Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference

Human Factors & User Experience in everyday life, medicine, and work

Edited by
Dick de Waard, Karel Brookhuis, Antonella Toffetti, Arjan Stuiver, Clemens Weikert, Denis Coelho, Dietrich Manzey, Berfu Ünal, Stefan Röttger, and Natasha Merat
ISSN 2333-4959 (online)

Please refer to contributions as follows:
[Authors] (2016), [Title]. In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference (pp. pagenumbers). Downloaded from http://hfes-europe.org (ISSN 2333-4959)

Available as open access

Published by HFES

Contents

TRANSPORTATION

MENTAL WORKLOAD

HUMAN MACHINE INTERACTION

Self monitoring – an age-related comparison

Peter Rasche, Matthias Wille, Sabine Theis, Katharina Schäfer, Christopher M. Schlick, & Alexander Mertens
Chair and Institute of Industrial Engineering and Ergonomics of RWTH Aachen, Germany

Abstract

Wearable devices like activity trackers, measuring motion and steps, enable users to monitor their behaviour and might support a healthier lifestyle. These wearables might also motivate elderly to be active and age healthy. Unfortunately these wearables are mostly designed for younger users and it is unclear if they are suitable for older adults, too. In order to answer this research question we conducted a four weeks lasting empirical study about individual motivation and the activity trackers' usability dependent on age groups. During the first session, participants put the activity tracker into operation without further training while a think-aloud-method was applied and mental effort was measured. During the whole study participants used the activity tracker according their personal needs. Usability was measured by PSSUQ at the introduction session, at the middle session after two weeks and at the final session after four weeks. Aspects concerning customer requirements identified during interviews were weighted by a pairwise comparison at the final meeting. Results show a comparison of younger and elder users regarding usability, requirements, motivation, mental effort and technical affinity.

Introduction

One main aspect in terms of healthy aging is daily physical activity and exercise. Performing sports regularly has been proven to reduce the risk of mortality and age-related morbidity and have a common positive effect on older adults' (+65 years) physical and mental wellbeing (Vankipuram, McMahon, & Fleury, 2012). Although all these advantages are well known and promoted by the World Health Organization (2010), more than 70% of older adults are inadequately active (Vankipuram et al., 2012). Thereby industrial societies have to investigate solutions for this problem. Main barrier for elder people to perform physical activity is their lack of information about their individual capabilities and limitations (Hirsch et al., 2000). A solution might be activity trackers which are proven to motivate young user to be more physical active (Clemes, Matchett & Wane, 2008; Steinert, Wegel & Steinhagen-Thiessen, 2015).

Scope of this paper

This paper presents results of a study about an age-related usability investigation of an activity tracker and corresponding daily activity. To identify age-related differences, two age groups ($x < 30$ years; $x > 60$ years) were compared. Aspects explored within this study are users' motivation and intention of use depending on age and whether activity trackers are emphasized to stigmatize a user. Further preferred wearing position is evaluated in relation to age. Positive and negative feedback about the use of the activity tracker within the first four weeks of contact were collected. Finally a pairwise comparison was performed to identify basic requirements for an activity tracker.

Method

This section provides a detailed overview about the method of this four week study under field conditions.

Experimental design

The general task of participants in this study was to use the provided activity tracker and as much as possible integrate it into their daily life. For a most realistic setting all participants were asked to put the activity tracker in operation by themselves without any special training or certain instruction in advance. They were just provided with the standard manual which is sold with the activity tracker. Participants were divided into two equally sized groups by age.

The usability was evaluated according to DIN ISO 20282-2. Therefore, usability tasks were defined which participants should accomplish while “thinking-aloud”. Besides this qualitative observation each usability task was evaluated by the Rating Scale of Mental Effort (RSME) to evaluate the mental effort necessary to accomplish the tasks (Zijlstra, 1993). All participants were familiarized with the RSME scale by appropriate daily live examples. For a more detailed insight into the usability participants fulfilled several Post Study System Usability Questionnaires (PSSUQ) (Sauro & Lewis, 2012) at different points in time: at the beginning, after two weeks and at the end of the study after four weeks. These three moments were chosen to investigate whether the perceived usability improves or at least changes during the first four weeks of contact with the product. The PSSU Questionnaire consists of 11 questions rated by 5-likert-Scale.

Furthermore, participants’ individual technical affinity was evaluated by a questionnaire containing 15 questions (Jay & Willis, 1992). This questionnaire was completed before initial contact with the activity tracker. People with a high technical affinity are supposed to accept and use technical products more frequently and easier than people without (Lee & Coughlin, 2014). The questionnaire was collected a second time at the end of this study after four weeks of activity tracker usage. Aim of this approach was to investigate whether the use of such an innovative product which is worn day and night influences a users’ attitude towards modern information and communication technology (ICT) in general. It might be possible that the users’ attitude changes due to the intensive contact with this product and certain feelings like stigmatization or surveillance.

To evaluate the perceived aesthetics of the activity tracker, stigmatization, wearing position and intention of usage, the MeCue questionnaire was used (Thüring & Mahlke, 2007). As a supplement, a self-developed questionnaire was used which was specially designed for evaluating activity trackers based on the work of Vooijs et al. (2014).

Finally, a pairwise comparison between relevant aspects of an activity tracker was performed in the final session (Pfeifer & Schmitt, 2014). The evaluated aspects were obtained from a market research about sold activity trackers. Aim of this method was to evaluate basic requirements for an activity tracker in the context of the investigated age groups.

Participants

In total 30 participants (14 female, 16 male) with no experience with activity trackers took part in this study. The group of younger users contained 15 participants (6 female, 9 male) recruited at RWTH Aachen University. The 15 older participants (8 female, 7 male) were recruited in a special technical lecture for elderly at RWTH Aachen University.

The age of the older participants reached from 60 years up to 78 years (mean age 68 years, SD = 5.29). All were retired for at least one year and up to 35 years (mean time 12.33 years, SD = 12.64) and used a smartphone in their daily life. 11 out of 15 elder participants described themselves as sporty. Each of these participant performed on average 5.2 h sport (SD = 2.957) per week. For no one sport was a medical requirement. All were interested in measuring their daily activity (measured by a seven point Likert scale (1 = I totally agree; 7 = I totally disagree)) the average answer was 1.71 (SD = 1.139).

The age within the younger group of participants reached from 19 years up to 30 years (mean age 25 years, SD = 2.59). All younger participants were users of a smartphone and 10 out of 15 described themselves as sporty. Within this group the average time of sport was 4.80 h (SD = 4.758) per week. Six out of 15 participants reported sport as a medical requirement in sense of staying healthy and feeling well. The interest in measuring daily activity was lower than in the group of older participants. It was in average 2.53 (SD = 1.356) on the seven point Likert scale.

Experimental apparatus

Prior to this study, different activity trackers were evaluated and compared. As technical features were mostly the same, finally the “ViFit connect Activity Tracker” by Medisana AG was chosen due to higher privacy policy standards. Furthermore, a german-speaking support would have been available to participants, although none of our participants made use of it. The “ViFit connect Activity Tracker” consists of two components: the tracker itself measuring steps and a silicon wristband necessary to position the tracker on the wrist (see Figure 1).

Figure 1. ViFit connect Activity Tracker and Vitadock+ app.

The tracker containing an accelerometer linked with an algorithm is able to measure and count steps as well as to classify user movements during their sleep. A daily activity overview can be displayed on a small liquid crystal display integrated in the activity tracker itself. A more detailed visualization of all recorded data could be viewed within the “Vitadock+” app after the activity tracker is synchronized with the user’s smartphone via Bluetooth 4.0. All trackers were used with anonymous online accounts to provide the highest possible level of privacy.

Experimental procedure

The Ethics Committee at the RWTH Aachen Faculty of Medicine authorized this study and its ethical and legal implications by their statement EK038/15 in February 2015. Afterwards participants were recruited for this study.

Each participant started the four week study individually and had his/her own initial meeting. During the initial meeting a semi structured interview was performed to evaluate the participant’s demographics and their experience with medical products in general. Then participants were asked to fill in the questionnaire to measure their technical affinity, as described in the experimental design. Afterwards, the usability tasks were performed. Therefore, the RSME Scala was introduced to the participant with five examples to familiarize them with this instrument. Then the participants performed certain usability tasks while “think-aloud” method was applied and an interview was executed at the end (see Table 1). The tasks were separated by the evaluation of the RSME.

Table 1. Usability Tasks.

Number	Task
1	Unpack the activity tracker
2	Install the App Vitadock+
3	Charge the activity tracker
4	Identify the battery status symbol
5	Commission the activity tracker by login into the app and synchronization between activity tracker and smartphone
6	Apply activity tracker and measure your steps

Following the usability tasks participants fulfilled the PSSUQ in order to evaluate their first usability impression of this product. Finally, participants were paid an expense allowance of € 15 for this initial meeting and the follow-up meetings were determined. Participants were asked to use activity tracker according their personal preferences during following weeks. Each participant got an activity goal of 10,000 steps a day according the WHO (World Health Organization) suggestion (2010).

After two weeks the second meeting was performed. First, participants fulfilled the PSSUQ for the second time. This was done at the beginning to avoid interference by discussing questions, problems or other topics with the participant. Afterwards a semi-structured interview was performed to record positive and negative aspects experienced in the first two weeks of usage. Afterwards the same usability tasks as in the initial meeting were asked to perform while the RSME was evaluated. Finally, participants were paid a second expense allowance of 15 € for this meeting and the final meeting was determined. Again participants were asked to use activity tracker according their personal preferences, also the activity goal per day kept the same.

