intention to stay behind the lead vehicle; dashed line: intention to turn behind the lead vehicle.

Participants experienced test scenarios in two different environments (turn right and turn left; see Figure 2). To reduce learning effects, filler scenarios were implemented in between the test scenarios. Environments did not differ from those in the test events, but the lead vehicle continued driving and did not decelerate at a certain point. All participants encountered four test scenarios and four filler scenarios that varied according to manoeuvre intention and traffic environment. To control for transition effects, the sequence of the eight events was permuted to four different sequences. Subjects were randomly distributed to the four sequences.

As dependent measures, the frequency of braking responses and the magnitude of speed reduction in response to the braking lead vehicle (and the FCA) were assessed. Brake pedal positions of 10% and more were counted as intentional braking response (Gold, Damböck, Lorenz, & Bengler, 2013; Zeeb, Buchner, & Schrauf, 2016). The magnitude of speed reduction was calculated as a result of the speed participants drove when the lead vehicle started braking minus their lowest
speed before they re-accelerated (in km/h). In case drivers did not decelerate at all, their speed at a pre-defined position in the test scenario was used as subtrahend.

Procedure
Upon arrival, participants received written instructions about the purpose and the procedure of the study and signed a consent form that informed them about their right to decline to participate and to withdraw from the study at any point and the method of data anonymization. The information given in the consent form covered Articles 13 and 15 – 21 of the EU General Data Protection Regulation.

After participants drove a five-minute practice drive, they started with the experimental drive. Overall, the experiment took about one hour per participant. At the end of the experiment, participants were thoroughly debriefed.

Data Analysis
Data analyses were carried out using IBM SPSS statistics software (Version 22). The impact of both independent variables on magnitude of speed reduction was analysed with a mixed analyses of variance (ANOVA). The influence of level of assistance on frequency of braking responses in situations with dissolving outcome of the conflict (= manoeuvre intention to turn) was calculated with a chi square test. The significance level was set at $\alpha = 0.05$. Magnitude of speed reduction values reflect the average between both test scenarios turn right and left.

Results
When the conflict dissolved due to the intention to turn behind the braking lead vehicle, chi-square tests for both test scenarios revealed a significant association between the level of assistance and the frequency of braking responses, $\chi^2(1) = 15.43, p < .001$ (Scenario 1), $\chi^2(1) = 18.71, p < .001$ (Scenario 2). The number of drivers who braked in response to the braking lead vehicle was higher in the group with unnecessary FCAs than in the group without FCAs (see Figure 3).

![Graph showing number of braking responses for dissolving conflict (turn behind lead vehicle) grouped by test scenario and level of assistance.](image)

Figure 3. Number of braking responses for dissolving conflict (turn behind lead vehicle) grouped by test scenario and level of assistance.

A 2x2 mixed ANOVA revealed a significant main effect of manoeuvre intention on magnitude of speed reduction, $F(1, 40) = 525.53, p < .001, \eta^2 = .93$. The magnitude
of speed reduction was significantly higher when drivers intended to stay than to turn behind the braking lead vehicle. Drivers reduced their speed to a significantly higher extent with FCAs than without, $F (1, 40) = 35.63, p < .001, \eta_p^2 = .47$. There was a significant interaction between manoeuvre intention and level of assistance, $F (1, 40) = 28.22, p < .001, \eta_p^2 = .41$. The impact of level of assistance on magnitude of speed reduction was only present when drivers intended to turn behind the braking lead vehicle, $t (40) = -6.15, p < .001$. Figure 4 displays the corresponding means and 95% confidence intervals.

![Figure 4. Magnitude of speed reduction in Experiment 1 grouped by manoeuvre intention and level of assistance. Error bars reflect 95% CI.](image)

**Discussion**

This experiment investigated short-term effects of unnecessary alarms on driving behaviour. Alarms were considered as unnecessary when the inter-vehicle conflict dissolved as the ego driver intended to turn directly behind the braking lead vehicle and as useful when drivers planned to stay behind it. Results have shown that drivers who received unnecessary collision alarms braked more frequently and reduced their speed to a higher extent than drivers who experienced the same situation without alarms. When staying behind the braking lead vehicle, drivers of both groups with and without FCAs reduced their speed to an equal extent. The results of the present study are in line with the findings by Lees and Lee (2007) and Zarife (2014) and contradict findings by Flannagan et al. (2016) and General Motors Corporation (2005). Most drivers who did not receive FCAs refrained from braking in situations with dissolving conflict, indicating that the anticipation of a safe outcome was principally possible. Accordingly, predictability did not seem to be the key factor that helped drivers to select an appropriate alarm response, particularly to suppress a braking response. In the short term, drivers might overreact to unnecessary FCAs (commission errors) because the process of cross-checking the validity of the alarm might be too demanding in the first phase. As already suggested by Flannagan et al. (2016) and General Motors Corporation (2005), drivers might reduce or suppress braking responses to unnecessary alarms with increasing experience.
Experiment 2

Method

Participants
Forty drivers (20 females) were recruited from the Wuerzburg Institute for Traffic Sciences GmbH (WIVW) driver test panel (mean age = 42.60 years, SD = 10.77). As experienced drivers have advantages in hazard detection and prediction compared to novice drivers (Crundall, 2016; Smith, Horswill, Chambers, & Wetton, 2009), only experienced drivers were recruited who held their drivers’ licence for 10 years or more ($M = 23.83$, $SD = 10.00$) and covered more than 8000 km per year ($M = 20.550$, $SD = 14.408$). Only five participants had prior experience with FCA systems, $M = 4.6$, $SD = 0.89$ on a 5-point scale (1 = very little, 5 = very much experience).

Apparatus
The study was conducted in the fixed-base driving simulator at WIVW GmbH (mock-up: Opel Insignia). The simulator provided a 300° horizontal and 47° vertical field of view. Five image channels seamlessly projected the simulation onto a flat screen. The rear-view was furnished by two LCD displays installed as the rear view mirror and the left side mirror. The driving simulation software was SILAB 6.

Experimental design and dependent measures
The experiment used a 2x2 mixed design with lead vehicle behaviour as within-subject factor and level of assistance as between-subject factor. The initial situation was the same for each test event. Participants followed the lead vehicle at 50 km/h. Just before an intersection, the lead vehicle activated its turn indicator for 400 ms before it suddenly braked to a standstill. TTC values between ego and lead vehicle always fell below two seconds. The braking lead vehicle either (a) stayed in the same lane as the ego driver or (b) turned and left the lane (see Figure 5). Three test scenarios with different environmental cues were realized. A pedestrian on a crosswalk in the right turning street (Scenario 1, see Figure 5), in the left turning street (Scenario 2), or a stopped bus at station at the beginning of the right turning street (Scenario 3) served to indicate that the lead vehicle will stay in the lane and the conflict will remain. In the other condition, there were no obstacles in the right or left turning streets. Thus, the lead vehicle remained stopped for only 150 milliseconds before it started again and turned at the intersection. Participants could have driven on at a constant speed of 50 km/h without causing a collision. Therefore, the conflict dissolved and alarms in these situations were considered as unnecessary. Level of assistance and alarm design were realized similarly to Experiment 1 (with alarms $n = 21$; without alarms $n = 19$).

To reduce learning effects, twelve filler scenarios were implemented in between the test scenarios. Environments were similar to those in the test events, however, the lead vehicle turned without prior deceleration at the intersection. Thus, all participants encountered six test events and twelve filler scenarios throughout the experiment. To control for transition effects, the sequence of the test events was permuted to four different sequences. The filler scenarios were randomly allocated
between the test events. To obtain groups of equal size for each sequence, participants were quasi randomly distributed to the four sequences. This means that the number of participants randomly assigned to each sequence was limited.

Dependent variables were similar to those of Experiment 1.

Procedure and data analysis
Procedure and data analyses were similar to Experiment 1. Magnitude of speed reduction values reflect the average between the three different test scenarios.

![Figure 5. Realization of lead vehicle behaviour in Scenario 1. Environmental cues that served to predict lead vehicle behaviour are marked by a frame: occupied versus free crosswalk.](image)

Results

When the lead vehicle turned and left the lane, there was a marginal significant impact of level of assistance on the frequency of braking responses in Scenario 1, $\chi^2(1) = 3.09, p = 0.079$. However, there were no significant effects for Scenario 2, $\chi^2(1) = 1.09$, n.s. and Scenario 3. In the latter scenario, in both groups with and without FCAs, all participants braked in response to the braking lead vehicle. In sum, the number of drivers who braked in response to the braking lead vehicle even though the conflict dissolved did not differ in the group without and with (unnecessary) FCAs.

A 2x2 mixed ANOVA revealed a significant main effect of lead vehicle behaviour on magnitude of speed reduction, $F(1, 38) = 163.38, p < .001, \eta^2 = .81$. Drivers reduced their speed to a significantly higher extent when the lead vehicle stayed in the lane than when it left the lane. The magnitude of speed reduction was
significantly higher in the group with FCAs than without, $F(1, 38) = 39.9, p < .001, \eta^2_p = .51$. There was a significant interaction between lead vehicle behaviour and level of assistance, $F(1, 38) = 38.34, p < .001, \eta^2_p = .50$. The impact of level of assistance on magnitude of speed reduction was only present when the lead vehicle turned and the conflict dissolved, $t(38) = -6.37, p < .001$. Figure 6 shows the corresponding means and 95% confidence intervals.

Discussion

As in the first experiment, unnecessary alarms were associated with a dissolving inter-vehicle conflict. Here, a lead vehicle braked to turn and cleared the participants’ lane. When the lead vehicle remained in the same lane for some more seconds, the alarm was considered as useful. Results have shown that drivers who did not receive alarms braked as often as drivers who received unnecessary alarms even though the lead vehicle left the lane. However, unnecessary alarms resulted in higher speed reductions compared to those of drivers who experienced the same situation without alarms. When the lead vehicle stayed in the lane, drivers of both groups with and without FCAs reduced their speed to an equal extent. Compared to drivers’ natural driving behaviour unnecessary alarms resulted in stronger driver responses. The conclusions are in line with those of the first experiment.

General discussion

The aim of the present research was to gain insights into short-term driver responses to unnecessary collision alarms in traffic situations with predictable outcome. In two driving simulator studies, drivers’ braking behaviour in response to unnecessary collision alarms was compared to natural driving behaviour in the same situations without FCAs. Unnecessary alarms were issued when TTC values fell below critical thresholds. Thus, present-day FCA systems would have judged these situations as hazardous. However, drivers were able to predict a non-hazardous dissolving outcome either because the ego driver had the intention to turn behind the braking lead vehicle (Experiment 1) or because the ego driver was able to predict that the
lead vehicle braked to turn within the next seconds (Experiment 2). The results of both experiments revealed that drivers who received unnecessary FCAs in dissolving inter-vehicle conflicts reduced their speed to a higher extent than drivers who experienced the same situation without alarms. In the first experiment, unnecessary alarms were associated with the frequency of braking reactions. However, the second experiment showed no differences between the groups with and without FCA with regard to braking reaction rates. Drivers reduced their speed to a higher extent in Experiment 2 than in Experiment 1 when experiencing dissolving inter-vehicle conflicts. These findings may be explained by the fact that drivers could easier predict their own manoeuvre intention (Experiment 1) than the intentions of other road users (Experiment 2). As drivers who did not receive unnecessary alarms in dissolving inter-vehicle conflicts mostly reacted without or with only minimal speed reductions, the outcome of the situations seemed to be predictable for drivers in both experiments. However, the predictability did not seem to help drivers to select an appropriate response to unnecessary alarms. The results of the present studies supported findings by Lees and Lee (2007) and Zarife (2014) that unnecessary alarms cause drivers to respond with moderate to strong braking in the short term. Without prior experience, appropriate response selection for unnecessary alarms seems to be more difficult than for false alarms (e.g. Lees & Lee, 2007). The validation of unnecessary alarms requires the evaluation of the apparent alarm trigger as hazardous or not. This process is assumed to be more demanding compared to the validation of false alarms. With increasing system and alarm experience, drivers might develop an understanding of typical traffic constellations that activate unnecessary alarms. This knowledge might simplify the process of validating an unnecessary alarm and to select an appropriate alarm response. Therefore, drivers’ responsiveness to unnecessary alarms is expected to decrease with increasing system experience as shown in the field studies by Flannagan et al. (2016) and General Motors Corporation (2005).

Future research should examine the process by which driver responses to unnecessary alarms change over time. Therefore, short- and long-term effects of a similar type of unnecessary alarm on driver responses need to be compared. In this context, it is interesting to investigate the cognitive load involved in the validation of unnecessary alarms with varying degrees of system experience. Additionally, there might be a complex interaction between false and unnecessary alarms and their influence on driver behaviour. This topic reflects another question for future research.

The findings of the present research underline the need to reduce the rate of unnecessary alarms in order to avoid driver overreactions as potentially critical short-term effects.

Acknowledgements

We would like to thank Norbert Schneider, Diana Adrian, Felicitas Muth, and Christoph Klöffel for their support during study preparation and execution. Furthermore, we thank Clara Beck and Jakob Kammerer for their support with regards to driving data preparation.
References


Towards an error-tolerant HMI approach to counteract registration errors in augmented reality

Vitalij Sadovitch, Michael Wittkämper, Fares Wallis, & Ina Petermann-Stock
Volkswagen AG, Wolfsburg
Germany

Abstract

An augmented reality head-up display (AR-HUD) enhances the driver’s view by a superimposition of virtual content onto the real environment to support the driving task. A frequently addressed technical issue in augmented reality are deviations within the registration of virtual elements in 3D space. In the context of an AR-HUD these registration errors are leading to a reduction in driving performance and subjective usability assessment. In this contribution, an error-tolerant approach for Human-Machine-Interfaces (HMI) is described to counteract occurring registration errors. The approach was evaluated in a stereoscopic driving simulator with 62 participants performing a driving and two secondary tasks. The subjective perception of augmented reality accuracy was slightly higher, when real objects were virtually marked by extended or diffuse visualizations rather than conventional visualization types with sharp edges. Meanwhile there were no negative effects observed on driving performance and mental workload due to the error-tolerant HMI solutions.