At the end of the fourth week the final meeting was held. This one also started with the participant fulfilling the PSSUQ. Afterwards technical affinity questionnaire as well as the semi structured interview from the second meeting was performed again. Then the participants were asked to perform the usability tasks a last time. Hereafter, they fulfilled the MeCue questionnaire and were asked some final questions about their attitude towards this product and its subjective use with in healthcare. In addition, participants accomplished the pairwise comparison of the different wearing positions, functions and requirements an activity tracker might fulfil. Before they left, the last expense allowance of 20 € was paid.

Results

All participants completed the full period of four weeks of activity tracker usage, which was validated by the recorded daily activity. Although this data might have inaccuracies due to the used activity tracker, data processing or the applied wearing position, this data does indicate that the participants regularly wear the tracker during the four weeks (see Figure 2). As an ANOVA revealed there is a significant main effect of the age group on the average activity in weeks two, three and four (see Table 2). As Figure 2 shows the older participant performed more steps a day then the younger ones.

Table 2. ANOVA results for average activity per week.

Week	df	F	p
2	1	7.161	.012
3	1	8.432	.007
4	1	9.686	.004

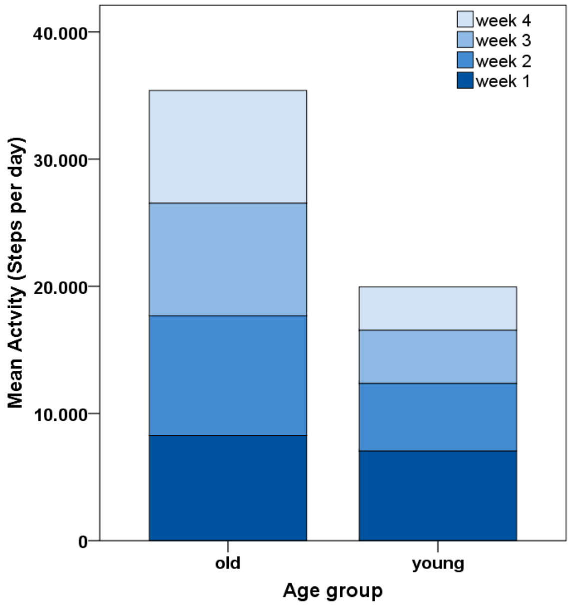


Figure 2. Average activity per week.

The PSSUQ values evaluated at the first (PSSUQ1), second (PSSUQ2) and last meeting (PSSUQ3) indicate no significant difference for the usability evaluation within the four weeks study (see Table 3).

Table 3. Paired t-Tests for PSSUQ.

Pair	old			young		
	t	Df	Sig.	t	Df	Sig.
PSSUQ1 – PSSUQ2	-0.219	14	0.830	-1.402	14	0.183
PSSUQ2 – PSSUQ3	0.904	14	0.381	0.408	14	0.690
PSSUQ1 - PSSUQ3	-0.325	14	0.750	1.281	14	0.221

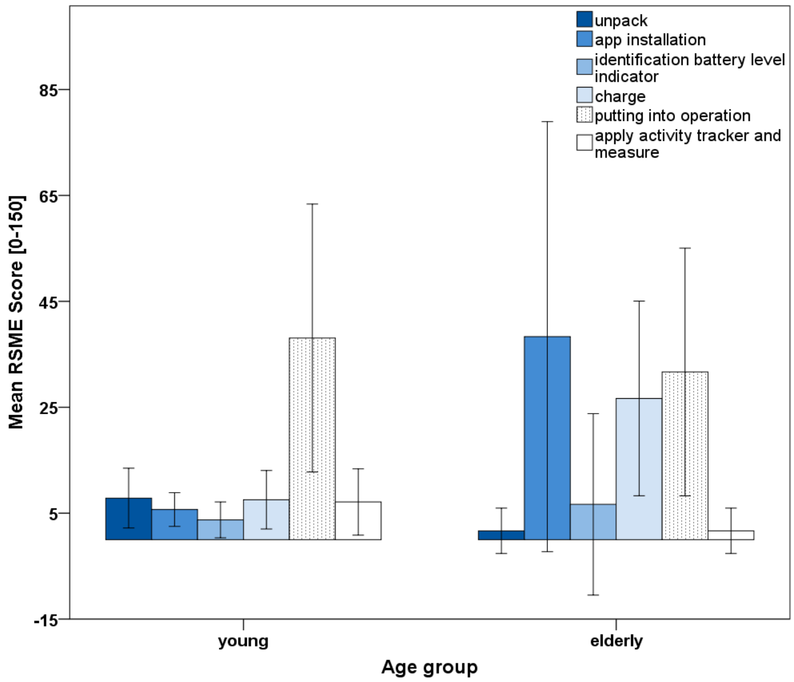


Figure 3. RSME-values for different steps during initial contact. Whiskers represent the 95% confidence interval.

Regarding subjective mental effort some differences can be seen although the mental strain levels are still in the lower third of the overall scale: As shown in figure three older participants reported the highest mental effort during the tasks "2 - app installation", "3 - charge" and "5 - putting into operation" (see Table 1). Within the younger group of participants just the task "5 - putting into operation" was evaluated with a higher mental effort. As a performed ANOVA reveals differs the mental effort for the tasks "app installation" and "charge" significantly due to the investigated age groups (app installation: $df = 1$, $F = 9.511$, $p = .007$; charge: $df = 1$, $F = 10.075$, $p = .006$). The evaluation of the RSME values within the second and third meeting revealed no differences and all participants rated the mental effort for the usability tasks to be zero, which indicates that the activity tracker is easy to use.

Based on the interviews participants felt comfortable using an activity tracker and they were willing to apply this kind of product to their everyday life. 14 out of 15 older participants reported in their last meeting that they would appreciate an activity tracker as therapy support due to motivational aspects (6 out of 14 participants) as well as an objective control (10 out of 14 participants). Within the group of younger participants 12 out of 15 participants reported this, too. Although they were unable to determine detailed reasons for their decision. Also in case of the intention to use the tracker for the next 12 weeks the elderly reported higher intentions. Measured by a seven points Likert scale (1 = I totally agree; 7 = I totally disagree) the average answer for elder participants was 2.57 (SD = 2.138) and for younger participants 3.87 (SD = 2.900).

To investigate if the technical affinity changes by the usage of an activity tracker, the questionnaire scores were compared between different points in time and age groups. But the ANOVA showed no significant differences of the technical affinity between the age groups (start: df = 1, F = .012, p = .912; end: df = 1, F = .817, p = .374). Further showed a t-Test between the two points in time no significant differences (see Table 4).

Table 4. Paired t-Tests for technical affinity.

Age group	Pair	t	df	p
Older	Start - End	1.465	14	.165
Younger	Start - End	-.312	14	.760

The pairwise comparison of wearing position (see Figure 4) shows quite similar preferences of the wearing position between the younger and elder group. A statistical analysis revealed no significant differences. Qualitative analyses revealed for both groups that the wristband is preferred.

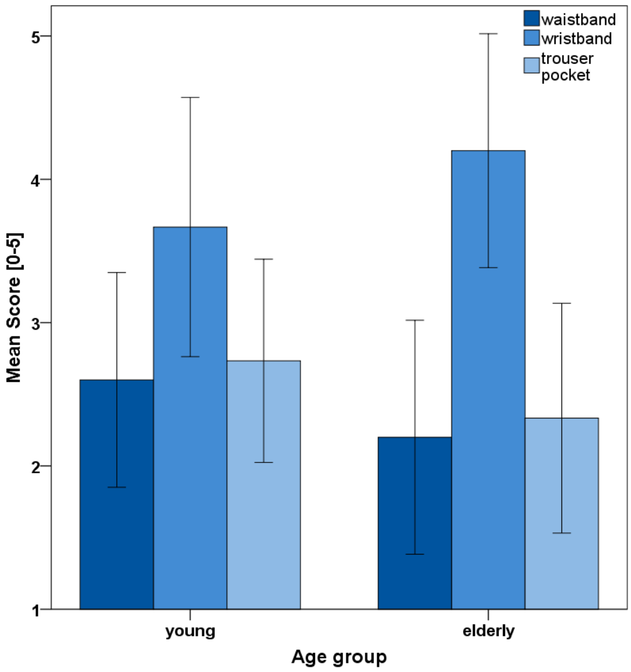


Figure 4. User's attitude regarding wearing position of an activity tracker based on a pairwise comparison. Whiskers represent the 95% confidence interval.

Figure 5 shows the pairwise comparison of basic requirements an activity tracker should fulfil. The Requirements "functional range" and "data accuracy" were rated highest by the older ones. Requirements like "price" or "design" were rated lowest (see Figure 5). The group of younger participants evaluated "functional range" and "comfort" highest. Requirements like "price" and "battery life" were lowest rated by them. Significant differences between the younger and elderly group are revealed by an ANOVA for the aspects "data accuracy" (df = 1, F = 7.383, p = .011) and "design" (df = 1, F = 6.201, p = .019).