Introduction

Augmented reality (AR) is considered as one of the most recent developments in human computer interaction technology (Billinghurst et al., 2015). It aims to combine reality and virtuality by aligning, respectively registering three-dimensional virtual content with the real environment, allowing interaction in real time (Azuma, 1997). There are three distinguishable displaying techniques to realize visual AR. Optical see-through (OST) superimposes the real environment with virtual content directly in user’s field of view by using semi-transparent mirrors as projection planes. In contrast, the real environment is shown in video see-through (VST) as a digital video feed on a separate display with an AR. These techniques can be realized as a stationary, a handheld or as a head-worn display (HWD), which is most common (Van Krevelen & Poelman, 2010).

Besides the recently growing usage of augmented reality in the field of consumer applications (e.g. smartphone games), Sanna and Manuri (2016) show some examples how AR is already used professionally in several industries (e.g. medicine, architecture, automotive). In the automotive industry engineers can be supported at
the product development process, as well as customers at the driving task. Modern vehicles are often equipped with a head-up display to show virtual information to the driver, directly in line of sight onto the real environment. To this purpose, the semi-transparent windshield is used as an optical see-through technology to reflect a generated image into driver’s eyes, who then perceives a virtual image behind the windshield. Originally, head-up displays were used for screen-fixed (unregistered) content like speed and navigation cues in the form of abstract icons. In contrast several publications show also the opportunity to use a HUD to present world-fixed (registered) augmented reality content in the field of navigation and other assistance systems (Gabbard et al., 2014). An exemplary comparison of a conventional HUD and an augmented reality head-up display (AR-HUD) is shown in figure 1.

![Figure 1. Comparison of a conventional HUD and an AR-HUD showing speed information, a navigation cue and an obstacle warning.](image)

The biggest benefit of head-up displays over traditional head-down displays is that there is no need to take the gaze away from the street to look at displayed information (Gish & Staplin, 1995). This leads to a reduction of reaction times in unexpected traffic situations (Horrey & Wickens, 2004; Kiefer, 1998). Furthermore, correctly superimposed augmented reality content is supposed to reduce mental effort and increase situation awareness, because the interpretation of information becomes easier due to the spatial context (Bengler et al., 2015).

However, an AR-HUD has much higher technical requirements than a conventional HUD. First, the virtual image needs to be placed in further distance to ensure that the driver is able to perceive real and virtual content simultaneously without the need of changing the focus constantly. While the virtual image distance of a conventional HUD is usually at 2 - 3 m, an augmented reality image plane should be placed in a distance of at least 7.5 m to the drivers’ eyes (Jachens et al., 2017). Furthermore, a bigger field of view of the head-up display is usually needed for augmented reality applications. Both requirements lead to a big increase in the installation space of the display unit.

Besides higher requirements for the display technology itself, augmented reality requires a constant stream of detailed information about the environment and a precise tracking of the user’s respectively the vehicle’s movement for an accurate registration of virtual content (Van Krevelen & Poelman, 2010). Modern vehicles are equipped with a detailed road map and several sensors for the self-localization (e.g. GPS) and the detection of other road users and attributes (e.g. camera and radar). Nevertheless, the precision of those systems is limited, which may lead to
deviations (registration errors) up to several meters in the superimposition of the environment with virtual objects (Bengler et al., 2015).

In general several types of registration errors can be differentiated. There are position errors which describe translational deviations in lateral, longitudinal or altitudinal direction. Additionally rotational deviations in yaw, pitch and roll are referred to as orientation errors (Holloway, 1997). Position errors result in a constant offset between a real and a virtual object in any of the three directions. In contrast when it comes to orientation errors, the degree of the offset varies with the distance to the virtual object. Furthermore, latency and signal noise might occur. Latency leads to a smooth swimming of the virtual object in up to all six degrees of freedom, whereas noise results in an unsmooth jitter (Sadovitch & Wittkämper, 2018).

Pfannmüller et al. (2014) investigated the effects of inaccurate positioned augmented reality navigation cues in a driving simulator study. The participants chose more often a wrong turn when the virtual arrow on the street had a position error of more than 3 m to the ideal position. Intuitively, the usability and other subjective assessment items were rated significantly better, when there was no or just a smaller registration error. In another empirical study participants watched movies of real driving scenarios with an AR-HUD, filmed out of the driver’s perspective. Significantly negative effects on subjective system acceptance were reported with orientation errors of around 0.5° in yaw and 0.2° in pitch (Sadovitch & Wittkämper, 2018).

To improve registration precision and reduce latency several tracking and rendering approaches exist. To take advantage of the benefits of various tracking technologies the hybrid use of several tracking sensors is advised (Van Krevelen & Poelman, 2010). For example, the localization of a vehicle via GPS can be optimized by a sensor fusion with the data of a camera (Jachens et al., 2017). Another approach is to use the sensor data to predict the future position and orientation (Azuma, 1995). Besides an optimized tracking, rendering should be improved as well to reduce latency. Due to a long rendering pipeline the tracking data on which basis an image is rendered, is not the latest anymore at the time the image is displayed. Zheng (2015) describes several approaches to reduce rendering latency such as image displacement. Here, a larger image is rendered than displayed to be able to shift the image according to the latest tracking data just before displaying. Another approach suggests even to hide the image for a few frames, when it comes to very high frequented orientation changes of the vehicle (Tasaki et al., 2013).

In addition to technical tracking and rendering improvements, it is also possible to adapt the interface design of augmented reality visualizations to counteract registration errors. One possible approach is the use of contextual cues to decrease ambiguity of inaccurate registered information. In a laboratory study imprecise registered virtual objects were rather recognized as belonging to a specific real object, if the objects were similar in their form and color (Wegerich, 2012). Furthermore, some scientists assume that the design of virtual elements may lead to a lower or higher saliency of registration errors. In a collaborative AR application sharp edges were avoided in the design of virtual objects. Authors used blurred visualizations instead to make a possible overlap of virtual and real object less
visible (Fuhrmann et al., 1999). In case of a virtual bounding box or a frame to mark a real object, overlapping due to registration errors may even be prevented by extending the size of the virtual element (MacIntyre & Coelho, 2000).

While the improvement of technical approaches is a main focus in augmented reality science, there is little scientific effort done yet to explore non-technical solutions for the compensation of registration errors. Unfortunately, the literature offers just a few proposals that are not even evaluated yet. In the following sections a simulator study will be described to investigate some of the introduced error-tolerant HMI solutions and some further ideas to counteract registration errors in augmented reality head-up displays.

**Methods**

**Experimental Design**

The experiment was conducted as a 4 (event type) x 5 (visualization type) within subject design in a static driving simulator with one primary task and two secondary tasks. As the main task subjects had to monitor an Adaptive Cruise Control (ACC) assistance system while driving a vehicle on the highway. The system adjusted the vehicle velocity automatically to maintain a safe distance from vehicles ahead. For that a preferred target speed and distance had to be set. Although, the driver usually sets these settings, in this case the target distance was predefined and the speed was determined by traffic signs automatically. The subjects had to steer and monitor the assistance system, which is categorized as level 1 automation (SAE J3016, 2016).

Most vehicles were detected accurately by the ACC system and thus the vehicle regulated the velocity and distance. However, sometimes the system was erroneous and vehicles were not detected, which lead to the system not reducing velocity, despite a slower vehicle driving ahead. Most of the later situations did not necessitate an intervention from the driver. But every trial there was one critical event where the vehicle ahead suddenly braked. To avoid a collision, subjects had to recognize the system failure and brake manually. Summarizing there were four types of events in every trial:

- Event type no. 1: No ACC relevance
- Event type no. 2: Correct ACC behaviour
- Event type no. 3: ACC failure in safe distance
- Event type no. 4: ACC failure in critical distance

Braking reaction time in critical ACC failure events was analysed and considered as an indication for driving performance. To avoid collision scenarios in the stereoscopic simulation setting, the upfront vehicle accelerated automatically if the distance got shorter than 10 m. Bad driving performance was interpreted as an evidence for driver distraction due to the type of augmented reality visualizations, which are described in the following paragraph.
Participants were supported in the monitoring task by an augmented reality visualization in a simulated head-up display, indicating which vehicle was detected by the ACC system. Vehicles ahead with a relevance for the ACC adjustments of velocity were marked with an orange frame directly in driver’s field of view. To simulate registration errors the augmented reality marking was always swimming around the vehicle with orientation errors varying dynamically in between of 0 - 0.5° in pitch and 0 - 1° in yaw. There were five different visualization types used for the augmented reality frame (see figure 2). The basic frame is based on a simulator study, where Pfannmüller (2017) observed negative effects of registration errors in an AR-HUD on driving performance and subjective system assessment. The basic frame consisted of sharp borders and was not optimized towards an error-tolerant HMI. The other four visualization types were assumed to be more error-tolerant by using an extended, a diffuse, a fragmented and a reduced frame to mark the detected vehicle ahead. The subjective evaluation of registration precision was measured by several items of a questionnaire.

<table>
<thead>
<tr>
<th></th>
<th>Basic (A)</th>
<th>Extended (B)</th>
<th>Diffuse (C)</th>
<th>Fragmented (D)</th>
<th>Reduced (E)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No error</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
</tr>
<tr>
<td>Registration error</td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
</tr>
</tbody>
</table>

*Figure 2. Types of augmented reality visualizations with and without a registration error.*

In addition to the driving performance and the subjective perception of registration precision the mental workload was measured through a secondary tactile Detection Response Task (tDRT). For this purpose, a vibration motor was fixated at the wrist of participant’s dominant hand. Every three to five seconds the motor started to vibrate. The task was to press a switch, which was attached at the thumb, against the steering wheel, as fast as possible. If there was no response for 2.5 seconds the vibration stopped automatically and the event was logged as a miss. This task is suited for sensitive measurements of mental workload in primary tasks with high mental load (Bruyas & Dumont, 2013). Since the primary task in this study was very monotonous it was necessary to implement another secondary task to increase the overall task load. Therefore the participants performed an auditive N-back task (N = 1) additionally (Kirchner, 1958). The following letters were played in three seconds intervals: c, g, h, k, p, q, t and w. If the same letter was played two times in sequence, subjects had to signalize it verbally by saying the german word “Jetzt” (English translation: “now”). A verbal response was choosed instead of a tactile one, because participants had already to react to the vibrations of the Detection Response Task with a tactile response.
Material

To realize a realistic visual experience of the augmented reality head-up display, the experiment was conducted in a stereoscopic driving simulator. Participants had to wear a pair of shutter glasses to perceive a 3D image of the projected Virtual Test Drive simulation. The head-up image, which was rendered with the game engine Unity 3D, was shown as an overlay with a second stereoscopic projector. However, the simulation was limited to a single front view without side views. The authors consider this circumstance as acceptable since the driving task was concentrated on the traffic in front of the participants. The front view had a field of view of 60° and the projection surface was approximately 3.80 m away of driver’s eyes.

![Experimental setup with a participant wearing shutter glasses and tDRT apparatus.](image)

Procedure

In the beginning, the participants were informed about the experiment procedure and the handling of personal data, which was saved separately. All data was anonymized and only used for the purpose of the present study. It was ensured that no conclusions about the person are possible from the results. After agreeing voluntarily, participants were asked to fill in a sociodemographic questionnaire and perform the Butterfly Stereo Acuity Test to gather information about stereo vision capabilities for post-hoc analyses.

Afterwards the participants were introduced to the ACC assistant system and the augmented reality visualizations. The main task as well as the two secondary tasks were trained separately and together. The actual experiment contained five simulated driving scenes with different HMI conditions. The order of the conditions was balanced according to the latin square (Bortz & Schuster, 2010). After finishing the experiment the participants were debriefed and received a simple gift as an expense allowance. The experiment was designed to take 90 minutes.
Participants

Twelve females and 50 males participated at the experiment. All of the 62 participants were employees of the Volkswagen AG, working in different departments and participating voluntarily during their free time.

The participants had an average age of 36.05 years (SD = 9.65). Although nobody achieved an inadequate result in the stereo vision test, the stereo vision of six people was slightly impaired (> 100 seconds of arc). Since there were no conspicuity noticed in the post hoc tests, these participants were not excluded from data analyses.

Results

Subjective error tolerance

The subjective error tolerance of the examined HMI solutions was addressed by three items in the questionnaire: perceived offset, perceived swimming and visual impairment of registration errors (see figure 4). A repeated measures ANOVA was conducted for each item. While there was no significant difference concerning the perceived offset (F (4, 244) = 1.95, p = .103), a main effect of the visualization type was observed regarding the perceived swimming (F (4, 244) = 3.93, p=.004). A post hoc test with Bonferroni correction showed that the effect was mainly based on a significant difference (p = .015) between the best rated HMI (diffuse frame) and the worst rated one (reduced frame). This did not lead to any significant differences in the visual impairment of registration errors (F (4, 244) = 1.79, p=.131).

![Figure 4. Subjective rating of visualization types regarding error tolerance.](image)

After the experiment, subjects were asked to sort the visualization types in a ranking regarding the error tolerance. A significant difference was observed ($\chi^2 = 12.66$, $p = .013$). Over 50% of the participants ranked either the extended or the diffuse HMI on the highest rank and the basic HMI on one of the two lowest ranks with the
lowest error tolerance. In comparison with the basic visualization type the extended (p = .004) and the diffuse (p = .042) depiction were rated as significantly more error tolerant. The fragmented frame was rated just slightly, but not significantly better (p = .418) and the reduced HMI slightly worse (p = .879) than the basic visualization type.

**Further subjective ratings**

Additionally participants were asked to rate the visualization types regarding three positive attributes (*usefulness, design, detectability*) and three negative attributes (*distraction, masking, workload*). Mean results are shown in figure 5. A repeated measures ANOVA showed significant effects of the visualization type for every item except of distraction (F (3.37, 205.42) = 0.767, p=.587).

The extended HMI was rated as most useful and detectable. Furthermore, it demanded the least perceived mental workload. However, participants preferred the design of the fragmented frame, which masked less of the environment than the other visualization types. The highest masking was produced by the diffuse HMI, which was also harder to detect than the sharp edged designs. Overall, the reduced frame was rated worst. In average it was perceived as less useful and harder to detect than the other visualization types. Moreover, the reduced HMI caused the highest perceived mental workload and participants did not like its design.

![Figure 5. Further subjective ratings of visualization types.](image)

**Driving performance**

The driving performance was measured by analyzing the break reaction time in critical ACC failure events. Reaction times faster than 300 milliseconds were excluded from further analysis, because they suggest a rather coincidental break activity. In hindsight to the automatic acceleration of the upfront vehicle to avoid collision scenarios, reaction times bigger than two seconds were defined as missed
towards an error-tolerant HMI approach for augmented reality

reactions. Since only four cases of missed reactions occurred in more than 300 trials, no further analyses were conducted.