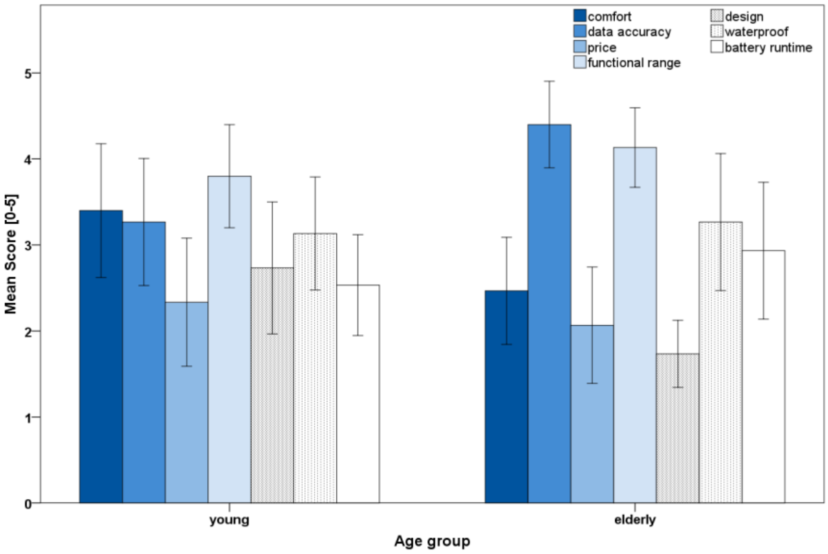


Figure 5. User's attitude regarding basic requirements of an activity tracker based on a pairwise comparison. Whiskers represent the 95% confidence interval.

Figure 6 shows the pairwise comparison for different key functions an activity tracker might have. Also in this case a statistical analysis revealed no significant differences. Therefore qualitative analysis was performed. In total elder participants weighted the functions "measuring pulse" and "counting steps" higher than the other available functions. Younger participants rated the functions "measuring distance" and "measuring sleep" highest.

The semi structures interviews support the results of the pairwise comparison. Participants stated that they prefer to use the activity tracker instead of a wristwatch. Therefore it should be waterproof, so users could wear the tracker in the shower or during swimming or aqua-fitness. A problem occurred with the interaction design: The activity tracker just had one button to interact with. Elder participants reported this interaction design to be difficult and annoying because they had problems to feel the button under the silicon wristband. Younger participants appreciated this interaction design as an easy one. The navigation on the activity tracker display works by pushing the mentioned button, so for example it is necessary to press twice to get the actual time, which was reported to be annoying for the above mentioned reasons, too: Both groups reported that the tracker should always display the time if no certain data like the number of steps or burned calories is accessed. Further participants reported that the provided tracker could only be worn on the left hand. If it was worn on the right hand the button could be accidently pushed by the participants' cloths as well as the data was displayed upside down. A positive aspect reported by all participants was the long battery life of the tracker: If fully charged the tracker could measure steps for up to seven days.

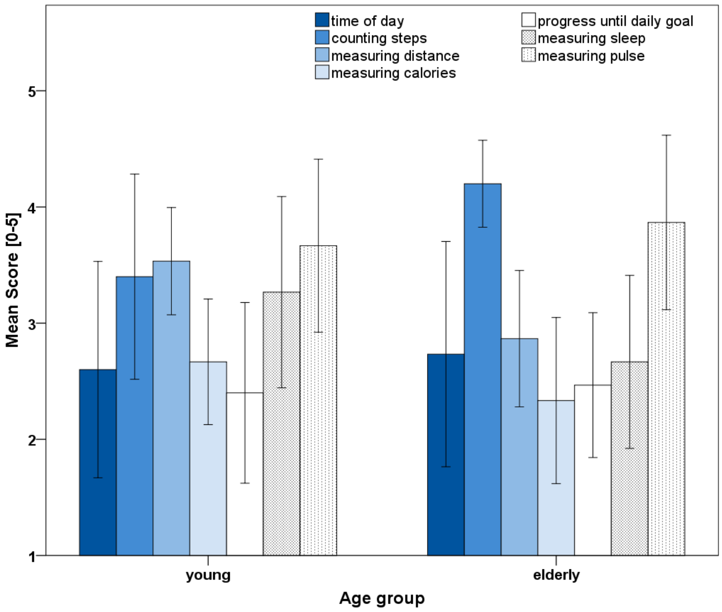


Figure 6. User's attitude regarding key functions of an activity tracker based on a pairwise comparison. Whiskers represent the 95% confidence interval.

Discussion

Our results show a motivating effect of activity tracker and thereby support the results of Clemes et al. (2008) and Steinert et al. (2015) in context of younger users. Further this research does extend these results for the group of older users. Participants reported to be motivated as well as to appreciate activity tracker as motivational and objective support during a therapy as well as daily life. All participants applied the activity tracker. Although the data revealed a higher activity level for elder participants it is unclear if the motivation by the activity tracker or other external circumstances were responsible for this effect (e.g. exams for the younger group at this time).

The results show that activity trackers' usability is commonly suitable for older users. All participants were able to use and put the activity tracker into operation without a certain training or support by the investigator. Nevertheless, usability could be improved: For instance the time should be displayed permanently on the trackers' display, so it could be used as substitution for a wristwatch. Furthermore, a tracker should be universal in case of the hand the activity tracker is worn on. Older users prefer to wear such a product on the left as well as the right hand. Displays as well as the visualization of the displayed information should be suitable for both wearing positions. Furthermore a clip to attach the tracker at the waistband is appreciated by the users to increase the individuality of the wearing position.

Especially women liked to wear the activity tracker in a trouser pocket or even at their ankle especially in cases where they wore jewellery at their wrists.

The results of the evaluated RSME values show that the mental demand differs between the elder and younger participants. Elder participants reported high mental effort for installing the app. Reported main reason for this high mental effort was a lack of training. Similar in both groups is the high mental demand for the task “putting the tracker into operation”. During first initiation process all participants needed to login within the app, with a provided anonymous account. Further they needed to follow the integrated user training within the app. This training was experienced as difficult because the training used full screen images which looked like the real interface. Therefore the participants tried to interact with these images. As they experienced no reaction of the system they started to feel insecure and tried even harder until they understand that this screens are just example images. A solution for this problem might be a proper usability testing of the app before realising it. Furthermore the app did demand a Bluetooth connection which was difficult to handle for most participants. The Bluetooth connection needed to be initialized within the app for security reasons. Most of the younger participants tried to initialize the connection within the standard Bluetooth menu of their smartphone until they realized that the tracker could just be connected via app and is invisible within the standard Bluetooth interface.

The results of the pairwise comparisons revealed a significant difference between the younger and older group regarding the basic requirements “data accuracy” and “design” of an activity tracker. Older participants were more concerned about the data accuracy then the design and for younger participants it was the other way around. Thereby elder users seems to have a more benefit oriented view of an activity tracker, whereby younger once might see them as a product to express their individuality.

One limitation of this study might be that the group of participants was not representative for all elderly users as they are very sporty and active compared to others in their age groups. Furthermore only participants were selected using a smartphone on regular basis. Therefore technical affinity as well as intent of use of an activity tracker might not be representative for elderly in general.

Outlook

Actually, six participants of each group decided to join a long time study and are still using an activity tracker. They are monitored and thereby questions like the product lifecycle as well as the long-time effects (Fritz, Huang, Murphy, & Zimmermann) will be investigated.

Further the ethical and legal aspects in the context of the usage of such products will be investigated more deeply in different workshops with these participants to identify besides usability barriers also more subjective aspects in terms of accepting such products in medical context.

Acknowledgment

This publication is part of the research project “TECH4AGE”, which is funded by the German Federal Ministry of Education and Research (BMBF, Grant No. 16SV7111) supervised by the VDI/VDE Innovation + Technik GmbH.

References

Clemes, S.A., Matchett, N., & Wane, S.L. (2008). Reactivity: an issue for short-term pedometer studies? *British journal of sports medicine, 42*, 68–70.

Fritz, T., Huang, E.M., Murphy, G.C., & Zimmermann, T. Persuasive technology in the real world: a study of long-term use of activity sensing devices for fitness. In M. Jones, P. Palanque, A. Schmidt, and T. Grossman (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 487–496). Toronto, Ontario, Canada: ACM Association for Computing Machinery.

Jay, G.M., & Willis, S.L. (1992). Influence of direct computer experience on older adults' attitudes toward computers. *Journal of Gerontology, 47*, 250-257.

Lee, C., & Coughlin, J.F. (2014). PERSPECTIVE: Older Adults' Adoption of Technology: An Integrated Approach to Identifying Determinants and Barriers. *Journal of Product Innovation Management, 32*, 747-759.

Pfeifer, T. &Schmitt, R. (Eds.). (2014). *Masing Handbuch Qualitätsmanagement* (6., überarbeitete Auflage). München: Hanser, Carl.

Sauro, J., & Lewis, J.R. (2012). *Quantifying the user experience: Practical statistics for user research*. Elsevier.

Steinert, A., Wegel, S., & Steinhagen-Thiessen, E. (2015). Selbst-Monitoring der physischen Aktivität von Senioren. *HeilberufeScience*, *6*, 115-120.

Hirsch, T., Forlizzi, J., Hyder, E., Goetz, J., Kurtz, C. & Stroback, J. (2000). The ELDer project: social, emotional, and environmental factors in the design of eldercare technologies. In *Proceedings on the 2000 conference on Universal Usability* (pp. 72–79). Arlington, Virginia, USA: ACM Association for Computing Machinery.

Thüring, M., & Mahlke, S. (2007). Usability, aesthetics and emotions in human–technology interaction. *International Journal of Psychology, 42*, 253–264.

Vankipuram, M., McMahon, S., & Fleury, J. (2012). ReadySteady: App for Accelerometer-based Activity Monitoring and Wellness-Motivation Feedback System for Older Adults. In *AMIA Annual Symposium Proceedings,Volume 2012, (pp.* 931–939). Chicago, Illinois, USA: AMIA American Medical Informatics Association.

Vooijs M., Alpay L.L., Snoeck-Stroband J.B., Beerthuizen T., Siemonsma P.C., Abbink J.J., Sont J.K., Rövekamp T.A. (2014). Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. *Interactive journal of medical research, 3*, e14.

World Health Organization (2010) *Global recommendations on physical activity for health*. (2010). Geneva, Switzerland.

Zijlstra, F. (1993). *Efficiency in work behaviour. A design approach for modern tools* (PhD thesis). TU Delft, Delft University of Technology, Soesterberg. Retrieved from http://repository.tudelft.nl/assets/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/ide_zijlstra_19931123.PDF [28.09.15]

Needs assessment for individuals ageing with impairment: Findings from subject matter expert interviews

Kimberly C. Preusse[1], Elena T. Gonzalez[2], Jenny L. Singleton[1], Tracy L. Mitzner[1], & Wendy A. Rogers[1]
[1]School of Psychology, Georgia Institute of Technology, Atlanta, GA, USA
[2]Center for Assistive Technology and Environmental Access, Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Adults with sensory and mobility limitations are living longer and thus likely to experience additional declines related to ageing that can negatively influence their independence and quality of life. Worldwide, millions of individuals with vision impairments (i.e., blind or low vision), hearing impairments (i.e., Deaf or hard of hearing,), or mobility impairments (i.e., using a wheelchair) are ageing into the senior population. These individuals who are ageing with a pre-existing impairment comprise a segment of the population that has received very little attention in the human factors community. The goal of this study was to identify the range of challenges in everyday activities for older adults ageing with pre-existing impairments in vision, hearing, and mobility. We present data from interviews with subject matter experts who represented a range of personal and/or professional roles and experiences with older adults with these specific impairments. Findings inform the development of a comprehensive assessment tool with which to interview members of these target groups. In addition, the results have implications for technology design, instruction, and use—ultimately improving technology interactions for this understudied sector of older adults who are ageing with sensory and mobility impairments.

Introduction

Individuals with impairments are living longer lives than ever before, with many ageing into older adulthood. The result is a growing population of older adults subject to complex difficulties that extend beyond what would be expected with normative ageing – the intersection of both having a pre-existing impairment and being above age 65. In the United States, a recent census report revealed that almost 40% of older adults (aged 65+) reported having one or more disabilities (U.S. Census Bureau, 2014), including vision (19%), hearing (40%), and ambulatory (67%) impairments. Human factor interventions have the potential to support

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

independence and successful ageing for older adults with vision, hearing, and mobility impairments.

For writing efficiency, we refer to "older adults with vision, hearing, and mobility *impairments*." We certainly recognize that some communities of such individuals reject the framing of their lifelong differences as "impaired," instead viewing them from a "diversity" or socio-cultural community perspective. For example, individuals with lifelong deafness often do not consider themselves impaired and identify instead as a member of Deaf culture. The way that individuals experience and view their sensory/mobility differences will be important especially as we consider the impact of age-related declines, and the acceptance of human factors interventions to mitigate everyday challenges.

Impairments alone do not necessarily lead to disability. The Person-Environment Fit framework suggests that disability is the result of a mismatch between an individual's functional abilities and the demands of their environment (Lawton & Nahemow, 1973). Age-related losses alone, such as declines in vision, hearing, and mobility, can create issues for older adults in maintaining everyday activities (Rogers et al., 1998). For older adults with long-term impairments, the addition of such declines could present compounding environmental difficulties and increased risk of disability (Harrington et al., 2015).

Successful ageing involves three goals: avoid disease and disability; maintain physical and cognitive function; and remain actively engaged in life (Rowe & Kahn, 1997). Goals for people ageing with pre-existing impairments might also include reducing age-related chronic disease, promoting independence, and social participation (LePlante, 2014; Minkler & Fadem, 2002). In addition, the majority of older adults would prefer to age in place, to maintain an independent lifestyle in the home of their choice (AARP, 2010). For many, lack of home accessibility and access to health and community resources inhibits their ability to do so. There is a need for research to identify and address the home and community barriers experienced by people ageing with pre-existing vision, hearing, and mobility impairments.

Technology can provide supportive solutions as individuals navigate age-related declines. Yet, complex issues with technology use and acceptance can arise as both the user and their familiar assistive technologies age (Agree, 2014). An appropriately designed and implemented assistive technology has the potential to alleviate or reduce many challenges and improve the quality of life for people with impairments (Kemp, 1999). However, that technology should also be designed to accommodate the user's additional declines brought on by ageing. For example, older adults with pre-existing impairments are can experience an interaction of sensory and physical impairments that can interfere with their ability to use a familiar device (Harrington et al., 2015).

Little attention has been paid to understanding the additional challenges ageing presents to individuals with a long-term impairment. Interviewing subject matter experts (SMEs) has been demonstrated to be a useful method for acquiring rich information in a variety of domains (e.g., Cullen et al., 2012; Lindgaard et al., 2006;

McBride et al., 2011), and is especially useful at the needs assessment stage. The present study features interviews with SMEs with extensive knowledge about and experience with older adults with vision, hearing, and mobility impairments. SMEs could speak to the range of issues because they interacted with multiple older adults in their respective populations. Through their perspectives, our goal was to identify the range of challenges experienced by these individuals. Specific emphasis was identifying challenges with technology, access to community resources, and housing.

Method

Participants

Interviews from nine SMEs are presented here. They were strategically recruited to represent different areas of expertise and perspectives, three in each group (i.e., vision, hearing, mobility; see Table 1). Each SME had personal and/or professional experience working with older adults in these groups and and were recruited through outreach to relevant ageing and disability organizations and professional societies.

Materials

A semi-structured interview script was developed to explore SME perspectives on challenges in the areas of technology; community and health resources; and ageing in place for individuals ageing with sensory or mobility impairments. Scripts are available from the authors upon request.

The interview script was divided into two main parts: SME background and perceived challenges among older adults with vision, hearing, and mobility impairments. In the first part, participants were asked about their relevant experiences with their respective SME population. The second part assessed their experiences and knowledge of the challenges of the older adults in four domains: (1) technology use and attitudes towards technology; (2) access to community and health resources; (3) housing; and (4) other challenges not encompassed by the first three domains.

Procedure

Three researchers conducted SME interviews for this study; each researcher was assigned to an impairment group and conducted all three interviews for that group. Each interview lasted 45 to 60 minutes; they were recorded and later transcribed.

Thematic Analysis

Transcripts were reviewed twice for emergent themes, first by the interviewer for that group of transcripts, and second by a different interviewer. The four categories of the original interview script were used for the initial organization of emergent themes (see Table 2). The transcript reviewers then conferred and reached agreement on the most salient and prevalent themes within and across transcripts.

Table 1. Description of Subject Matter Expert (SME) Participants

Group	*Participant*	*Description of Expertise and Relevant Experiences*
Vision	V1	- Has a visual impairment - Has a parent with a visual impairment - Instructor for computer classes for older adults with visual impairments - Has experience working as a blind/low vision rehabilitation engineer, usability consultant, and researcher
	V2	- Provides care for older adult mother, who is blind (age-related macular degeneration) - Occasionally provides care for another older adult with vision impairment - Licensed certified nursing assistant (CNA)
	V3	- Has a visual impairment (blind) - Leads workshops for people aging with visual impairments for living communities and community centres - Worked as a rehabilitation counsellor for people with visual impairments (typically 55+) - Provided care for legally blind wife (now deceased)
Hearing	H1	- Child of two Deaf parents - Provided care for her older Deaf mother (recently deceased) - Business executive in healthcare industry
	H2	- Provides care for older Deaf father - Sign language interpreter
	H3	- Organizes Deaf senior groups at a Deaf community social service agency - Provides care for older Deaf father
Mobility	M1	- Occupational therapist in a geriatric hospital - Owner of a home accessibility consulting business that helps older adults and their families identify accessibility solutions (e.g., home modifications) that can help them remain at home - Certified aging in place specialist
	M2	- Professional caregiver; working long-term with an older adult with Spinal Cord Injury (SCI) - Family caregiver for her older adult husband who has a mobility impairment - Advocacy Specialist for people with spinal cord and brain injuries
	M3	- Geriatric care manager at a private aging care management service company specializing in senior advocacy, needs analysis, and care coordination - Geriatric social worker in skilled nursing facility-Geriatric Care Manager at a private aging care management service company specializing in senior advocacy, needs analysis, and care coordination - Geriatric Social Worker in skilled nursing facility

Results

Participants

The nine SMEs represented a range of backgrounds. Most worked directly with older adults with vision, hearing, or mobility impairments (e.g., rehabilitation counsellor, sign-language interpreter, occupational therapist). Many also worked as a volunteer or advocate on behalf of those individuals (e.g., instructor for computer classes for older adults with vision impairments, organizer of Deaf senior community events, advocacy specialist for individuals with spinal cord or brain injuries). In terms of personal experiences, some SMEs had the specific sensory or mobility impairment themselves, and some were caregivers for their ageing parent/spouse who had a sensory or mobility impairment. All were well informed in their respective areas and were able to provide rich accounts of the challenges faced by older adults ageing with sensory or mobility impairments.