The smallest mean reaction time of 1,001.41 ms (SD = 230.66) was observed for the reference frame. Reactions were just slightly slower in the reduced HMI (M = 1,006.41 ms, SD = 238.37) and the diffuse HMI condition (M = 1,022.39 ms, SD = 255.98). Even the highest mean reaction times of 1,050.30 ms (SD = 234.39) and 1,062.78 ms (SD = 224.70) with the visualization types fragmented and extended frame were just slightly worse than the fastest reactions. Based on this data a repeated measures ANOVA showed no significant main effect for the visualization type (F (4, 180) = 0.83, p = .508) regarding the driving performance.

Secondary task performance

The performance in the n back task was not recorded, because the task was only implemented to increase overall task load. For an adequate interpretation of the performance data in the tDRT, stimuli were assigned post hoc to the different types of events. Participants received most of the vibration stimuli in unspecified events and only few stimuli in critical failure situations (see table 1). Since the augmented reality marking of detected vehicles is rather designed to increase situation awareness in general than to warn in critical situations, the tDRT performance in the event type no. 4 was not further analyzed.

Table 1. Average amounts of vibration stimuli in each event type per visualization type.

<table>
<thead>
<tr>
<th>Event type</th>
<th>Event type 2</th>
<th>Event type 3</th>
<th>Event type 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic HMI</td>
<td>35.51</td>
<td>8.08</td>
<td>3.43</td>
</tr>
<tr>
<td>Extended HMI</td>
<td>35.87</td>
<td>8.30</td>
<td>3.48</td>
</tr>
<tr>
<td>Diffuse HMI</td>
<td>36.08</td>
<td>8.72</td>
<td>3.41</td>
</tr>
<tr>
<td>Fragmented HMI</td>
<td>35.87</td>
<td>8.31</td>
<td>3.30</td>
</tr>
<tr>
<td>Reduced HMI</td>
<td>35.28</td>
<td>8.30</td>
<td>3.23</td>
</tr>
</tbody>
</table>

To analyze secondary task performance a two factors (event type x visualization type) repeated measure ANOVA was calculated. The results confirmed the sensitivity of the tDRT to measure differences in the mental workload within the present setting. A significant main effect of the event type was observed for the hit rate (p < .001) as well as the amount of false alarms (p = .008) and the reaction times (p < .001). The average reaction times for the separated types of events are shown in figure 6.
Only slight differences were found between the HMI conditions regarding secondary task performance. No significant main effect of the visualization type could be observed neither for hit rates \( F(4, 216) = 0.67, p = .612 \), nor for false alarms \( F(4, 216) = 0.975, p = .422 \) or reaction times \( F(3.57, 185.81) = 0.57, p = .664 \). Furthermore, there were no significant interactions of the factors event type and visualization type. This leads to the conclusion that there is no evidence for differences in the mental workload due to the type of the evaluated augmented reality visualizations.

**Discussion**

Limitations in the precision of tracking technology and maps lead to registration errors in augmented reality. Concerning an augmented reality head-up display, this might affect driving performance, usability (Pfannmüller, 2017) and subjective system acceptance (Sadovitch & Wittkämper, 2018). There are several technical approaches discussed in literature to improve tracking and rendering regarding precision and latency (e.g. hybrid tracking, predictive tracking, image displacement). Additionally the problem can be addressed by the use of an error-tolerant HMI, which does not improve the actual but the perceived registration precision.

The approach was evaluated using the example of an augmented reality adaptive cruise control assistance system in a stereoscopic driving simulator. Besides a basic frame to mark a detected vehicle ahead, four different potentially error-tolerant visualization types were tested regarding a counteracting effect on the perception of registration errors (see figure 2).
Subjective data showed that the participants perceived the highest registration precision with the use of an extended and a diffuse HMI, although the actual precision was not better than in the other trials. The extended frame was also very easy to detect and it was rated as very useful. Though subjects did not like its design as much as the fragmented ones for instance. A combination of these visualization types should be considered.

In contrast, the diffuse HMI revealed some more disadvantages. Participants criticized the design as well as the masking effects and a bad detectability. Therefore an optimization of this visualization type should be aimed. In total only very slight differences could be observed in the subjective assessment of error-tolerance. Overall registration precision were not assessed as poor as authors have assumed before. Therefore it should be taken in consideration to increase registration errors for further evaluations. Especially as the occurrence of stronger errors is to be expected in real driving conditions.

The reduced visualization type showed the lowest error-tolerance and usefulness as well. This might be explained by a missing connection to the marked vehicle. Participants remarked that they had even problems to determine the position in depth of the virtual element. For this reason the authors advise not to use too simple and abstract elements in augmented reality that have no direct connection to the environment.

There were no indications found for any differences in driving performance and mental workload due the visualization type. This allows implementing further augmented reality visualizations that are less similar to the basic frame. Even some dynamic patterns should be considered and evaluated additionally to the examined static visualization types.

References


Pfannmüller, L., Walter, M., & Bengler, K. (2014). Lead me the right way?! The impact of position accuracy of augmented reality navigation arrows in a contact analog head-up display on driving performance, workload and usability.


Effects of driving scenario on subjective workload and secondary task performance

Anna Pätzold¹, Rami Zarife¹, Michael Wagner¹, & Josef F. Krems²
¹Opel Automobile GmbH, ²Technische Universität Chemnitz
Germany

Abstract

Facing the digital revolution, new and extended functions will be available both on the smartphones and in in-vehicle information systems. To investigate the interaction effects of driving scenarios and the modality of a secondary task, a fixed-base simulator study was conducted (N = 42). Driving scenarios differed in perceived willingness to engage in a secondary task. Modalities of secondary task were visual-manual, Surrogate Reference Task, and cognitive-auditory, n-back task, both in three levels of difficulty versus no secondary task. Subjective workload was found to significantly differ between driving scenarios and the secondary task modalities. Meeting the predictions, scenarios subjectively eliciting a lower subjective engagement willingness were perceived higher in workload. Moderate task difficulty provoked the highest subjective workload in the driving scenario of a low perceived engagement willingness. Reaction time for the visual-manual task was highest in the driving scenario with the highest subjective workload. Ratio of hits, independent of difficulty, yielded no significant effect between the driving scenarios, indicating compensatory behaviour.

Introduction

As technology increases throughout our daily lives, we are faced with new chances and challenges, also as car drivers. Extended availability and functionality of in-vehicle information systems (IVIS) and driver assistance systems in addition to mobile device functions tempt the driver into distraction from the primary driving task. European drivers were found to engage in secondary tasks for 10.2% of the driving time, mostly in using their mobile phones (Carsten et al., 2017). It was found that 22% of all secondary tasks performed were texting, representing the third most obtained secondary task. Naturalistic Driving Studies both in Europe and the USA revealed that drivers’ engagement in non-driving related tasks differed between driving situations and manoeuvres (Carsten et al., 2017; Dingus et al., 2016; Huemer & Vollrath, 2017). Regarding the demand of the driving environment, medium, e.g. reading, and difficult tasks, e.g. engaging in the phone, were more likely to be executed in demanding situations and also longer in duration (Carsten et al., 2017). Engagement in secondary task was most likely on urban roads and country motorways, but little on country roads. Huemer and Vollrath (2017) and Ferreira et
Pätzold, Zarife, Wagner, & Krems (2012) found drivers to rather engage in their smartphones on highways than in cities. Further, modality of the executed secondary task influences the probability of a crash. Whereas hand-held mobile phone usage (visual-manual) increased the crash risk by an odds ratio of 3.6 (Dingus et al., 2016), speaking on the phone or with a passenger (cognitive-auditory) did not have a negative effect on crash risk and was even found to decrease the crash risk’s odds ratio to below 1.0, indicating a protective effect (NHTSA, 2012; Young, 2018). According to the Multiple Resource Theory (Wickens, 2003), the primary task of driving demands visual-manual resources, hence, simultaneously executing tasks demanding the same resources decreases performance in both tasks. As a visual-manual task, driving interferes negatively with additional visual-manual tasks, such as texting or dialing, leading to poorer driving behaviour (e.g. Reimer, 2009; Tsimhoni et al., 2004, Wandtner et al., 2016) and an increase of crash risk (Dingus et al., 2016; Fitch et al., 2013). Tasks of a different modality, e.g. talking to a passenger as of cognitive-auditory modality, do not directly interfere with driving and can therefore better be executed simultaneously. A less varying lateral and longitudinal vehicle control was found for the parallel execution of cognitive-auditory tasks (e.g. Engström et al., 2005). In comparison to no secondary task, Young (2015) found visual-manual secondary tasks to increase reaction times up to 450 ms and cognitive-auditory tasks up to 200 ms.

**Task Demands and Drivers’ Capability**

If task demands, evoked by both primary and secondary driving tasks, exceed the driver’s capability, a loss of control can occur and potentially result in a crash (Fuller, 2000, 2005, Figure 1).

Figure 1. Task Capability Interface Model (modified and extended based on Fuller, 2000, 2005).

According to the task capability interface model (Fuller, 2000, 2005), driving task performance is determined by the driver’s road position, trajectory and speed. In case there is an incongruence of the driver’s capability and the task demands, a
crash can be the consequence. The driver’s capability is determined by biological characteristics, e.g. information processing speed and capacity, physical constraints, and acquired characteristics, such as knowledge and skills (Fuller, 2005). These factors are vulnerable, as by the driver’s state. Drowsiness, stress, sickness and emotions are only some factors affecting the driver’s capability. Task demands in the driving task emerge from the environment, the vehicle and the speed. Bad road conditions and driving in bad weather possess further demand on the driver. Regarding the vehicle, the status of driver assistance and information systems as well as factors inside the vehicle, such as passengers, play a role. Active driver assistance systems can hereby lower the task demand by taking parts of the primary driving task over. As Fuller constituted, speed is a determining factor, restricted by a speed limit, traffic conditions and other road users as well as the driver. Since the driving task is self-paced, speed plays a crucial role as being controllable by the driver. As an extension of Fuller’s model, secondary tasks add to the driving task demands depending on their modality, frequency of execution and difficulty. As previously mentioned, secondary tasks interfering with the primary driving task lead to a loss in task and driving task performance. Under high task demands performance both in the driving and secondary task suffer (Engström et al., 2005; Young, 2015). Even if the task demands exceed the driver’s capability, a crash can still be avoided by compensatory behaviour by the driver or other road users. On the driving task side that can be adapting speed, road position and trajectory.

Driving scenario × Secondary tasks

In order to drive safely, drivers adopt their driving behaviour as a response to the circumstances the driving task invokes; they self-regulate their behaviour (Wandtner et al., 2016, Young et al., 2008). Given the opportunity, drivers chose to execute a secondary task in less demanding driving situations and showed better lateral and longitudinal vehicle control than drivers not being given the opportunity to decide themselves (Wandtner et al., 2016). The probability to self-regulate is influenced by some biological characteristics, such as age and gender. Older drivers were found to avoid driving under difficult circumstances as in bad weather, high traffic density or poor road-surface conditions (Charlton et al., 2006, Hakamies-Blomqvist, 1994, Stalvey & Owsey, 2003). Gwyther and Holland (2012) also identified an influence of age on self-regulating behaviour when controlled for driving experience, that is younger and older drivers rather reported self-regulating behaviour. Further, women rather engaged in self-regulation than men.

Pre-Studies

A previous online survey (N = 384) investigated the drivers’ perceived willingness to engage in a secondary task depending on both the task’s modality and the driving scenario (Pätzold et al., 2018). Driving scenarios were defined by road type, landscape, traffic density, weather and day time. As secondary tasks, read a text message, type a text message, watch a video, talk on the phone hands-free, make a shopping list mentally and adjust music volume were to be rated. A choice-based conjoint analysis was used to assess participants’ perceived willingness to engage in a secondary task. Across all tasks, road type and weather were identified as the two main factors influencing the perceived willingness. Participants assumed to rather
decide against engaging in a secondary task on a city road, whereas it was most likely on country roads. Regarding the task modality, visual-manual tasks (e.g. read a message) and cognitive-auditory tasks (e.g. talk on the phone hands-free) differed in the degree of influence of the context factors. Road type was found to have the biggest impact, except for the cognitive-auditory task. For the cognitive-auditory task, weather influenced the decision the most. The percentage of participants assuming to engage in the secondary task in all driving scenarios can be found in Table 1.

**Table 1. Perceived willingness to engage in the secondary task in every driving scenario**

<table>
<thead>
<tr>
<th>secondary task</th>
<th>percentage of participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>read a text message</td>
<td>6.7 %</td>
</tr>
<tr>
<td>type a text message</td>
<td>2.5 %</td>
</tr>
<tr>
<td>watch a video</td>
<td>2.1 %</td>
</tr>
<tr>
<td>talk on the phone hands-free</td>
<td>58.1 %</td>
</tr>
<tr>
<td>making a shopping list mentally</td>
<td>32.0 %</td>
</tr>
<tr>
<td>adjust music volume</td>
<td>70.4 %</td>
</tr>
</tbody>
</table>

Based on the choice-based conjoint analysis, five scenarios were chosen for the current study and the perceived willingness to engage in a secondary task was calculated, shown in Figures 2 and 3.

**Figure 2. Perceived willingness to engage in typing a text message while driving.**

*Note: A high total utility indicates a low perceived willingness to engage in the secondary task.*
Figure 3. Perceived willingness to engage in making a phone call hands-free while driving.
Note: A high total utility indicates a low perceived willingness to engage in the secondary task.

Aim of the study

The current study investigated the effects of secondary task engagement in different driving scenarios on the subjective workload and the secondary task performance. It was of interest whether the previously assessed perceived willingness to engage in the secondary task is reflected in the subjective workload and secondary task performance. Based on these findings it was assumed, that:

1. The higher the perceived willingness to engage in the secondary task, the lower is the subjective workload.
   For visual-manual tasks, it is the lowest in scenario A (dry country road) and the highest in scenario E (snowy highway). For cognitive-auditory tasks, it is the lowest in scenario A and the highest in scenario D (rainy city).

and

2. The higher the perceived willingness to engage in the secondary task, the better is the secondary task performance.
   For the visual-manual task, it is best (lower reaction time, higher ratio of hits) in scenario A and worst in scenario E. For cognitive-auditory tasks, it is best (higher ratio of hits) in scenario A and worst in scenario D.
Method

Sample

A total of $N = 42$ ($n = 11$ females) took part in the fixed-base driving simulator study. Participants were recruited from Opel’s internal participants pool, all being experienced in driving in a simulator. All participants held a valid driver’s license, drove $M = 22\,000$ km ($SD = 13\,000$ km) per year, and were $M = 42.7$ years ($SD = 9.1$, Range = 24-59 years) old.