Technology use and acceptance

Table 2 provides the themes of challenges for older adults with vision, hearing, and mobility impairments. For each domain, the themes were classified as being a crosscutting across impairment groups or unique to older adults with a specific impairment. One crosscutting challenge was the size of technology, particularly concerning cell phones and smart phones (e.g., small keyboards). Similarly, complex interfaces and functions of devices were a shared challenge. SMEs suggested the need for simple, streamlined devices that are easy for older adults to understand and operate. Lack of technology training and support was another challenge that recurred throughout the interviews. As H3 explained "*They get trained one way, and then [it] changes six months later. It can be challenging.*"

Some themes regarding technology use and acceptance were specific to older adults with certain impairments. For example, screen readers were discussed as very difficult for older adults with vision impairments; screen readers have a high learning curve, more functions than desired by most users, and are taxing on memory. SMEs noted that assistive technologies can draw unwanted attention, but that this was alleviated as assistive technologies become more typical in appearance or are integrated into everyday technologies (e.g., iPads). SMEs discussed the benefits of voice recognition software but also discussedconcerns about technologies used in financial transactions.

For Deaf older adults, SMEs explained that technology has a time and a place. For instance, while video remote interpreting (VRI: a videoconferencing technology that brings a remote sign language interpreter into the local situation via laptop and webcam) is useful in most situations, it might not be appropriate in high stress situations, such as the emergency room where *"they're worked up, and they just want a body in front of them to interpret.*" SMEs also reported that many technologies (e.g., medication reminders, phones) rely on audio-based cues and do not have an adequate alert system for people who are Deaf. Thus, although technologies were being adopted, additional considerations could make them more useful and usable.

Technology challenges among older adults with mobility impairment were primarily related to privacy concerns. SMEs described how the threat of identity theft and scams discouraged people from using the Internet. Being monitored too closely was another perceived threat among older adults with mobility impairment, especially with many health-monitoring technologies (e.g., activity trackers, medication monitoring systems) imposed on an individual by caregivers and family members. Fear of being spied on and not fully understanding the technology can make people distrust these technologies ("*people don't like to feel like they're under watch*").

Perhaps this concern arose for mobility SMEs because technologies that monitor movement between rooms, and in case of falls notifies help, are likely more prevalent amongst those with mobility impairments. The technologies were typically described as often being designed for the caregiver instead of the user.

Table 2. Themes from Subject Matter Expert Interviews: Challenges Among Older Adults with Vision, Hearing, or Mobility Impairments

	Crosscutting Challenges	*Challenges for Older Adults with Specific Impairments*		
	Vision, Hearing, & Mobility	***Vision***	***Hearing***	***Mobility***
Technology Use and Acceptance	- Need for simple, streamlined technologies that are more intuitive	- Assistive technologies for people who are blind/low vision can be large and attract unwanted attention	- Use of video remote interpreting can be inappropriate in high stress situations and is inadequate when the camera/video phone is not in the needed location	- Privacy concerns about using the internet (e.g., identity theft)
	- Size of technologies often too small, creating usability issues	- Screen readers can be difficult for older adults to use; provide an overwhelming amount of information and functions	- Difficult to keep up with frequent changes and updates to technology	- Health technologies often designed for the caregiver instead of the user
	- Technology training and support is insufficient	- Touchscreens without buttons can present issues for people who are blind/low vision because they often rely on tactile cues to make associations; voice recognition software may be a beneficial alternative	- Need for technologies to provide adequate alerts for individuals who are Deaf/hard of hearing, (e.g., vibration, lights)	- Resistance toward sensor/monitoring technologies; concerns about privacy and family members or caregivers spying
	- Older generations tend to be more fearful and resistant to technologies	- Financial transactions can be challenging; difficult to know if they are paying or being charged the correct amount		

Table 2 (continued). Themes from Subject Matter Expert Interviews: Challenges Among Older Adults with Vision, Hearing, or Mobility Impairments

	Crosscutting Challenges	*Challenges for Older Adults with Specific Impairments*		
	Vision, Hearing, & Mobility	***Vision***	***Hearing***	***Mobility***
Access to Community and Health Resources	- Lack of accessible public transportation; finding and utilizing it requires substantial planning	- Privacy concern in releasing personal information to community resources	- Lack of interpreter services in medical settings; emergency services may be unable to communicate with Deaf older adults	- Lack of awareness about aging and disability resources; many older adults use phone books instead of internet resources
	- Private transportation options are expensive and often unaffordable	- Lack of accessible exercise options	- Need to make events accessible by providing an interpreter	- Difficult to visit doctors and healthcare providers; opportunity for telehealth services with for both patients and staff
	- Resources and services tend to be in major cities; need for satellite resources in suburban and rural areas	- Many people fall into the gap of having just enough money where they do not qualify for aid, but are unable to afford services themselves		- Many people fall into the gap of having just enough money where they do not qualify for aid, but are unable to afford services themselves
				- Lack of accessible exercise options
Housing	- Sensory and mobility impairments are misunderstood by nursing home staff	- Concerns about kitchen safety (e.g., cuts, burns); need for accessible appliances to enable independence	- Limited access to wireless technology needed for communication devices (e.g., in assisted living settings)	- Concerns about telling family members about falls, out of fear they will have to leave their home
	- Affordability of home modifications, technologies, and services are a major barrier to aging in place		- Few homecare providers or caregivers in assisted living facilities know sign language	- More likely to need extensive support with basic activities of daily living (e.g., bathing, toileting), especially if they need assistance with transfers

Accesss to community and health resources

The lack of accessible public transportation was a crosscutting concern across groups. SMEs commented that finding and utilizing accessible public transportation required substantial planning. Para transit options, such as those offering curb to curb service, often required several days' notice. Mass transit options were more readily available but impose a high cognitive load; in addition, issues with walking abilities and fear of being lost were common. Private transportation options, such as taxis, were described as expensive alternatives with ongoing concerns about accessibility (e.g., ease of summoning, risk of injury, complexity of financial transactions). The cost of private transportation is even more profound with wheelchair users in purchasing and maintaining an accessible van. Transportation challenges were discussed as having broad and serious impacts, particularly on one's ability to be an active part in the community and to stay healthy.

Another cross-cutting challenge was that community services and resources – particularly those intended for individuals with vision or mobility impairments – were primarily located in downtown areas of major cities, requiring transportation to reach them. Both vision and mobility SMEs stressed the need for satellite resource centres in suburban and rural areas. In terms of exercise and recreation, SMEs also commented on the lack of exercise options; gym facilities and exercise classes were generally considered inaccessible and inappropriate for older adults with vision or mobility impairment: "*If you're older, there might be an ageing exercise course, but if you're older and you have a vision impairment, there's zero.*" This concern was echoed for older adults with mobility impairments, whose physical capabilities vary widely and for whom exercise would need to be appropriately modified. Across community and health resources, there is a need for better information dissemination for older adults, especially for those with mobility impairments. Many did not know where to look for information about cost, available equipment, and training. As one explained, "*unless you have an advocate for you, you're pretty much on your own*."

Housing

Most of the challenges elicited involved a need to modify, outfit, or bring services to the home to enable a certain task. The ability to complete kitchen-centred activities independently and safely (i.e., without cuts or burns) was addressed as a challenge for older adults with vision impairments. Accessible appliances that provide audio or tactile cues were mentioned as useful, but expensive options. Having assistive technologies throughout the home was also noted as a need among Deaf older adults. Video relay service (VRS) phones are typically only installed in one room of the house, which can be inconvenient and potentially dangerous for Deaf older adults (i.e., if an emergency arises and the person is in a different room). Finding caregivers for Deaf older adults was described as particularly difficult because most in-home caregivers typically do not know sign language.

The SMEs affirmed that ageing in place is the overwhelming preference of older adults, but challenges remain. Older adults with mobility impairments tend to need extensive home support, mostly with Activities of Daily Living (ADLs). Older adults may not tell caregivers or family members about falls or injuries in the home out of fear they would be forced to move to long-term care residences.

When asked about challenges in finding assisted living options, SMEs from both the vision and hearing, groups reported a lack of understanding among staff about sensory impairments. One SME discussed how staff at her mother's assisted living community did not seem to understand the difference between deafness and age-related hearing loss. Thus, employees would scream to try to get her Deaf mother to hear them. Another spoke of a similar misunderstanding, and potential stigmatization, wherein assisted living facilities assigned older adults with visual impairments into higher levels of care than needed. In both cases, SMEs felt broader education for care providers was necessary. The hearing-specific SMEs also described limited access to computers and videoconferencing technologies because of slow adoption in long-term care residences, which in turn could limit access to health and community information as well as result in increased social isolation among residents. In sum, the SMEs did not think that most long-term care residences were ideal (or even satisfactory) communities for older adults with sensory or mobility impairments.

Other: Psyco-social barriers

The interviews yielded a category of challenges not captured by the pre-specified domains; namely, psycho-social barriers. Several SMEs discussed how older adults with impairments are at high risk for isolation and depression. SMEs explained that older adults might self-impose psychosocial barriers (e.g., choosing not to socialize as frequently) or be isolated because of barriers in their environment. For example, one explained that her father's nursing home had group events for residents, but he would not go because they would not hire an interpreter: *"He can't communicate with the people around him."* SMEs suggested that psycho-social barriers are pervasive across all the domains, such as transportation challenges contributing to social isolation or lack of exercise opportunities exacerbating depression.

Discussion

A thematic analysis of the nine SME interviews identified a range of everyday needs for technology, community resources, and housing for individuals aging with vision, hearing, or mobility impairments. Many themes were shared across groups, such as a need for larger technologies; better technology training and support; more accessible and affordable transportation options; a greater understanding of resident needs at assisted living residences; and the potential increase in psychosocial concerns that may result from experiencing barriers in these domains. Nonetheless, some themes were specific to a particular group, with SMEs of older adults with vision impairments describing a need for user-friendly screen readers, exercise options, and privacy concerns. Deaf older adult experts described a need for appropriate use of remote-video compared to in-person sign language interpreters in healthcare, and a general need for sign language accessibility for community resources and events. SMEs for older adults with mobility impairments detailed a need for easy-to-use technologies, resource information dissemination, and a better method for sharing personal information with their children that allows them to retain some privacy.