Driving simulator and driving scenarios

The driving simulator uses Silab 5.0 (WIVW, 2015) as simulation software. Three displays in front of the integrated Opel Insignia give a $180^\circ$ view, one display in the back provides the back view and LCD displays are used as side mirrors (Figure 4).

Figure 4. Fixed-base driving simulator at Opel Automobile GmbH.
Five driving scenarios that differed in perceived engagement willingness were chosen (Pätzold et al., 2018). Screenshots of the scenarios from the driving simulator and their descriptions are shown in table 2.

*Table 2. Driving scenario descriptions*

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
<th>Speed limit</th>
<th>Scenario length</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Dry country road</td>
<td>70 km/h</td>
<td>5.83 km</td>
</tr>
<tr>
<td>B</td>
<td>Snowy country road</td>
<td>70 km/h</td>
<td>5.83 km</td>
</tr>
<tr>
<td>C</td>
<td>Dry highway</td>
<td>120 km/h</td>
<td>4.17 km</td>
</tr>
<tr>
<td>D</td>
<td>Rainy city</td>
<td>50 km/h</td>
<td>10.00 km</td>
</tr>
<tr>
<td>E</td>
<td>Snowy highway</td>
<td>120 km/h</td>
<td>4.17 km</td>
</tr>
</tbody>
</table>
Secondary tasks

In order to test different task modalities, the Surrogate Reference Task (SuRT, ISO14198, 2012) as a visual-manual, and the n-back task (Mehler et al., 2009, 2011) as a cognitive-auditory task were chosen.

The SuRT is a visual searching task, requiring participants to search for the biggest circle within 50 distractor circles (Figure 5). Here, participants were instructed to select the half of the display where the biggest circle appeared. The task was presented in three levels of difficulty, with a ratio of circle sizes of 150/80 as the easy, 150/100 as the moderate and 150/110 as the hard condition (Schwalm, 2009).

Figure 5. Surrogate Reference Task (SuRT, Petzoldt et al., 2014, p. 9).

The n-back task is a cognitive task, requiring participants to recall numbers from their short term memory. Following Mehler et al. (2009), participants were presented sequences of ten synthetically vocalized numbers ranging from 0 to 9 with 2.25 seconds pause in between the single numbers and 30 seconds between the sequences. A beep tone was played 2.5 seconds before the first number of a new sequence. The task was varied in three levels of difficulty, with 0-back as the easiest, 1-back as the moderate and 2-back as the hard condition. The easiest required participants to repeat the aforementioned number directly, which required a low level of cognitive attention. In the moderate condition, participants had to repeat the number that was presented before the current one, inducing a moderate load of cognitive attention by recalling an old and processing a new number in the short term memory. The hard required participants to repeat the number that was presented two values before the current one, provoking a high load on cognitive attention by recalling an old and encoding two new numbers in the short term memory.

Study Design

The order of conditions was balanced across the participants. The secondary tasks and their difficulty levels were randomly assigned to the driving scenarios. Participants were instructed that driving safely was the first priority, and that they shall maintain their position to the centre of the lane. They were asked to pay attention both to the driving and the secondary task. Within each scenario, the secondary task was executed for approximately five minutes while driving. Before each trial, the secondary task was practised for two times. After each scenario, participants were asked to fill in the questionnaire. In total, the study took 90 minutes.
Measures

Participants rated their subjective workload after each trial on the Driving Activity Load Index (DALI, Pauzié, 2008). The DALI measures workload on the subscales attention, visual load, auditory load, temporal load, interferences between the secondary and driving task and the perceived stress. As behavioural data, secondary task performance was measured. For the SuRT, the ratio of total number of hits in the total number of executed tasks (ratio of hits) and reaction times were assessed. The n-back task performance was measured as ratio of hits.

Results

Data was analysed using IBM SPSS statistics software (Version 22). The impact of the driving scenarios and the secondary task modality on subjective workload and secondary task performance was analysed with a repeated measures analysis of variance (rmANOVA). Further, the correlation of subjective workload and secondary task performance was analysed. The significance level for the analyses was defined as $\alpha = 0.05$.

Subjective Workload

A significant main effect for secondary task was found, $F(2, 74) = 107.85, p < .001, \eta_p^2 = .745$, Figure 6. Subjective workload was significantly higher when the SuRT, $p < .001$, or the n-back task, $p = .000$, were executed compared to no secondary task. Further, a significant main effect for driving scenarios was found, $F(4, 148) = 5.439, p < .001, \eta_p^2 = .128$. Overall, scenario D (rainy city), with a low willingness for both tasks, was rated significantly higher in workload than the other scenarios ($p < .032$).

Surrogate Reference Task

For the SuRT, a significant main effect for driving scenario was found, $F(4, 27) = 3.663, p = .017, \eta_p^2 = .352$. Both the scenarios with a high (B, snowy country) and moderate willingness (C, dry highway) were perceived significantly lower in subjective workload than the low willingness scenario (D). No significant main effect of task difficulty on the subjective workload was found, $F(2, 27) = 0.207, p = .815, \eta_p^2 = .015$. The interaction of driving scenario and task difficulty yielded a significant effect, $F(8, 27) = 3.293, p = .009, \eta_p^2 = .494$. For the easy task condition, the low willingness scenario (D) was rated higher in workload than the moderate willingness scenario (C). For the moderate task condition, the moderate willingness scenario (C) was higher in subjective workload than the high willingness scenario (B), and the low willingness scenario (D) was higher than the lowest willingness scenario (E, snowy highway).

n-back task

For the n-back task, a significant main effect for driving scenario was found, $F(4, 27) = 3.267, p = .026, \eta_p^2 = .326$. The scenario of the highest willingness (A, dry country road) was perceived significantly lower in workload than the scenario of the lowest willingness (D, $p = .019$). Further, task difficulty significantly influenced subjective workload, $F(2, 27) = 5.927, p = .007, \eta_p^2 = .305$. The hard condition was rated significantly higher in workload than the easy condition ($p = .006$). The
interaction of driving scenario and task difficulty yielded no significant effect, $F(8, 27) = 1.812, p = .119, \eta_p^2 = .349$.

Figure 6. Subjective workload scores (DALI) for perceived willingness in driving scenarios and secondary tasks. Error bars reflect Standard Error.

Secondary task performance

Surrogate Reference Task

Reaction times differed significantly between the driving scenarios, $F(4, 35) = 11.459, p < .001, \eta_p^2 = .647$, Figure 7. In scenarios of a high (A, B) and low (E) willingness, reaction time was significantly shorter than under the lowest willingness (D, $p < .001$). No significant interaction effect with task difficulty, $F(2, 25) = 0.360, p = .701$, $\eta_p^2 = .028$, or driving scenario and task difficulty, $F(8, 25) = 0.940, p = .502$, $\eta_p^2 = .231$, was found.

The ratio of hits for the SuRT did not differ significantly between driving scenarios, $F(4, 25) = 1.424, p = .255, \eta_p^2 = .186$. Task difficulty had neither a significant effect on the ratio of hits for the SuRT, $F(2, 375) = 2.889, p = .068, \eta_p^2 = .135$, nor on the interaction with the driving scenario, $F(8, 25) = 2.038, p = .083, \eta_p^2 = .395$. 
workload of driving scenario and secondary tasks

Figure 7. Reaction times for the SuRT. Error bars reflect Standard Error.

Relation with subjective workload

As shown in Figure 8, there is no clear indication of a relation between the ratio of hits, $r = .043$, or reaction time, $r = -.122$, and the subjective workload in the scenarios.

Figure 8. Relation of total number of hits and subjective workload scores (DALI).

n-back task

The ratio of hits revealed no significant effect of the driving scenario, $F(4, 27) = 0.728, p = .580, \eta^2_p = .097$. Task difficulty significantly influenced the ratio of hits, $F(2, 27) = 3.693, p = .038, \eta^2_p = .215$, but not the interaction with the driving scenario, $F(8, 27) = 0.402, p = .909, \eta^2_p = .107$. The ratio of hits was significantly lower for the hard difficulty (2-back) versus the easy (0-back) condition ($p = .037$).

Relation with subjective workload

No significant correlation of n-back ratio of hits, DALI score and driving scenario could be obtained, $r = -.238$, Figure 9.
Figure 9. Relation of total number of hits and subjective workload scores (DALI).

Summary

Subjective Workload
The hypothesis that the willingness to engage in the secondary task perceived in the driving scenarios predicts the subjective workload cannot fully be accepted. For the visual-manual task, the hypothesis was only partly supported. Contrary to the hypothesis, the scenarios inducing the lowest (scenario E, snowy highway) and the highest perceived willingness (scenario A, dry country road) were rated equally high in subjective workload. Workload was perceived the highest in the scenario of low engagement willingness (scenario D, rainy city), and following the assumptions being significantly higher than in the scenarios of high (scenario B, dry highway) and moderate (scenario C, snowy country road) willingness. Supporting the hypothesis for the cognitive-auditory task, workload was rated the highest in the rainy city scenario (D) and the lowest in the dry country road scenario (A). The hypothesis, that the driving scenario dependent willingness determines the subjective workload for a cognitive-auditory task was supported. Drivers seem well capable of predicting their engagement willingness based on the workload they perceive within that driving scenario.

Secondary Task Performance
The hypothesis that secondary task performance depends on the perceived engagement willingness in that driving scenario was only partly supported. Ratio of hits for both the visual-manual and the cognitive-auditory secondary task yielded no significant difference between driving scenarios. Reaction times for the SuRT yielded a significant effect between driving scenarios. Partly supporting the hypothesis, reaction times in the scenario of low engagement willingness (D, rainy city) were significantly higher than in the other scenarios. The results of ratio of hits for both the secondary-manual and cognitive-auditory task indicate compensatory behaviour. Drivers were able to sustain their secondary task performance regardless of the driving scenario’s demand and secondary task difficulty.
Since the results on subjective workload and reaction times supported the assumptions made on the online survey results, it can be assumed that drivers are capable of predicting their willingness on engaging in a secondary task.

**Limitations and further research**

There are some limitations to this study. First, it is important for the interpretation and comparison of secondary tasks that the possibility of self-regulation differed between the two tasks. Whereas the execution of the SuRT was self-paced by the drivers, the n-back task was system-paced. In real life, some cognitive-auditory secondary tasks can be controlled, e.g. by pausing before replying to the interlocutor or suppressing the conversation. Therefore, the results for the n-back task can only be transferred within limits to other cognitive-auditory task. The same accounts for the SuRT, since it was a low-involving visual-manual task, it cannot be applied unrestrictedly to a visual-manual task of high involvement, such as typing a text message.

Second, participants were instructed to execute the secondary task throughout the whole trial while prioritising the primary driving task. This way, extreme behavioural adaptations could be obtained, but further limited self-regulation. Thereby, it was not possible to measure the participants’ strategy and engagement in secondary task execution under real-world conditions (Young, 2018).

Third, due to differences in subjective workload but the sustained secondary task performance, effects of the driving scenario and secondary task performance on driving and gaze behaviour should be further investigated.

Fourth, the relationship between the online study results and the current simulator study and its reliability should be investigated. Therefore, the assessed utility values should be investigated as predictive factors for both the subjective and behavioural data.

**Conclusion and recommendations**

In order to be able to make an appropriate comparison, subjective workload for the secondary tasks was compared to the no-secondary-task-condition (Young, 2018). It was shown, that drivers seem to adopt to the driving scenario. Supporting the Multiple Resource Theory (Wickens, 2008), the cognitive-auditory task lead to less subjective workload than the visual-manual task in some driving scenarios, whereas in others, the visual-manual secondary task lead to less subjective workload. Due to the different effects of task modalities on subjective workload depending on the driving scenario, an IVIS should consider the task-and-scenario-varying workload. Since drivers adopt their driving behaviour in order to avoid a collision when task demands exceed his or her capability (Fuller, 2000, 2005), that compensatory behaviour shall be supported. In case of an overload by a visual demanding driving scenario and an interfering visual demanding IVIS task, the task modality can either be changed or the information within the IVIS can be adapted in quantity and position based on the relevance of the information for the driving scenario. Further research is needed to investigate the potential of adapting IVIS concepts.
References


workload of driving scenario and secondary tasks

DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).


When does the driver benefit from AR-information in a navigation task compared to a Head-Up Display? Results of a driving simulator study

Kassandra Bauerfeind\textsuperscript{1}, Julia Drüke\textsuperscript{1}, Lennart Bendewald\textsuperscript{1}, & Martin Baumann\textsuperscript{2}  
\textsuperscript{1}Volkswagen Aktiengesellschaft - Group Research, \textsuperscript{2}Ulm University, Germany

Abstract

With Augmented Reality (AR), additional information can be presented to the driver, which is directly related to the environment. Thus, the driver can perceive and understand the information more easily. Especially when driving with a navigation system, this visual support can lead to an improved understanding of the route in ambiguous situations. The aim of the driving simulator study was to examine the effectiveness and acceptance of AR-information in a navigation task compared to a Head-up Display (HUD). While driving in an urban area, the driving task was to identify the correct destination street by means of navigation information presented either in the AR display or the HUD. The decision-point for the right destination street, navigation errors and subjective data were recorded (N = 59; 37.1 years, SD = 11.1 years). Results showed that while driving with the AR display, participants found the correct destination street significantly earlier compared to the HUD. Furthermore, the AR display led to reduced navigation errors compared to the HUD. The AR display was evaluated as “more useful” and as “a greater reduction” regarding mental load than the HUD. Recommendations for an adequate use of AR-information for navigation can be derived from this study.