Some of the broad challenges revealed in this study aligned with the general ageing literature, but the results revealed details for individuals who are ageing with a pre-existing impairment. For example, a Deaf individual may have not used a medication reminder system prior to ageing. However, due to increased medication use or memory changes with ageing, such a system may now be needed but typical audio alerting tools would not suffice. As another example, technology training is established as a need for older adults (Czaja & Sharit, 2012; Mitzner et al., 2008). However, for Deaf older adults, there is the additional challenge of communication accessibility of such trainings (i.e., needing a sign language interpreter). An understanding of these barriers is the primary contribution of this study.

Similarly, transportation challenges are commonly reported in the ageing literature, brought upon by older adults having to limit or stop driving (Anstey et al., 2005). However, this difficulty is exacerbated for older adults with vision, hearing, or mobility impairments. For instance older adults with vision impairments not only need to rely on others or a service for getting places, but also may be experiencing age-related declines in balance or mobility which could create challenges for getting out of a vehicle. Moreover, they need to be able to identify the location and height of a curb, to prevent a misstep and risk not recovering their balance. Similarly, with age, walking speed can decline, requiring more extensive planning and travel time for older adults with pre-existing vision, hearing, and mobility impairments, many of whom must already arrange accessible transportation days in advance. In all cases, private transportation options are typically too expensive for regular use. If these concerns are echoed in the European population of older adults with sensory and mobility impairments, interventions that alleviate transportation issues may best be piloted in Europe, where the public transportation system as a whole is well-developed.

Implications for design

Our results highlight a variety of barriers faced by older adults with sensory and mobility impairments. With respect to technology, these findings suggest the need for participation of SMEs as well as individuals with sensory and mobility impairments in technology design. By including user groups who have a diverse set of capabilities and limitations early in the design process, technologies are likely to be more usable for all users. Indeed, several challenges identified in this study were cross-cutting among the vision, hearing, and mobility impairment groups, suggesting that designers should consider the interaction of multiple impairments, which can create compounding difficulties for a user. Taken together, these findings further support the potential value of universal design, in which environments and products are designed to be usable regardless of the age or limitations of the user (Sanford & Stark, 2014). Thus, human factors considerations should not just include ageing *or* sensory/mobility impairments, but also those who are *both* ageing and have a sensory or mobility impairment. As a general guideline, technologies should be developed such that the loss of one sense will not disable the user from utilizing a beneficial technology. For example, technology training should be available through multiple media (not just user manuals, but also audio support for individuals with visual impairments and help-lines should have VRS access). Similarly, small, flat buttons that are difficult for individuals with vision impairments to differentiate, might also be difficult for individuals with limited motor control to accurately press. Alternative inputs could help with this issue (as voice commands have helped older adults with vision impairments use communication technology).

For individuals ageing with an impairment, universal design could improve their ability to fully participate in the community (Sanford, 2012). The design of public transportation, buildings, and services should consider and accommodate needs of a wide range of users. By determining the challenges faced by older adults with sensory and mobility impairments, interventions can be developed. For example, equipping taxis with money readers may alleviate payment concerns for older adults with visual impairments. In other cases, it may simply be the laggard adoption of people who are not older adults with impairments that can be changed. For example, if assisted living facilitates adopted VRI and computers for residents in the same way they adopted board games and telephones, information and communication could be much more accessible to Deaf older adults. In such cases, interventions would require the involvement of a greater range of stakeholders. When designing universally or specifically for older adults with impairments similar interventions could ideally be implemented based on a taxonomy of everyday support needs for older adults with sensory and mobility impairments. A taxonomy could also guide the development of technologies and systems not just for older adults with impairments, but for older adults in general, who may later experience age-related sensory and mobility declines.

Limitations and future directions

As in all research studies, limitations must be considered in interpreting the data. First, there is a need to recruit SMEs from a greater range of experiences, particularly professional caregivers in assisted living communities. Such individuals are very difficult to recruit, perhaps because they have a heavy workload and high turn over, and possibly because those who care for multiple older adults with a sensory or mobility impairment are rare. In either case, given the difficulty the researchers had in finding such individuals, it follows that it may also be difficult for those seeking care to find those individuals. Additionally, these data revealed that not all challenges faced by older adults with sensory or mobility impairments are barriers to functional needs. Social challenges and psychosocial challenges are also substantial. Further, technological advances may modify some of the challenges reported here. For example, as VRI and voice-activation input become more standard in technology, some challenges reported by SMEs might decrease. However, if VRI and voice-activation continue to become popular, but not procedurally standard across the technologies that use them, training may become a greater burden with users having to distinguish how to access the feature on one product versus another product. Likewise, the increased use of flat buttons might make technologies less accessible for users with vision and mobility impairments. Additional investigation into the specifics of challenges and how older adults with sensory and mobility impairments respond to them is necessary to provide additional human factors guidance.

Strengths of the study

The design and results of this study provide useful guidecance. First, the results demonstrate the need for involvement of older adults with sensory and physical impairments, as well as their advocates, caregivers, family members, and relevant professionals in the design process. Knowledge gained from SMEs can be used to guide further investigation with target user groups. The SME interviews allowed for a comparison across three common impairments (vision, hearing, mobility), while still capturing a broad range of experiences. SMEs will be vital points of contact for future work with user groups because they can facilitate the recruitment of a diverse sample and can help to assure potential participants of the credibility of the researchers.

Conclusion

The goal of this study was to identify the range of challenges experienced in everyday activities for older adults ageing with pre-existing impairments in vision, hearing, or mobility. Interviews with SMEs with personal and/or professional experience with these individuals revealed challenges that were unique to each population as well as shared across impairment groups. By better understanding these challenges, the human factors community, alongside key stakeholders (e.g., physicians, community groups), can jointly provide solutions that promote successful ageing for individuals with sensory and mobility impairments.

Acknowledgements

This research was supported in part by a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (Department of Health & Human Services, Administration for Community Living) Grant 90RE5016-01-00 under the auspices of the Rehabilitation and Engineering Research Center on Technologies to Support Successful Ageing with Disability (TechSAge; www.techsage.gatech.edu). The contents of this publication were developed under a grant from the Department of Health & Human Services, Administration for Community Living. However, those contents do not necessarily represent the policy of the Department of Health & Human Services, Administration for Community Living, and you should not assume endorsement by the Federal Government. We also thank Jaudale Banks, Kevaghn Hicks, Yeiryun Park, Mimi Phan, and Alessandra Zhang, for assistance with the transcriptions.

References

AARP. (2010). *American Community Survey Reports: Home and community preferences of the 45+ population*. Retrieved from http://www.aarp.org/livable-communities/learn/research-trends/info-12-2012/aarp-home-community-preferences.html

Agree, E. (2014). The potential for technology to enhance independence for those ageing with a disability. *Disability and Health Journal, 7*, 33-39.

Anstey, K.J., Wood, J., Lord, S., & Walker, J.G. (2005). Cognitive, sensory and physical factors enabling driving safety in older adults. *Clinical Psychology Review*, *25*, 5-65.

Cullen, R.H., Smarr, C.A., Serrano Baquero, D., McBride, S.E., Beer, J.M., & Rogers, W.A. (2012). The smooth (tractor) operator: Outcomes of a knowledge engineering project. *Applied Ergonomics, 43,* 1122-1130.

Czaja, S.J., & Sharit, J. (2012). *Designing Training and Instructional Programs for Older Adults.* Boca Raton, FL: CRC Press.

Harrington, C.N., Mitzner, T.L., & Rogers, W.A. (2015). Understanding the role of technology for meeting the support needs of older adults in the USA with functional limitations. *Gerontechnology, 14*, 21-31.

Kemp, B. (1999). Quality of life while ageing with disability. *Assistive Technology, 11*, 158-163.

Lawton, M.P., & Nahemow, L. (1973) Ecology and the ageing process. In C. Eisdorfer & M.P. Lawton (Eds.), *Psychology of Adult Development and Ageing* (pp.619-674). Washington, D.C.: American Psychological Association.

LePlante, M.P. (2014). Key goals and indicators for successful ageing of adults with early onset disability. *Disability and Health Journal, 7,* 44-50.

Lindgaard, G., Dillion, R., Trbovich, P., White, R. Fernandes, G., Lundahl, S., & Pinnameni, A. (2006). User needs analysis and requirements engineering: Theory and practice. *Interacting with Computers 18*, 47-70.

McBride, S.E., Beer, J.M., Mitzner, T.L., & Rogers, W.A. (2011). Challenges for home health care providers: A needs assessment. *Physical and Occupational Therapy in Geriatrics, 29*, 5-22.

Minkler, M., and Fadem, P. (2002). "Successful Ageing": A disability perspective. *Journal of Disability Policy Studies, 12,* 229-235.

Mitzner, T.L., Fausset, C.B., Boron, J.B., Adams, A.E., Dijkstra, K., Lee, C.C., Rogers, W.A. & Fisk, A.D. (2008). Older adults' training preferences for learning to use technology. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *Volume 52* (pp. 2047-2051). Santa Monica, CA: Sage Publications.

Rogers, W.A., Meyer, B., Walker, N., & Fisk, A.D. (1998). Functional limitations to daily living tasks in the aged: A focus group analysis. *Human Factors, 40*, 111-125.

Rowe, J. & Kahn, R. (1997). Successful ageing. *The Gerontologist, 37,* 433-440.

Sanford, J.A. (2012). *Universal design as a rehabilitation strategy: Design for the ages*. New York: Springer.

Sanford, J.A. & Stark, S. (2014). Universal design as a workplace accommodation strategy. In I. Söderback (Ed.), *International Handbook of Occupational Therapy Interventions Second Edition (*pp. 363-379), New York: Springer Publishing.

U.S. Census Bureau. (2014). *Older Americans with a disability: 2008-2012.* Retrieved from https://www.census.gov/content/dam/Census/library/publications/2014/acs/acs-29.pdf

Shared Usability - a support mechanism to product and service system design for older adults

Linda Shore, PJ White, & Hilary Dempsey,
designCORE, Institute of Technology, Carlow, Ireland

Abstract

Globally the Older Adult population is increasing; people are living longer, often with physical or functional limitations whilst remaining in their own home. This indicates a requirement of responsibility by Associated Stakeholders to support ageing in place. The concept of shared usability proposes that Older Adults can maintain independence, choice and empowerment, with mutual agreed levels of support from Associated Stakeholders when using products or services. Research was conducted as a means to identify and explore shared usability in the context of a user centred design process. Qualitative research methods with an ethnographic approach were conducted over a nine-month period. The fieldwork involved observing and understanding everyday life for the Older Adult in their own home, with specific enquiry and task observation of eight areas. Design methodologies of ideation, sketching and iterative sketch models were applied in order to select one specific area for design conceptualisation. Further to this, brainstorming sessions involving participants using storyboard and feedback were used to evaluate proposed concepts. The product concept outcome highlights how product and service systems can be developed with inclusion of shared usability. The fieldwork offers recorded and detailed enquiry of the experience of ageing. Finally, a definition of shared usability is proposed as a tangible consideration during the process of design that facilitates the user being supported by a network of Associated Stakeholders.

Introduction

People are not only living longer, but often living longer and independent with some functional limitation. The growing ageing population directs a need for designers to engage with research specific to Older Adults & ageing. The intention of design research must be to improve and endorse the choice and autonomy older adults deserve when using products or services.

This paper discusses in three parts, the approach and delivery of conceptual outcome using shared usability as a mechanism that offers mutually agreed levels of usability between a user and associated stakeholders when implemented as part of a user centred design method.

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Part One – Literature review: The initial enquiry was to understand and record quantitative data that displayed areas of relevance to understanding the context of research. This data was critical to identifying and understanding the scope and limitations of the research. There were numerous supportive documents published by Global and research agents researched in order to develop the areas of enquiry for fieldwork. (i.e. United Nations, European Commission, International Organisation for Standardisation, Centre for Ageing Research and Development in Ireland) It reviews the ethical considerations and concludes by sharing the strategy to prepare for fieldwork and the recruitment of Older Adult participants. Concluding with the eight areas of enquiry selected for fieldwork

Part Two – Fieldwork Methodologies: Part two discusses the methodologies that were used, and highlights design ethnography as the research method selected. It details the fieldwork sessions undertaken during life-logging and task observation sessions.

Part Three – Research outcomes: The conceptual phases of design to product and service system concept outcomes of the research are discussed. As a result, the concept of the 'Smart*Share* System' was created. This concept promotes and highlights how a User (the Older Adult) will select the levels of engagement they have in managing heating and fuel efficiency in their home. The paper concludes by displaying an infographic that highlights the journey of this research through to the research outcomes.

Part One – Literature review

This research is a progression of findings from previous design research titled: "Designer as Ethnographer: A Study of Domestic Cooking and Heating Product Design for Older Adults" (White, PJ. 2012). White highlighted the potential for shared usability' as a supportive method of intervention between Older Adults and Associated Stakeholders when using products or services.

The broad intent of this research was to identify unmet product and service needs within the day to day lives of Irish Older Adult participants and integrate methods to a user centred design process. There was a need to define the research methods to be conducted, for example, what areas of day to day life of Older Adults that would be researched within their home environment. Literature reviewing assisted understanding of the limitations people have in living at home in later life. Ageing can present a decline in sensory function, mobility, balance and memory and therefore impact on our ability to remain independent, (Farage, Miller et al. 2012).

The Madrid plan of action on ageing states the requirement to support the desire an Older Adult has to age in a home of their selection and type (United Nations, 2002). These factors, combined with reduced fertility and birth rates determined the requirement to explore the viability of shared usability' for Older Adults. The International Classification of Functioning, Disability, and Health (ICF – World Health Organisation; 2001) offered a frame of reference to understand this (see Figure 1). This classification gauges' individual's health or disability in context to their environment or ability. It offered support and guidance to the research by stating definitions and limitations to activities and experience a person may have throughout life. The classification is outlined in the 5 points as follows:

1. Activity: the execution of a task or action by an individual.
2. Participation: involvement in a life situation.
3. Activity limitations are difficulties an individual may have in executing activities.
4. Participation restrictions are problems that an individual may experience in life.
5. Environmental and Personal factors make up the physical, social and attitudinal aspects of the user. Defining the areas to observe the day to day life for Older Adults was identified further by the Information matrix published by World Health Organisation.

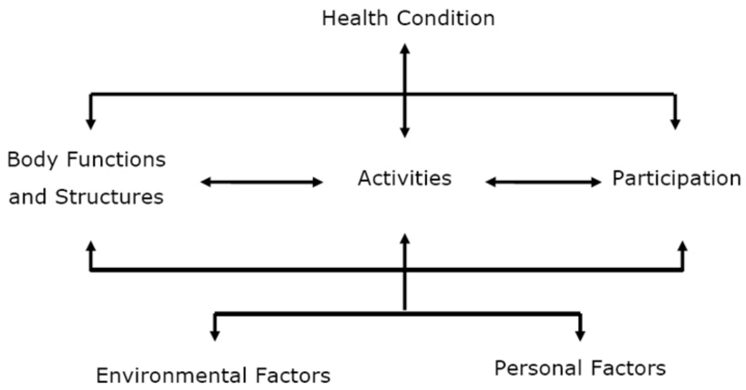


Figure 1. Framework for ICF - WHO, 2001.

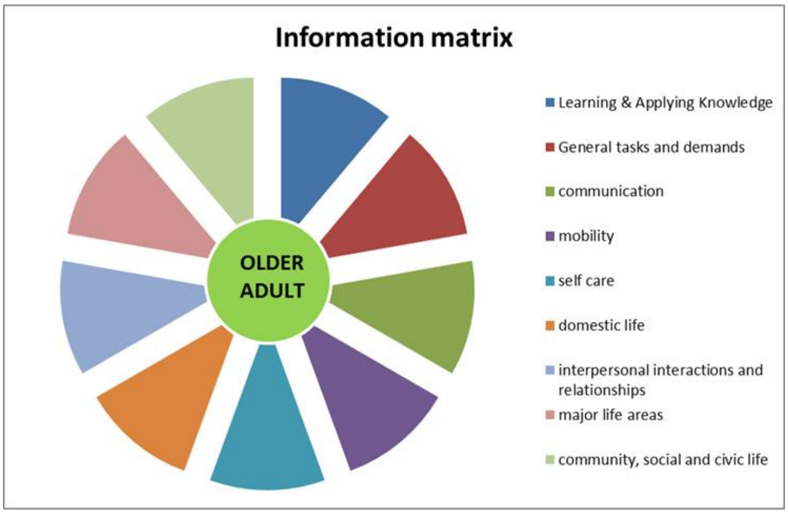


Figure 2. Information matrix as listed per ICF 2001 - WHO as interpreted visually by Author.

In addition to the International Classification of Functioning, Disability, and Health (WHO; 2001), the Information matrix (Figure 2) offered classification guidelines to consider the human factors need for this design research. It highlighted the potential enquiry areas associated with Activities and Participation, and how these can relate to contextual needs of the environment and person. This would assist the developing of the enquiry template that would be used as a memo tool during the fieldwork.

Finally, Parker and Thorslund's study of disabled elderly people in Sweden further assisted with defining areas of enquiry. It discussed the use of technical aids as a facilitator to ageing independently (Parker & Thorslund., 1991). Figure 3 was designed by the researcher as a means of interpreting the requirement needs of fieldwork for this project and was adapted as per the areas of enquiry conducted by Parker and Thorslund.

Figure 3. Adapted from 'The use of technical aids among community based elderly' Parker,M.G; Thorslund, M; 1991.

The first six areas to explore in fieldwork were seen as direct activity, and necessary to function independently:

- Dressing
- Bathing & Toileting
- Cooking
- Communication, TV &radio
- Mobility
- Access

The remaining two areas were considered more selective to choices and priorities people place in their day to day activities:

- Interests & Activities
- Physical Functions/Limitations

The opportunity to develop and define the concept of shared usability now had a format and frame of enquiry. A format that would allow the Older Adult participants narrate their day to day life and experience. This format would allow capturing 'uncertainty' and allowing the 'user' to be involved in the process of research and design (Papanek; 1985, Demirbilek 1999, Button 2000, ISO; 2002, Norman 2002, DreyfussH; 2012 Ed.)
The literature review provided the basis to plan the fieldwork and address ethical considerations to recruiting and engaging with participants. Together with important statistical evidence to progress with this research the literature reviewing clarified the following areas:

- Identified and defined the Older Adult as purpose User to be studied for this research.
- Acknowledged areas that can be problematic for Older Adults (i.e. fuel poverty, pressure ulcers)
- Identified a qualitative method of enquiry using ethnographic methods as a means to understand day to day life for Older Adults.
- Defined a need to seek ethical approval within Institute of Technology, Carlow for the parameters of fieldwork to be conducted.
- Highlighted a need to conduct Pilot Studies as the precursor to the main body of fieldwork.
- Assisted deep understanding to specific design philosophies that explore usability and consider more than one user (i.e. Universal Design, Inclusive Design, Transgenerational Design)
- Shared insight to various areas of understanding people and the psychology of experience and behaviour when using products or services.