Introduction

In the late 1950s, the first Head-Up Displays (HUD) were in development by the military to present flight relevant information in the pilot’s field-of-view (Newman, 1995; Prinzel & Risser, 2004). Because of this head-up transparent screen, a concurrent scanning of the provided information and the environment is possible. In 1988, this technology found its way into the automotive field (Weihrauch, Meloeny & Goesch, 1989). The projection of light into the windshield or on a specific combiner screen enables the presentation of information into the driver’s field-of-view. Instead of looking down to gather relevant information in the instrument cluster, like speed and navigation information, the driver is able to keep his “head up” and receive all the necessary information on the bottom of his visual field. Since there is no need of looking down, driver glances away from the road can be reduced. Kim et al. (2013) indicated that glances away from the road can lead to divided attention, which can distract from the driving task and may therefore increase the risk of accidents. However, the presented information in a HUD are not directly
related to the environment. The driver has to map the information, hovering above the bonnet, to the real traffic situation. For the mapping a focal re-accommodation is necessary. Through the new technology of Augmented Reality displays (AR display) this mapping is highly facilitated. Augmented Reality enables to enrich the environment with virtual information, which is perspective-correctly superimposed with the relevant objects in the environment. The user gets the impression that specific objects can be marked in the environment. Because of spatial closeness of real and virtual objects, a fast and easy comprehensibility of information is possible. Furthermore, eyes can dispense with a focal accommodation beyond the distance of 10 meters (Cutting & Vishton, 1995). This is why an accommodation is not necessary in the interaction with AR-information.

Kim et al. (2013) and Tönnis (2008) determined that AR displays are suitable for warning systems since reaction times can be decreased. Rusch et al. (2013) found out that AR displays are efficient in leading drivers' attention to critical traffic events and objects as well as supporting them in detecting them. Because of the superimposition, the mapping of the presented information to the real traffic situation is facilitated. According to Bengler et al. (2015) and Pfannmüller et al. (2015) AR displays are supposed to require less mental effort regarding understanding and interpreting information in comparison to HUDs or map material (Kim & Dey, 2009), since the virtual information is already placed into the environment. As AR enables the driver to dispense with switching between the relevant information and the real traffic situation, this technology is also suitable for presenting navigation information to the driver. Because of the appearance of spatial proximity of the presented navigation information and the real street, the driver should benefit especially in a navigation task from the visual assistance through AR. According to Kim and Dey (2009) AR displays can help to reduce navigation errors and divided attention. It is expected that AR displays are particularly beneficial in ambiguous situations, for instance complex intersections or turning scenarios with many possible turns, which are very close to each other.

It is to be investigated to which extent the visual support of AR can lead to an improved understanding of the route especially in ambiguous situations. Thus, if finding the right destination is simplified with an AR display comparing to a HUD. Furthermore, it has to be determined whether navigation errors can be reduced with an AR display. It is assumed that drivers are able to decide earlier for the destination street in a navigation task than when driving with a HUD. Drivers are assumed to make less navigation errors (deciding for the wrong destination) with an AR display, than when driving with a HUD. The experiment was designed to examine the effectiveness and acceptance of an AR display compared to a HUD in an ambiguous navigation scenario. While driving in an urban area, the driving task was to identify the correct destination street by means of navigation information presented in the AR display or in the HUD.
Method

Navigation displays: HUD and AR display

In this study the two display types HUD vs. AR display were compared (see Fig. 1). The HUD is the current model version from AUDI with a perceived projection distance of about 1-2 meters, hovering above the bonnet. The projected picture in the HUD appears in 2D. In the study 300 m before the destination street the HUD presented navigation information in terms of a straight blue arrow to the driver (see Fig. 1, left). The turning direction (right or left) was presented to the driver once 80 m before the destination street is passed (see Fig. 1, left (2) & (3)). In contrast, with the AR display the driver experienced the augmented picture to be superimposed in his driving environment (see Fig. 1, right). The picture seems to be placed into the specific area in the surroundings, giving the impression to be a turquoise 3D arrow. This arrow was visible 300 m before the destination street and did not change when approaching the destination.

![Figure 1. Left: HUD as a current version of an AUDI HUD. Right: AR display gives the impression of placing a 3D navigation arrow into the driving environment.](image)

Driving scenario and driving task

In the study an urban driving scenario was used. Participants started by entering an urban area with many intersections on a straight road without any traffic lights. The intersections were 40 m apart. There were no other road users on the streets. Participants were asked to respect the road traffic regulations (50 km/h in cities). Each drive took about 1 min and 30 sec. The task was to identify the correct destination street by means of navigation information presented in the display (HUD or AR display). Because of the minor distance between the intersections the navigations task was ambiguous. The participants were asked to find the correct destination street. To detect this moment of identification, they had to switch off the display by pressing a specific button on the steering wheel (see Fig. 2). The experimenter emphasized to be able to ensure hitting the right destination street,
when switching off the display. Furthermore, the participants had to navigate to the destination. They received feedback concerning the correctness of their choice (right or wrong) by the experimenter. In case of a navigation error, they were not told which street was the correct one and had to repeat the scenario.

Additional task

In addition to the navigation task, the participants had to perform an auditory cognitive, spatial non-driving-related task. This was done to raise the drivers’ mental load, as the driving scenario was very easy when driving on a straight main road without other traffic. In the non-driving-related task the participants heard positive numbers from one to nine out of two speakers, one in front and one behind the mock-up. The task was to rate every number by answering “right” when hearing even numbers from the front speaker as well as odd numbers from the back speaker. Participants had to say “wrong” when hearing odd numbers from the front speaker as well as even numbers from the back speaker. The interval of the numbers was two seconds. The correctness of this task was recorded. The results of the non-driving-related task is not focused in this paper.

Figure 2. The red circle marks the button on the steering wheel, which the participants needed to press to switch off the display when deciding for the correct destination street.

Static driving simulator

The study was conducted in a static driving simulator with a mock-up from the Group Research of Volkswagen Aktiengesellschaft, equipped with an automatic gear and one projection screen 3.8 m in front of the mock-up. The projection screen’s width was 3.8 m and the resolution of the projector was 1920 x 1200 pixels. A 56° field of view was covered. The simulation was implemented with the software Virtual Test Drive (VIRES Simulationstechnologie GmbH, 2014). The two display types (HUD and AR display) were realised with the software Unity and were projected by a separate projector. The participants were able to experience the simulated world as well as the navigation information in the display in 3D due to a stereo simulation and shutter glasses.
Participants

The sample included 61 drivers. Because of simulator sickness and the lack of 3D-vision, data of 59 participants were analysed (18 female). The participants were on average 37.1 years old ($SD = 11.1$ years) and drove 18274.7 km per year ($SD = 11368.9$ km). 44% of the participants had gained experience with a HUD, whereas the majority indicated to rarely use it. The participants were recruited from the test driver pool of Volkswagen Group Research.

Experimental design

In the study a within-subjects design was used (see Fig. 3). Half of the participants started with the HUD in seven drives, followed by seven drives with the AR display. The other half of the group drove the two display types the other way around. Each drive lasted about 1 minute and 30 seconds.

![Figure 3. Experimental setup of the driving simulator study.](image)

In the study (1) the decision-point of the correct destination street, (2) navigation errors and (3) subjective data about the evaluation of the display types were recorded (see Fig. 3). The decision-point defines the distance to the destination street, by switching off the display. For the analysis, the last three drives out of seven drives per participant were considered. This was done to ensure that the participants were familiar with the task of switching off the display as soon as they identified the correct destination street. For analysing this data, a $t$-test for dependent samples was used. Furthermore, the navigation errors were determined. Deciding for the wrong destination street was counted as a navigation error and the participant had to repeat this drive at the end of the correspondent display type session. For the analysis, a Wilcoxon-signed-rank test was used. After the seven drives per display type, participants got an evaluation questionnaire, consisting of closed and open questions. The closed questions were answered with a 15-Point rating scale (Heller, 1982) (see Table 1 and Fig. 4). In addition to the evaluation questions, participants also rated the non-driving-related task as well as the driving situation after each drive. In the beginning of the study, they stated socio-demographic information,
such as age, gender, use of visual aids and right- or left-handed and gave information about their condition. Data were analysed using repeated measures Analysis of Variance (rmANOVA) to examine the effects of display type (HUD vs. AR display) and the order of test condition (starting with HUD vs. starting with AR display) as between subject factor.

Table 1. Examples of items from the evaluation questionnaire regarding the display types.

<table>
<thead>
<tr>
<th>Category</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>appropriateness</td>
<td>“How appropriate was the display for the navigation task?”</td>
</tr>
<tr>
<td>reduction of mental load</td>
<td>‘How strong did the display reduce your mental load?’</td>
</tr>
<tr>
<td>usefulness</td>
<td>‘How useful was the display for the navigation task?’</td>
</tr>
<tr>
<td>like/dislike</td>
<td>‘How much do you like the display for the navigation task?’</td>
</tr>
<tr>
<td>comprehensibility</td>
<td>‘How comprehensible was the display for the navigation task?’</td>
</tr>
<tr>
<td>comfort</td>
<td>‘How pleasant was it with the display in the navigation task?’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>not very useful</th>
<th>not useful</th>
<th>moderate</th>
<th>useful</th>
<th>very useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>+</td>
<td>0</td>
<td>-</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Figure 4. 15-point rating scale according to Heller (1982) with regard to the question: How useful was the display for the navigation task?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Procedure

In the beginning of the experiment, the participants stated the socio-demographic questions. They were also tested on stereo vision with the Butterfly Stereo Acuity Test (VAC Vision Assessment Corporation, 2007) and on colour vision. The participants were allowed to wear their own glasses underneath the shutter glasses, which enabled them to see the simulation in 3D.

The navigation task and the non-driving-related task were explained to the participants. Furthermore, they were instructed concerning the button at the steering wheel to switch off the display when they were sure they found the right destination street. After training the navigation task, they practiced the non-driving-related task without driving. Finally, they practiced both tasks simultaneously. The participants fulfilled the training till they were sure to manage the navigation task with each display type. After the training the experiment began. Half of the participants started with the HUD, followed by the AR display and the other half started with the AR display. They completed the two sessions, each with seven drives. In case of a navigation error (identifying the wrong destination street), they had to repeat the drive at the end of the respective display session. After each drive, participants also rated the non-driving-related task as well as the driving situation. Furthermore, they evaluated the two display types by the questionnaire after each display session. The participants received a present for their participation.
Results

Decision-point for finding the right destination street

The t-test for dependent samples showed that on average participants with the AR display switched off the display 69.5 m (SE = 2.8 m) before the destination street compared to 39.9 m when driving with the HUD (SE = 1.8 m), t(56) = -13.71, p < .001, r = .88 (see Fig. 5).

Figure 5. Boxplots of the decision-point for finding the right destination street of all participants for the last three drives per display type (N = 59).

Navigation errors

Navigation errors were not normally distributed, as assessed by the Shapiro-Wilk-Test, p < .001. Results of a Wilcoxon signed-rank test showed that driving with an AR display (Mdn = 0.10) lead to significant less navigation errors in comparison to driving with the HUD (Mdn = 1.03), z = -4.24, p < .001, r = -.39 (see Fig. 6). With an AR display 53 participants made no navigation error, whereas with a HUD 31 participants were error-free. Furthermore, participants never made more than one navigation error when driving with an AR display.

Subjective evaluation

The participants rated the AR display significantly better than the HUD (see Fig. 7a-f). There was a significant main effect concerning appropriateness (F(1, 57) = 138.90, p < .001, η²_p = .71). Thus, participants evaluated the AR display as more appropriate for the navigation task than the HUD (AR display: M = 12.7, SD = 1.6; HUD: M = 8.0, SD = 2.7) (see Fig. 7a). The interaction between display type x order of test condition was not significant (F(1, 57) = 0.90, p > .10, η²_p = .02). Furthermore, there was a significant main effect concerning the reduction of mental load (F(1, 57) = 130.39, p < .001, η²_p = .70; interaction display type x order of test condition, F(1, 57) = 1.33, p > .10, η²_p = .02). Participants rated the AR display as a
greater reduction of mental load for the navigation task than the HUD (AR display: $M = 11.2, SD = 2.2$; HUD: $M = 7.3, SD = 2.4$) (see Fig. 7b). There was also a significant main effect concerning usefulness ($F(1, 57) = 74.37, p < .001, \eta^2_p = .57$; interaction display type x order of test condition, $F(1, 57) = 0.08, p > .10, \eta^2_p = .00$).

Participants evaluated the AR display as more useful for the navigation task than the HUD (AR display: $M = 12.5, SD = 2.0$; HUD: $M = 8.6, SD = 2.9$) (see Fig. 7c). In addition, there was a significant main effect concerning like/dislike ($F(1, 57) = 214.74, p < .001, \eta^2_p = .79$; interaction display type x order of test condition, $F(1, 57) = 2.36, p > .10, \eta^2_p = .04$). Participants liked the AR display more for the navigation task than the HUD (AR display: $M = 11.9, SD = 2.3$; HUD: $M = 6.2, SD = 2.5$) (see Fig. 7d). Concerning comprehensibility there was also a significant main effect between the display types ($F(1, 57) = 67.33, p < .001, \eta^2_p = .54$; interaction display type x order of test condition, $F(1, 57) = 0.13, p > .10, \eta^2_p = .00$). Participants rated the AR display as more comprehensible for the navigation task than the HUD (AR display: $M = 12.9, SD = 1.8$; HUD: $M = 9.8, SD = 2.7$) (see Fig. 7e). Furthermore, there was a significant main effect concerning comfort ($F(1, 57) = 106.73, p < .001, \eta^2_p = .65$; interaction display type x order of test condition, $F(1, 57) = 0.00, p > .10, \eta^2_p = .00$). Participants evaluated the AR display as more pleasant for the navigation task than the HUD (AR display: $M = 12.4, SD = 1.8$; HUD: $M = 8.3, SD = 2.7$) (see Fig. 7f).

Discussion

The aim of this driving simulator study was to examine the effectiveness and acceptance of AR-information in a navigation task compared to a HUD. Since it is possible to present additional information to the driver, which are directly related to the environment, this technology could support the driver in particular in ambiguous situations, like complex intersections or close turning opportunities. Thus, the drivers should be able to find the correct destination street much earlier than when
driving with a HUD. As the marking of specific objects in the environment seems to be possible, this visual assistance provides an intuitive way of presenting and understanding the information. In the study it was also assumed that with an AR display less navigation errors will occur compared to the HUD.

Figure 7. Boxplots of subjective evaluation regarding appropriateness, reduction of mental load, usefulness, like/dislike, comprehensibility, and comfort (N = 59), (1: low consent, 15: high consent on the specific item).

Results of the study presented here confirmed that with the AR display the participants found the correct destination street much earlier compared to the HUD. According to the decision-point the participants found the correct street approximately 70 m before the intersection, in contrast to the HUD (40 m before the intersection). Because of the appearance of spatial proximity of the presented navigation information and the real street, the mapping is facilitated with AR. Participants seem to have an improved orientation in the traffic situation with the AR display than driving with the HUD. This can be confirmed regarding the
navigation errors: Participants showed less navigation errors when driving with the AR display. In the study of Kim and Dey (2009) the focus was on elderly drivers benefitting from AR-information regarding way finding. They made significantly less navigation errors with AR compared to map material in the centre console. It is also to note, that in the study presented here participants did not do more than one navigation error with the AR display. The AR display proved to present the information in an intuitive way, which implies this fast and easy comprehensibility of augmented information. Furthermore, in the study presented here participants preferred the AR display and rated this display type as more appropriate for the navigation task. This could be due to the fact, that they also evaluated the AR display as more comprehensible. As a result, they rated the AR display as more useful and as more pleasant than the HUD. The participants stated that driving with the AR display was a greater reduction of mental load. This is assumed by Bengler et al. (2015), Kim and Dey (2009) and Pfannmüller et al. (2015).

In regard to the limitations of this study, it has to be mentioned that in the driving scenario there was no other traffic on the street. The participants could decide for the right destination street without paying attention to other vehicles or pedestrians. This also implies that there was no masking of the destination street due to parked cars or heavy vehicles such as trucks. It would be interesting to investigate, whether drivers still benefit from AR-information, which are interfering with other road users. It has to be examined, whether the overlay of AR-information onto other road users still serve as a visual aid in navigation or might rather produce an information overload. Thus, the interaction with AR-information in a more realistic navigation scenario has to be examined. Object of a future study will be the investigation of the effectiveness and the acceptance of AR-information in a navigation task with different traffic density conditions. Results should also be validated in an on-road driving study.

According to further limitations of the study, the timing of presenting turning information in the two display types has to be considered. The HUD was based on a current version, which implied the time of presenting turning information to the driver (turning right or left). In the study, this information was presented once 80 m before the destination street was passed. In contrast, an AR display presented information concerning the turning direction from the very first moment the display turns visible (300 m before the destination street). Presenting relevant information directly at the place of the event, thus as early as possible, is a benefit of Augmented Reality, which makes the comparison with a HUD difficult. However, despite this early moment of presenting turning information in the AR display, the participants had to approach the ambiguous situation to perceive the exact localization of the AR-arrow. Future studies should concentrate on the time-wise comparability of presenting relevant information in these two display types.

**Conclusion**

In conclusion, this study contributes to a better understanding of drivers’ interaction with AR-information in a navigation task in comparison to a HUD. Results have confirmed that drivers can benefit from AR-information as this visual assistance provides an intuitive way of presenting information and thus a fast and easy
benefitting from AR-information in a navigation task

understanding for the driver. Especially in ambiguous situations, this technology supports the driver by a better orientation in the navigation task. The results have shown that the drivers recognized the right path much earlier and they made less navigation errors than with a HUD. Overall, this study supports the acceptance and effectiveness of an AR display in an ambiguous navigation task. Recommendations concerning the time-wise presentation of AR-information for an adequate use in navigation can be derived from this study. However, further research is needed to investigate other aspects of environmental factors on drivers’ interaction with AR-information. For instance, the effect of traffic density on the effectiveness of AR-information is necessary to examine. Furthermore, other information types such as warnings or system intervention (e.g. from advanced driver assistance systems) need to be investigated in future studies, as this study focused on navigation. Only with an overall understanding of the role of AR during driving, recommendations for an adequate use to support the driver will be possible.

References


Occurrence of motion sickness during highway and inner-city drives

Elisabeth Schmidt¹, Birte Emmermann², Joost Venrooij¹, & Klaus Reinprecht²

¹BMW Group, Research, New Technologies, Innovations
²Inspectio Forschungsinstitut UG
Germany

Abstract

Recent concepts for automated vehicles point towards future interior designs that allow passengers to consume media on in-vehicle displays during travel. It is widely accepted that the engagement in non-driving-related tasks, especially those that impede a view of the outside world, can lead to motion sickness. Two field studies were conducted investigating the occurrence of motion sickness while driving in two distinct environments: highway (n₁=296) and inner city (n₂=294). Both studies employed a between-subject-design in which the seating direction and non-driving-related task were varied. Seating direction was varied between forward- and backward-facing. The task varied from watching a movie on a handheld tablet, positioned on the passengers' laps, to watching a movie on a tablet mounted at eye-level in front of the passengers and a baseline condition in which the passengers did not watch a movie. The effect of seating direction, task and driving environment on the occurrence of motion sickness was evaluated using subjective misery scale ratings. Results showed that the occurrence of motion sickness was low across all conditions. Neither seating direction nor driving environment significantly affected the occurrence of motion sickness during both highway and inner city drives. These results indicate that the risk of motion sickness in automated vehicles may be smaller than is often assumed.

Introduction

While the technical developments within autonomous driving are progressing at a fast pace, central questions regarding the man-machine-interaction remain insufficiently addressed. Among these is the question of how severely motion sickness (MS) will affect the wellbeing of passengers. MS has been defined as “a condition characterized by pallor, nausea, and vomiting. It is brought about by exposure to real or apparent, unfamiliar motion to which the individual is not adapted” (Benson, 1992). According to the sensory rearrangement theory by Reason
and Brand (1975), MS is caused by stimulation that evokes “conflicting position and motion information arriving simultaneously from […] the vestibular system, the eyes and the non-vestibular receptors”. While driving, visual-inertial rearrangements are most likely to cause MS, e.g. when viewing static contents such as text in a book that are in conflict with the sensation of movement by the inertial receptors. In the context of autonomous driving the issue of MS gains importance because it enables settings in which MS is more likely to arise compared to manual driving. First, all vehicle occupants are now in a passenger’s role. Compared to the driver, the passengers are more prone to develop MS because they lack the control over the upcoming driving manoeuvres (Rolnick & Lubow, 1991). Second, automated driving of SAE (Society of Automotive Engineers) levels 3 and higher enables the passenger to engage in non-driving-related tasks (NDRT) because the system carries out the longitudinal and lateral control and monitors the driving environment (SAE, 2014). The engagement in NDRT such as watching a video or reading has been shown to increase MS in previous studies (Schoettle & Sivak, 2009). Third, several concept cars suggest a vis-à-vis seating arrangement in which some passengers are facing backward, which has also been shown to provoke more MS (Turner & Griffin, 1999a).

In order to understand the severity of the problem of MS in the context of automated driving, two on-road studies were conducted that investigated the occurrence of MS under varying conditions in terms of time-on-task, seating direction, NDRT and driving environment.

**Related Work**

MS has been previously studied both inside and outside the vehicular context. Therefore, insights have been generated in terms of causal factors (Reason & Brand, 1975), symptoms (Benson, 1992) and measurement techniques (Bos et al., 2005; Kennedy, 1975).

Regarding the measurement of MS, subjective and objective methods have been developed. Most important among the subjective measures is the single item 10-point MIserY SCale (MISC) by Bos et al. (2005). Another frequently used measure is the motion sickness questionnaire (MSQ) by Kennedy (1975), in which single symptoms of MS such as headache, stomach awareness, fatigue, blurred vision, nausea, cold sweating and vertigo are assessed. MS susceptibility of individuals can be assessed with the motion sickness history questionnaire (MSHQ) of Griffin and Howarth (2000) which consists of the retrospective report about sickness occurrence in different means of transport.

Previous research suggests that MS affects physiological responses which can therefore serve as objective indicators of MS. The study of Cowings et al. (1986) showed that skin resistance decreased during MS stimulation in people who were highly susceptible to MS. Similarly, Wan et al. (2003) observed a positive correlation between subjective MS ratings and phasic and tonic skin conductance levels. Decreased skin temperature has been found to decrease during MS (Bertin et al., 2005; Min et al., 2004).
In addition to the research on MS indicators, studies have been performed to investigate the occurrence of MS under various vehicular and behavioural conditions. Among the vehicular conditions, display position, seating direction and driving parameters have been shown to affect MS.

A display mounted at eye level (head-up) induced less MS than a display mounted below the glove compartment (head-down) (Diels et al., 2016; Kuiper et al., 2018). The reason for this was that the head-up condition allowed the passenger to visually perceive the direction of travel to a larger extent, resulting in smaller sensory conflicts and associated MS. Overall, both studies showed that symptoms of MS occurred after as little as 15 minutes of driving. A backward seating direction has also been shown to increase MS in survey participants because it impedes a forward view (Turner & Griffin, 1999a).

The influence of driving environments has been studied by Turner and Griffin (1999b) who found that nausea was higher with increased exposure to lateral motion at low frequencies (<0.5 Hz) and illness was found to be greater on cross-country routes. In addition, O’Hanlon and McCauley (1973) reported a dependency of MS on vertical acceleration and frequency.

In terms of behavioural conditions, Schoettle and Sivak (2009) found through a survey that NDRTs such as viewing a video or reading a text are related to MS occurrence. Similarly, the severity of MS was found to be lower when viewing a video compared to reading a text. MS was the lowest when passengers neither watched a video nor read a text.

The review of the above literature suggests that MS develops quickly under certain conditions and that vehicular interior arrangements affect the development of MS. The studies investigating the influence of certain vehicular and behavioural conditions have been conducted on test tracks (Kuiper et al., 2018) or using provoking tasks such as reading (Diels et al., 2016) during which MS is likely to arise. Therefore, it remains unknown how severe the problem of MS will be when autonomous driving will first be introduced on the highways and in cities and when passengers are conducting little provocative tasks.

Research question and hypotheses

In order to gain insights on the severity of the problem of MS in autonomous driving, two studies were conducted on real roads. In addition to evaluating the severity of the problem of MS in passengers during driving, the studies were designed to answer the following research question:

*Which effects do time-on-task, seating direction, NDRT, and driving environment have on subjective indicators of MS?*

The hypotheses were:

*H1: MS increases with time-on-task.*
H2: MS is higher for backward seated passengers than for forward seated passengers.

H3: MS is higher for watching a movie head-down compared to watching a movie head-up or watching no movie at all (baseline).

H4: MS is higher in the inner city than on the highway.

Method

The two studies were conducted investigating the occurrence of MS while driving in two distinct environments: highway and inner city. The methods of both studies were identical except for minor differences explained in the following.

Participants

In the highway-study 296 people participated of which 50 % were female. The average age of the participants was 43.9 years (SD=15.3 years; range=18-72 years). In the inner city-study 294 people participated and, again, 50 % were female. Here the average age was 43.8 years (SD=15.3 years; range=18-72 years). The participants were recruited by an independent recruitment agency. To avoid self-selection effects, the experiment description used for recruitment neither revealed the experiment’s goal, nor included any mentioning of MS, backward seating or NDRT. All participants were compensated financially.

Apparatus

The studies were conducted in a Volkswagen Multivan. The passenger cabin was equipped with four passenger (P) seats, of which the front two seats were rotated to face backwards. A table equipped with four adjustable tablet mounting points was positioned in the middle of the passenger cabin. The seating arrangement is shown in Figure 1. The table, the tablet mounting points, and the seats were positioned such that the tablets were centred with respect to the seats, at an eyepoint distance of approximately 620 mm, at a height of about 450 mm above the passenger’s hip joint. During the experiment, the tablets (Surface Pro 4.0, Microsoft Corporation) were used to present the NDRT and questionnaires to the participants. The experiment leader used an additional tablet to control the experiment, e.g., start and stop the measurements, monitor participants and data quality.

Vehicle accelerations, velocities and positions were measured using a 3-Degrees of Freedom Inertial Measurement Unit (Racelogic Ltd.). An ARCOS datalogger (CAETEC GmbH) was used to record obtained data. Physiological data of each participant was recorded using an E4 wristband (Empatica Inc.), which measured galvanic skin response and skin temperature.
occurrence of motion sickness during highway and inner city drives

Test routes

The test routes for both studies (highway and inner city) were located in and around Munich, Germany (Figure 2). The start and end of the highway-route were defined to be on the feeder road to the highway and on the exit of the highway, respectively, and served as a landmark at which data collection was started and ended. The route had a length of about 26 km and took about 26 minutes.

Figure 2. Routes for highway (left) and inner city (right) (Google Maps, 2018a,b).

The inner city-condition included urban streets in Munich. The route had a length of 11 km. Similar to the highway-condition, set landmarks determined the start and end of data collection for each drive. The average commute time to work in Germany is 24 minutes per day (Bach et al., 2007) which led to the decision to perform drives of similar durations.
Experimental design

Both studies employed a 2x3 between-subject-design in which the seating direction and NDRT were varied (Table 1). Seating direction was varied between forward and backward-facing. The NDRT varied from watching a movie (a documentary on the Bahamas) on a handheld tablet, positioned on the passengers’ laps (head-down), to watching a movie on a tablet mounted at eye-level in front of the passengers (head-up) and a baseline condition in which the passengers did not watch a movie. Each participant was assigned to only one condition. Figure 3 shows the independent variables.

Table 1. Experimental design showing the number of participants assigned to the different conditions (NDRT=non-driving-related task).

<table>
<thead>
<tr>
<th>NDRT</th>
<th>Baseline</th>
<th>Head-up</th>
<th>Head-down</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Highway</td>
<td>Inner</td>
<td>Highway</td>
<td>Inner</td>
</tr>
<tr>
<td>Seating</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>direction</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Forward</td>
<td>49</td>
<td>46</td>
<td>47</td>
<td>56</td>
</tr>
<tr>
<td>Backward</td>
<td>51</td>
<td>46</td>
<td>47</td>
<td>56</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>92</td>
<td>94</td>
<td>112</td>
</tr>
</tbody>
</table>

Figure 3. Non-driving-related task “Baseline”, “Head-up”, “Head-down”.

Procedure

The experiments on the highway and in the inner city took place in September 2017 and January 2018, respectively. The experiment consisted of several drives in which four participants could be tested simultaneously. In each drive, the task was identical for all participants and the four seats were assigned randomly.
Upon arrival, the participants received written instructions, describing the experiment process, tasks and safety instructions (Figure 4). The participants were informed that they can abort the experiment at any time and for any reason without repercussions. If a participant had wished to exit the vehicle, the driver would have stopped at earliest safe location. To motivate the participants in the head-up and head-down condition to direct their attention to the movie, they were told that they would be quizzed on the movie’s content after the drive.

The participants wore headphones which were connected to their tablets. Once the experiment leader had taken the passenger seat next to the driver, the experimental drive began. During the experimental drive, an audio cue interrupted the NRDT every three minutes and the MISC was prompted on the screen. In order to avoid extreme symptoms of MS the participants were automatically asked by the programme to contact the experiment leader once their MISC rating reached a value of 6 or higher.

![Figure 4. Procedure of experiments.](image)

<table>
<thead>
<tr>
<th>Table 2. Chronological order of questionnaires (BFI=Big Five Inventory, MSHQ=Motion Sickness History Questionnaire, MISC=Misery Scale, MSQ=Motion Sickness Questionnaire, NDRT=non-driving-related task).</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre</strong></td>
</tr>
<tr>
<td>Socio-demographical questionnaire</td>
</tr>
<tr>
<td>Recent/current illnesses</td>
</tr>
<tr>
<td>BFI-10</td>
</tr>
<tr>
<td>MSHQ</td>
</tr>
<tr>
<td>MISC</td>
</tr>
<tr>
<td>MSQ</td>
</tr>
</tbody>
</table>

*checkboxes for dizziness, cold/warm, headache, stomach/throat awareness, sweating, blurred vision, yawning, burping, tiredness, salivation (in case of MISC values from 2-5)
Questionnaires

To evaluate the subjective feeling of MS, ratings of the MISC and MSQ were taken in pre- and post-experiment-questionnaires (Table 2). The pre-test questionnaire also included questions regarding the participants’ socio-demographic information, recent and current illnesses, and the MSHQ. To control for the effects of personality, the Big Five Inventory-10 (BFI-10, Rammstedt et al., 2013) was included. The post-test questionnaire consisted of the MISC, MSQ, and questions about the content of the movie (for NDRT head-up and head-down).

Results

Of the initially N=590 datasets some had to be excluded from analysis due to technical issues with the data recording. The exact number of datasets considered will be reported for each procedure. To ensure comparability between the data recorded on highway and inner city, the first six measurements of MISC during the experimental drive were considered for the statistical analysis. This is because the duration of the experimental drive was shorter in case of low traffic density, leading to missing data in some of the drives after 18 minutes.

Descriptive analysis

Table 3 shows the percentage of participants experiencing different degrees of MS. MISC-values equal or higher than 6 were only reported by 0.25 % of all participants. The majority (about 90.39 %) neither had problems nor felt uneasy (MISC-rating between 0 and 1). The remainder of the participants (9.36 %) reported vague to severe symptoms of MS (MISC-rating between 2 and 5).

Table 3. Distribution of MS across factors for pre-rating, rating between minutes 0 to 18, and post-rating (MISC=Misery Scale).

<table>
<thead>
<tr>
<th></th>
<th>No problems / uneasiness (MISC 0-1)</th>
<th>Vague to severe symptoms (MISC 2-5)</th>
<th>Slight to severe nausea, retching, vomiting (MISC 6-10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (N=583)</td>
<td>90.39 %</td>
<td>9.36 %</td>
<td>0.25 %</td>
</tr>
<tr>
<td>Forward (n=291)</td>
<td>90.41 %</td>
<td>9.20 %</td>
<td>0.39 %</td>
</tr>
<tr>
<td>Backward (n=292)</td>
<td>89.61 %</td>
<td>10.35 %</td>
<td>0.04 %</td>
</tr>
<tr>
<td>Head-up (n=205)</td>
<td>88.75 %</td>
<td>10.95 %</td>
<td>0.31 %</td>
</tr>
<tr>
<td>Head-down (n=187)</td>
<td>87.87 %</td>
<td>11.79 %</td>
<td>0.34 %</td>
</tr>
<tr>
<td>Baseline (n=191)</td>
<td>93.47 %</td>
<td>6.53 %</td>
<td>0 %</td>
</tr>
<tr>
<td>Highway (n=289)</td>
<td>88.61 %</td>
<td>11.26 %</td>
<td>0.13 %</td>
</tr>
<tr>
<td>Inner city (n=294)</td>
<td>92.12 %</td>
<td>7.58 %</td>
<td>0.30 %</td>
</tr>
</tbody>
</table>
Statistical procedure

A multivariate (6x3x2x2; time of measurement, NDRT, seating direction, driving environment) repeated-measures MANOVA was conducted with N=565 participants with the first factor being the repeated measure. In the following, univariate results are reported to answer the respective hypotheses. If interactions were found, they are reported. All analyses assume multivariate significance.

Hypothesis 1: Motion sickness and time

It was hypothesised that MS changes across the duration of the experimental drive. As visualized in Figure 5, hypothesis 1 is confirmed ($F(6,546)=12.135$, $p<.000$; Hotelling’s $T^2=0.133$, partial $\eta^2=.118$): MS significantly differs between each time point, except time point 1 to 2. However, on the 10-point MISC, mean MS was very low ($M=0.42$). Therefore, based on the results of this study, the occurrence of MS can be deemed negligible for practical application. Nevertheless, participants occasionally experienced higher levels of MS (Table 3).

![Figure 5. MISC-rating (=Misery Scale) (M, SD) over time. (*p<.05)](image)

Hypothesis 2: Motion sickness and seating direction

Hypothesis 2 assumed that the experience of MS is influenced by the seating direction of the participants. However, no significant difference (Figure 6) can be
found between the conditions \((F(1,553)=0.783, p>.05, \text{partial } \eta^2=.001)\). Thus, the occurrence of MS cannot be explained by the direction the participants are facing while the vehicle is moving.

**Figure 6.** MISC-rating (=Misery Scale) \((M, \text{SD})\) for different seating directions.

**Hypothesis 3: Motion sickness and non-driving-related task**

Evaluating Hypothesis 3, the difference in MS due to the NDRT is examined. As visualized in Figure 7, a significant difference between BL \((M=0.22; SD=0.672)\), head-down \((M=0.42; SD=0.909)\) and head-up \((M=0.43; SD=0.125)\) was found \((F(2,553)=4.579, p<.05, \text{partial } \eta^2=.016)\). Subsequent Bonferroni-corrected post-hoc analysis revealed that the identified difference lies between the baseline and the head-down condition \((-0.23, 95\% \text{ CI}[-0.43, -0.03], p<.05)\) and between the baseline and head-up condition \((-2.40, 95\% \text{ CI}[-0.43, -0.05], p<.01)\).
The results indicate that the occurrence of MS is affected by whether the passengers watched a movie or not. However, if the participant’s attention was allocated...
towards the movie, the magnitude of MS was not influenced by the vertical positioning of the tablet.

**Hypothesis 4: Motion sickness and driving environment**

Hypotheses 4 assumed a difference in MS in dependence of the driving environment, inner city or highway. Statistical analysis showed no significant difference in the occurrence of MS \((F(1,563)=0.200, \ p>.05)\) between the driving environments (highway and inner city) (Figure 8). However, a difference in the MS for the course of the experimental drive between highway and inner city can be observed: While participants in inner city reported lower levels of MS than the participants in highway, this tendency turned after about 12 to 15 minutes. This might be explained by the different driving characteristics of inner city which cause MS after a certain time in traffic.

![Figure 8. Mean absolute lateral and longitudinal accelerations for different driving environments (M, SD) (*p<.05)](image)

Additionally, a comparison of the driving characteristics of inner city and highway revealed significant differences in longitudinal and lateral acceleration confirming...
that the inner city-drive was associated with higher average vehicular accelerations ($F(4,1900)=591.424, p<.05$, partial $\eta^2=.555$) (Figure 9).

Results from the other questionnaires, like BFI, MSQ and MSHQ and the physiological measurements are not reported in this paper due to constraints of length.

Discussion

The most striking result of this research was that only little degrees of MS ($M=0.42$) were measured and that a very small percentage of participants experienced nausea (percentage of people with MISC of 6 or higher $<1\%$). This could be due to the selected NDRTs which were watching a movie (head-up and head-down) and no task (baseline). Based on the observations of Schoettle and Sivak (2009), higher MS is likely to occur when reading a text. While watching a movie, the visual input is not as static compared to reading a text. This might have caused a lower discrepancy between the visual and inertial system and hence low levels of MS.

The study provided several insights with regard to the research question and the four hypotheses of which only one could be confirmed. As expected, MS increased with travel time. This result has been observed in previous studies (Kuiper et al., 2018) and McCauley et al. (1976) once described MS to increase monotonically under constant motion.

In contrast with results of the survey of Turner and Griffin (1999a), this study showed that MS was not affected by the seating direction. One reason for this difference could be that the seating direction in their survey came along with different impediments of the outside view. In this study, however, the view was impeded similarly in both forward and backward seating because of the armatures for the tablets and the NDRT. This is an important difference between this study and the one of Turner and Griffin (1999a), however, the lack of effect of seating direction on MS is still surprising and should be further investigated.

In alignment with the study of Schoettle and Sivak (2009), this study found that the NDRT affected the occurrence of MS. Like in their study, MS was higher when passengers watched a video (head-up and head-down) compared to no task (baseline). The result on the effect of display position of Diels et al. (2016) and Kuiper et al. (2018), however, could not be reproduced in this study because there was no difference between the head-up and head-down condition. The reason for this could be that in the previous studies, the display in the head-up condition was mounted directly behind the windscreen and the participant sat in the passenger’s seat of the car, thus, allowing peripheral view to the road ahead. In our study, the view in the head-up condition was obstructed to a larger extent, because of the vis-à-vis seating arrangement in the back of a van. This larger obstruction in the head-up condition could have played an important role for the rejection of the hypothesis.

Contrary to the results of Turner and Griffin (1999b) who found increased MS on cross-country routes compared to highways, the driving environment did not affect MS in this study. Even though the analysis of driving data could confirm larger
lateral and longitudinal accelerations in the inner city, the MS occurrence was not significantly higher compared to the highway. An explanation for the lack of the effect could be that the MISC-ratings were very low in general, suggesting that in both environments, the motion was not sufficiently provocative for most of the study participants. Another reason for the lack of alignment of this study to the one of Turner and Griffin (1999b) could be that their cross-country routes were associated with higher lateral and longitudinal accelerations than the inner city-route of this study, and hence leading to significant differences compared to highways.

The study came along with several limitations. The first limitation is that the study was conducted in different seasons, and hence the ambient air temperature varied between the highway and inner city-study. Additionally, naturally occurring changes in temperature were observed within the studies which were not sufficiently compensated by the vehicle’s air conditioning which was permanently set at 20 °C. According to Turner and Griffin (1999a), though, ambient temperature was not found to influence MS occurrence.

Another limitation was found to lie in the subjective assessment of MS using the MISC. Whereas the MISC-values were increased significantly over the duration of the experiment, the measurement taken before the experimental drive showed a higher average MISC-rating among the participants. Likely, this did not occur from MS itself but rather as a relic of nervousness and heightened attention towards the participant’s own wellbeing, or, as the participants were seated in the experimental vehicle shortly before the questionnaire was taken, anticipated MS as a reaction to the experimental setup. Accordingly, it is suggested that the MISC does not only measure symptoms related to MS in the lower range of the scale.

**Conclusion**

The key finding of the study was that – contrary to the often made assumption of MS in automated vehicles – the magnitude of MS has been shown to be fairly low. Descriptive analyses showed that only 9% of the n=583 participants felt distinctive symptoms of MS and less than 1% reported nausea. The mean rating of 0.42 on a scale from 0 to 10 indicated that on average almost no MS was experienced during the course of the study.

The studies further showed that even drives in the inner city, associated with higher lateral and longitudinal accelerations compared to highway drives, did not provoke a significant increase in MS. Similarly, the seating direction did not affect MS. Based on the findings of this study, new concepts of interior design do not need to be restricted to the classical seating arrangements of current vehicles but can comprise more divergent options including backward seating. Even when occupied with NDRTs passengers who were faced backwards experienced close to no symptoms of MS.

However, the ratings of MS were found to be significantly influenced by the factors manipulated in the experiment: time-on-task and NDRT. In both cases, the effects were very small. Long driving durations could still bear the risk of increased MS. While the chosen duration of approximately 26 minutes in the studies reflected the
average commute time to work in Germany (Bach et al., 2007), it remains unknown how MS would develop when driving longer in these environments.

As discussed above, different NDRTs associated with more static visual input will probably lead to higher MS. Therefore, it is worthwhile to investigate on the one hand longer times-on-task and on the other hand more provoking NDRTs in future studies.

References

Kennedy, R.S. (1975). Motion sickness questionnaire and field independence scores as predictors of success in naval aviation training. Aviation, space, and environmental medicine, 46, 1349-1352.


Preferences of European cyclists towards passive and active systems with audio-visual and with handlebar vibration warnings

Marco De Angelis¹, Federico Fraboni¹, Gabriele Prati¹, Davide Giusino¹, Marco Depolo¹, Bruna Zani¹, Luca Pietrantoni¹, Jeremy Shires², & Daniel Johnson²
¹Department of Psychology, University of Bologna, Italy,
²Institute of Transport Studies, University of Leeds, United Kingdom

Abstract

In Europe cyclists suffer a disproportionate share of serious injuries and fatalities, and in recent years that disadvantage has been growing. To minimise accidents, different types of collision warning systems have been proposed for motor vehicle drivers. Few studies are available on on-bike systems. The H2020 EU-project XCYCLE developed systems aimed at improving active and passive detection of cyclists informing both drivers and cyclists of hazards at junctions. In the present study, preferences for four different on-bike systems were investigated: 1) passive system without warning; 2) active system with audio-visual warning; 3) active system with handlebar vibration warning; 4) active system with both audio-visual and vibration warning. A sample of 2381 European cyclists from six countries (Italy, Spain, Hungary, Netherlands, Sweden, United Kingdom) answered the online questionnaire. The passive system was the most preferred among the respondents (67.8% would use). Cyclists aged 35-46 showed higher preference for all systems, while those aged 47 and older expressed the lowest preferences for all systems. Participants used to cycle 1-3 day per month exhibit a statistically significant lower intention to buy any type of active on-bike systems.

Introduction

The number of road accidents is increasing and causing significant losses in many countries, expecting to become the seventh major cause of death by the year 2030 (World Health Organization, 2018). In Europe cyclists suffer a great share of serious injuries and fatalities, partly because they can be considered as vulnerable road users (VRUs) due to a lack of physical protection in the event of a crash with motorised vehicles (Prati et al. 2018). According to the report of the European Road Safety Observatory (ERSO, 2017), while between 2006 and 2015 the number of both overall road fatalities and cyclist fatalities in EU decreased steadily, the number of cyclist fatalities has remained almost unchanged since 2010 and the percentage of cyclists of all road fatalities increased from 7% in 2006 to 8% in 2015. More people in Europe are using the bicycle as an unexpansive, convenient and environmentally safe mode of transport, however it is a fact that cyclists might be more exposed to fatalities (i.e. more exposure, more likelihood to incur in a critical accident).
As reported by Prati et al. (2017), the type of opponent vehicle is one of the most influential factors impacting the cyclists’ crash severity. A situation especially dangerous for cyclists is a conflict with a heavy goods vehicle (HGV) due to their difference in speed and mass (Kröyer, 2015). Previous studies also showed that the increasing levels of van, large automobile, and truck traffic are associated with higher collision risk (Vandenbulcke et al., 2009; Ackery et al., 2011).

In order to improve road safety, in recent years the development and use of intelligent transport systems and advanced assistance systems have received more and more attention (Jamson et al., 2013) whilst vulnerable road users have seldom been addressed as main target (De Angelis et al., 2017). The H2020 EU-project XCYCLE (http://www.xcycle-h2020.eu) developed on-bike warning systems to inform truck drivers and/or cyclists (depending on the type of system) about the crash risk at intersections which are considered the third most dangerous scenario for bicycles compared to other modes of transport (ERSO, 2017). The passive on-bike system (also called “bike tag”) warns the truck driver of the presence of a cyclist in the proximity of the truck. On the other hand, the active on-bike system warns the cyclists, alerting them about a dangerous situation. This is done either by means of audio-visual warning (through a device mounted on the handlebar and equipped with a buzzer and a number of LEDs), tactile warning (by means of a vibrating device concealed in the handlebar), or by the combination of these modes (i.e. trimodal warning with acoustic, visual, and tactile warning). The combined active and passive on-bike system warns both the cyclist and the truck driver about the collision danger.

While previous studies evaluated preferences and reactions of drivers to different type of warning modes (e.g. Dettman & Bullinger, 2017; Naujoks et al., 2016; Toffetti et al., 2014), to the authors’ knowledge, only few studies aimed at exploring cyclists’ preferences and reactions to intelligent safety devices with different warning modalities (Engbers et al., 2016; Prati et al., 2018). Particularly, Engbers et al. (2016) evaluated the effects on behaviour, mental effort and acceptance of an electronic rear-view assistance system between two modalities (i.e. visual and haptic) and found less mental effort and significantly more correct decisions regarding a safe left turn when using the rear-view assistant. Similarly, Prati et al. (2018) evaluated users’ acceptance of an on-bike system warning about potential collisions with motorised vehicles, as well as its influence on cyclists’ behaviour, and found that participants were relatively likely to accept the on-bike system, as well as likely to decrease their speed in case of warning of the on-bike system.

Grounding both on classical socio-psychological theoretical contributions (Ajzen, 1991) and the Technology Acceptance Model (TAM; Davis et al., 1989), it is known that the acceptance of the system has an influence on its future use (Ghazizadeh et al., 2012). Therefore, it is of interest to study the preferences of cyclists in terms of warning modes, especially during the design process of an assistance system (Van der Laan et al., 1997). The aim of the present study was to investigate the willingness to purchase warning modes of regular cyclists, as defined as the likelihood that an individual intends to purchase this product (Dodds et al., 1991). In particular, the preferences of warning modes of regular cyclists in relations with age,
preferences of EU cyclists towards active system warnings

...cycling frequency and country of residence were studied. Indeed, previous studies investigating users’ preferences of advanced driver-assistance systems (ADAS) have shown the effect of age (Souders et al., 2017) on the evaluation and the acceptance of such road user support systems (e.g. blind spot detection system). Results showed that increased age was associated with a higher valuation for the Active Blind Spot Detection System, even though older adults generally have less familiarity with new technologies (Czaja et al., 2006). To the best of authors’ knowledge, the existing literature has rarely been focused on investigating age differences towards on-bike warning systems. Furthermore, since familiarity positively affect the acceptance of a new system (Chang, 2010), a higher willingness to purchase each type of safety device might be expected from regular cyclists who use to cycle every day, and from cyclists who live in countries where cycling level is popular (i.e. the Netherlands and Sweden).

Method

Procedure

A web-based survey was administered to an online panel of respondents who have agreed in advance to take part in the data collection and resided in six countries (Hungary, Italy, Spain, Sweden, Netherlands, United Kingdom). Since the aim of this study was to target regular cyclists, selection criteria were used for each country. To be included in the study dataset, all respondents had to make on average at least 1 cycle trip per month, at least 50% of them had to be regular cyclists (i.e. make on average more than 2 cycle trips per week), at least 30% had to be female, and at least 10% of the sample had to be aged fifty years or more. Before the survey, the participants watched the video showing the functionality of the examined on-bike systems and the warning modes they employ.

Measures

Warning mode preference. As part of a larger set of Stated Preference items, participants were asked by four items “I would not purchase [system type]”, with possible options 1 (yes) and 2 (no). The preference for each warning mode was calculated as percentage of participants’ willing to purchase the system. Also, because participants expressed their preference independently for each system type, total percentages reported do not sum up to 100.

Cycling frequency and Country of residence. To measure cycling frequency, participants were asked “How many months a year do you normally cycle?”. This prompted them to consider their cycling behaviour only during these months and allowed us to account for local geographical differences in terms of weather limitations for bicycle use. To the second item, “In general, during these months, how often do you cycle?”, the participants responded using a five-points frequency scale with options 1 (daily), 2 (3 or more days per week), 3 (1-2 days per week), 4 (1-3 times per month), and 5 (less than once per month). Furthermore, country of residence was considered in the analysis to explore the potential effect of a different national cycling level as defined as the average share of cycling as main mode of
transport (Directorate-General for Communication of the European Commission, 2015) on the willingness to purchase a new on-bike safety system.

**Demographics.** The first part of the questionnaire asked participants about their age, gender, and nationality.

**Data analysis**

We created three age bands using tertiles, these were 18-34 years, 35-46 years, and 47 years and more. These categories were created to obtain subsamples with similar sizes. Once the descriptive analysis for demographic data and warning mode preferences had been performed, Friedman's non-parametric ANOVA analysis was carried out to explore significant differences among warning mode preferences. In a second step, a chi-square analysis with Bonferroni correction (i.e. the level of significance was adjusted by dividing the original level of 0.05 by the number of multiple tests performed during each comparison) was conducted to explore any significant differences between age bands, cycling frequencies and national cycling levels.

**Participants**

The questionnaire was completed by 2389 respondents. The mean age was 42.75 (SD = 14.34), with the range from 18 to 86 years, with 764 participants with age between 18 and 34 years, 761 participants with age between 35 and 46 years, and 856 participants aged 47 or older. The whole sample consisted of 1210 females (50.6%), 1171 males (49%), and 8 (0.3%) identified themselves as transgender. Concerning the bicycle use, 365 (15.3%) participants used bicycle 1-3 times a month, 707 (29.7%) participants cycled 1-2 days a week, 872 (36.6%) did so 3 or more days a week, and the remaining 437 (18.4%) cycled daily. Out of the whole sample, there were 2381 cyclists in total, i.e. Sweden (n=392), Netherlands (n=395), Hungary (n=399), United Kingdom (n=396), Spain (n=396), and Italy (n=403). In light of these data, all the selection criteria were met.

**Results**

**Warning mode preference**

Overall, results revealed a statistically significant differences on warning mode preference ($\chi^2_{3,2381} = 33.27, p<0.001$) with the passive bike tag (without warning) as the most preferred among the respondents (67.8%), followed by active system with audio-visual warning (65.9%), active system with handlebar vibration warning (65.5%) and active system with trimodal warning (62.8%). To explore more in depth the differences between the four type of warning modes, the Wilcoxon signed-rank test showed significant differences among all the warning mode (p< .05) except for the preferences between the active system with audio-visual warning system and the active one with handlebar vibration warning ($Z=-.51, p=.60$). The passive tag obtained higher preferences compared to the audio-visual warning ($Z=-2.06, p=.03$), the system with handlebar vibration warning ($Z=-2.48, p=.01$) and the trimodal warning ($Z=-4.92, p<.001$).
Also, the relation between age groups and warning mode preference was examined. The passive system was the most preferred among cyclists from the 35-46 age group (72.0%) and the 18-34 age group (70.5%), whereas in the group aged 47+ the mode preference was lower (61.7%). These differences were statistically significant ($\chi^2_{1, 2381} = 23.51$, $p<0.001$). The same mode preference pattern was seen for the active system with tactile warning ($\chi^2_{1, 2381} = 42.03$, $p<0.001$), for the active system with audio-visual warning ($\chi^2_{1, 2381} = 33.93$, $p<0.001$) and for the active system with trimodal warning mode ($\chi^2_{1, 2381} = 42.63$, $p<0.001$). In general, cyclists from the 35-46 age exhibited a statistically significant highest willingness to purchase each type of system ($p<.01$) whereas the older participants represented the group with the statistically significant lower intention ($p<.001$). See Figure 1 for summary.

![Figure 1. Percentage of participants willing to purchase on-bike systems by age group.](image)

In addition, differences in preferences between groups defined by cycling frequency were examined. In general, there was a positive pattern between cycling frequency and willingness to use each type of on-bike system (i.e. the higher the cycling frequency, the higher the willingness to use the system), except for the trimodal one. The Bonferroni post-hoc correction revealed that those who are used to cycle 1-3 day per month exhibit a statistically significant lower intention to buy any type of active on-bike systems compared to the other groups ($p<.001$). For summary of these results, see Figure 2.
Furthermore, cyclists from different national cycling level expressed their preference for each system type. Specifically, compared to the rest of the sample, cyclists from the Netherlands and Sweden (i.e. 36% and 17% respectively) expressed statistically higher preferences for each type of system (p<.001). Conversely, cyclists from countries with low national cycling rate (i.e. Italy and Spain, 6% and 3% respectively), exhibited a statistically significant lower willingness to purchase each type of on-bike systems (p<.01). Results are summarized in Figure 3.
Discussion

The aim of the present study was to assess regular cyclists’ preferences for different warning modes of on-bike systems addressing traffic safety. The results show that the passive tag was overall the most preferred system type. This seems to suggest that the regular cyclist expects being unnoticed by the truck drivers in traffic and therefore is more concerned about being noticed than about seeing the truck potentially crossing her or his trajectory. Therefore, it seems that while being seen is perceived as a priority, warning about a potential collision (mostly audio-visual signal) could be considered as complementary safety measure that has the potential to further improve feelings of safety. Additionally, it emerged how participants exhibited tendentially lower preferences for the haptic and the trimodal warning on-bike systems. Conversely to previous study on car drivers who reacted more positively to the tactile warnings in comparison to an auditory and visual system (Scott & Gray, 2008) or who preferred a combination of warnings designed for shorten their reaction times (Dettmann & Bullinger, 2017; Toffetti et al., 2014), in the present study a lower preference could be due to the fact that tactile warning can be confused with vibrations originating with the contact of the wheels with uneven road surface. Furthermore, the poor willingness to purchase the trimodal warning in comparison with the others seems to indicate that it might perceive as an unnecessarily redundant warning mode.

In general, each type of on-bike system was most preferred by younger adults (i.e. under 47 years), than older adults. A plausible explanation could be that older cyclists may show less willingness to purchase an on-bike system due to lower perceived usefulness, as older adults are willing to overcome barriers such as cost or time to learn only if a presented system has obvious personal benefits (Trübswetter & Bengler, 2013). Other feasible explanation might be older adults’ resistance to interact with new and unfamiliar technology (Czaja et al., 2006; Lee, & Coughlin, 2015), or also the fact that older cyclists expected the barriers to be high, possibly because they were not given the opportunity to try the systems in the field.

Differences in preferences between groups defined by cycling frequency were examined. The statistically significant influence of the cycling rate on the general preference seems to suggest that the more one has experience with the bicycle, the more one perceives the potential of the device, being aware of the relative problem of cycling safety showing, in turn, a greater willingness to buy an on-bike system dedicated to increasing comfort and safety. There is evidence that for the acceptance of new technologies familiarity plays a role (Chang, 2010). Indeed, high national cycling level could facilitate the development of a cycling culture, improving overall awareness of cycling safety issues and familiarity with different type of new technological safety device already available on the market. This implies that the possible introduction in the market of the on-bike systems might have more success in countries with higher cycling rates (i.e. the Netherlands, Sweden and Hungary). However, the higher preferences exhibited by cyclists from England, where the average share of cycling as mean mode of transport is poor (i.e. 3% according to the Directorate-General for Communication of the European Commission, 2015) in comparison with cyclists from countries with low cycling level means that other
factors could affect the willingness to purchase an on-bike safety device such as the level of cycling infrastructure as well as accessibility and national cycling safety awareness. Further studies are encouraged to understand the mechanisms explaining these relationships. Within this perspective, the implementation of a technology that can be accepted by different groups of cyclists from different national cycling level should be encouraged to further support designers and decision makers in the field of mobility and cycling safety measures.

This study has limitations which should be recognised. Firstly, the applicability to other segments of the regular cyclist population was limited due to the requirement for e-mail and Internet access. Furthermore, the extension of the findings is limited because the study population is self-selective (i.e. online panel). Finally, the survey data are based on self-reported information and are open to recall bias and reporting errors (Stone et al., 2002). Finally, participants were asked to evaluate the on-bike systems without having experienced them, which may have influenced the results due to a lack of understanding. Further investigations could consider virtual and augmented reality to improve the quality of the experience, but still make it more affordable than real prototyping (Lawson et al., 2016).

Conclusions

The preference of warning modes for on-bike collision avoidance systems was explored in a sample of regular cyclists. The study showed that during the design process of an on-bike assistance system, based on a user centred approach, age bands, cycling frequency and national cycling level could be significant dimensions to consider, in the perspective of bringing the system to the market with a higher probability of acceptance by users.

Acknowledgement

This work was supported by the European Commission under the Grant [number 635975]. Co-founded by the Horizon 2020 Framework Programme of the European Union (2014-2020) under contract number: 635975 Project XCYCLE.

References