The research explored three areas that were identified and stated both in the research title and learning outcomes from the Literature review. In addition, the three areas of focus created research questions:

- The Older Adult
  Research Questions:
  What is an 'Older Adult'?
  How can day to day activities and experiences be learned and understood?

- Shared Usability
  Research Questions:
  What is 'Shared Usability'?
  How can 'Shared Usability' be developed?

- Product Design
  Research Questions:
  What is Product design?
  Who are 'users?'
  How can Shared Usability become part of Design process?

The research hypothesis was developed as an outcome of the literature review and prior to the fieldwork.

'It is possible to empower Older Adults through Shared Usability by mutually agreed intervention with other stakeholders when using products or services.'

The researcher pursued enquiry with a tacit knowledge that was enhanced further by the narrative shared by the participants during the Pilot studies and fieldwork. The research developed at a pace that often required reflective periods. This was to assess and consider the previous stages of research while anticipating the potential development for future stages and outcomes. The objective of this consideration and reflection supported the researcher during sessions that required strategy and planning.

Design thinking is an intrinsic feature to design research. It offered the researcher an opportunity to explore and analyse the project or situation and deliver creative outcomes that are not detached segments but connected sequences to the 'whole' of the project (Brown,T. 2009). The research methodologies facilitated the iterative and non-linear nature of the design research as an exploratory process. The acceptance of this exploratory process was not to indicate a chaotic or disorganised approach; instead it displayed a creative approach undertaken by the researcher. This displayed the researcher's ability to share insight from observing actual experience and behaviour of people as a means to identify unmet needs.

Part Two – Fieldwork Methodologies

Fieldwork was conducted as a means to define unmet needs within eight areas ensuring a comprehensive record of Older Adult behaviour and experience. The fieldwork methods of observation, interview and task analysis within the day to day life for Older Adults revealed in-depth insight.
In addition, three Pilot studies were conducted offering new knowledge and insight into Associated Stakeholder involvement in Older Adult day to day activity.

Eighteen Older Adult participants and three associated stakeholders (Family member, Occupational therapist & healthcare manager) engaged with the researcher conducting fieldwork over a period of nine months.

The researchers own experience and building of knowledge throughout the fieldwork would support the development of the research. Interaction between the participants and the researcher would provide beneficial insight that would develop empathy as an objective measure to research outcomes (Denzin, N.K., & Lincoln, Y.S. 2005).

Ethnography *"involves the researcher participating, overtly or covertly, in people's daily lives for an extended period of time."* (Atkinson, P., & Hammersley, M., 2007, p.3).

Qualitative methodologies with an ethnographic approach were applied to observe and develop understanding of Older Adult day to day experience and behaviour as a means to understand Older Adult behaviour in two contexts:

- Life-Logging
- Task Observations

Life-logging

Life-Logging was conceived by Steve Mann as a method to record daily activity. Life-logging was used in this research to support an ethnographic method that could passively record Older Adult behaviour within the natural setting of their home. There was a total of sixteen life logging sessions conducted as part of this enquiry.

Firstly, a template was created for the Life-logging sessions. This was used to memo and document all notes or sketches during the sessions. The format of the template was structured into eight areas of enquiry with an informal approach that relied on a series of 'random words' (Collins, H., 2010) listed with each area of enquiry. This supported a strategy to explore the eight areas with open-ended questions that encouraged rapport, trust and storytelling with participants.

Task Observations

Observation is *"…the fundamental base of all research"* as discussed by Angrosino (Denzin and Lincoln 2005. p.729)

Task observation sessions were arranged with eight participants. Each of the participants would conduct a physical task linked with one of the eight areas of enquiry. The eight task observations were conducted in an unstructured format and led by the participants. The researcher discussed the proposed task observation with each participant prior to the activity. This was to ensure the participant was in agreement and also to discuss any other considerations necessary to the task (i.e. weather permitting for task outdoors)

The Task Observation sessions were an extremely rich source of contextual enquiry. They recorded human factors and ergonomic considerations. The participants were observed manoeuvring steps, furniture and fittings in order to complete tasks. During these observation sessions the researcher used a non-directed approach with the participants with the intent of conducting sessions that did not interfere with the participant's activity being observed. However, during sessions there were moments of direct interaction between the researcher and the participant. This direct interaction was seen as a positive development in the research because it supported the research hypotheses, whereby the Older Adult was inviting the researcher to engage or assist with the task, but on their terms. Shared usability was presenting itself to the researcher through the direction of the Older Adults.

The Older Adults naturally immersed themselves with the researcher within the focus of the task. Often the narrative from the participant would digress from the task being conducted to other subject matters important to them. This was beneficial to further understanding other aspects of day to day life. An example of this was when one participant shared the story of a house extension he and his wife decided to add to their home a number of years ago. At the time of construction, they decided to incorporate ramps as part of the outside access paths into the home. This insight demonstrated forward thinking of this participant and his wife. The rationale being that should they require mobility devices in the future, the familiarity of the ramps will be less intrusive as they adjust to a new means of mobility and independence.

At this point of the research validation of fieldwork findings was required. Furthermore, it had to be analysed with consideration to the research hypothesis:

'It is possible to empower Older Adults through 'Shared Usability' by mutually agreed intervention with other stakeholders when using Products or services.'

Krippendorff & Butter refer to a '*network of stakeholders*' as one of the four conceptual pillars that support Human Centred Design. (2008) Krippendorff & Butter discuss how, in addition to the user there are various stakeholders that become the 'network of stakeholders' relevant to the design outcome.

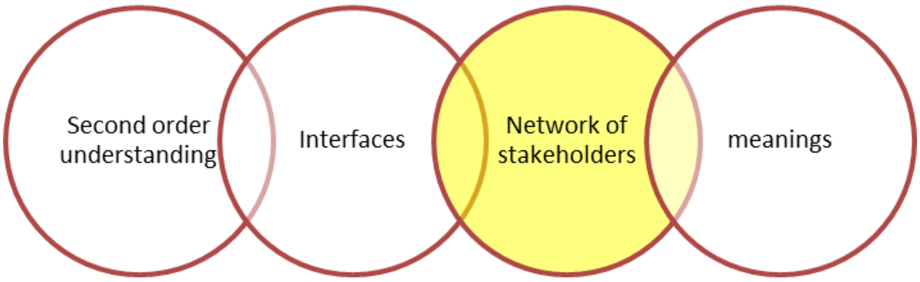


Figure 4. Four Pillars that support Human Centred Design as interpreted from Krippendorff & Butter 2008.

The User is described by Krippendorff & Butter as almost a figment built out of a "*rhetorically convenient illusion that designers offer their clients in justifications of their design*" (2008, p.358). There is a hierarchy of priority placed around the other considered stakeholders from clients who represent the business, financiers, engineers, market researchers, merchants, governmental agencies, buyers (not the user), repairpersons, recyclers, ecological activists, and others who will "*variously experience a design and collectively affect its fate.*" (2008, p.358)

> *"Human-centred designers must acknowledge the critical role of stakeholders – supporters and opponents – welcome their active roles in bringing a design to fruition, and see themselves not as masterminding the process, but as active participants in such networks as well."*
>
> - (Krippendorff & Butter, 2008, p.358)

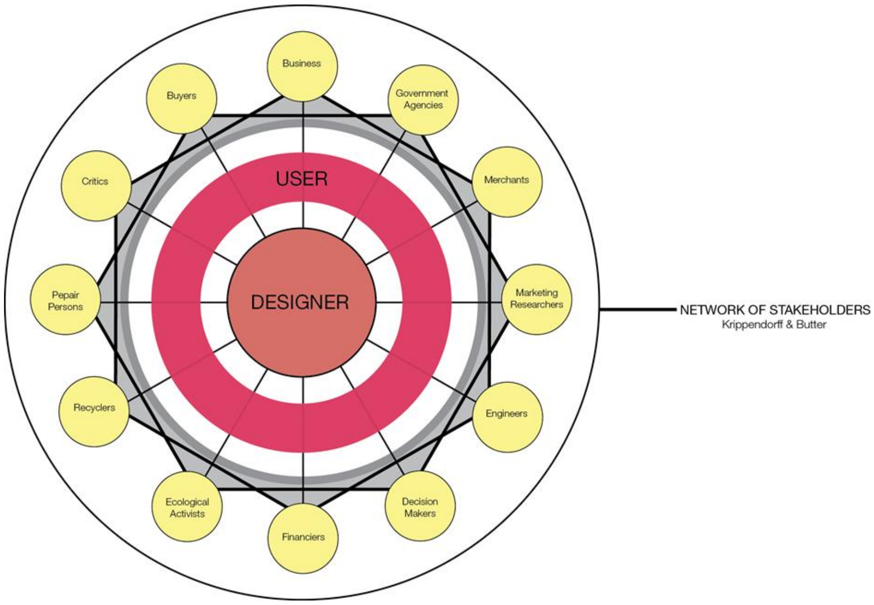


Figure 5. Interpretation of the network of stakeholders as per Krippendorff & Butter, 2008.

Krippendorff & Butter's *network of stakeholders* expresses the responsibility of the designer to consider more than the user in the process of design, but from the perspective of the stakeholders involved in the development and delivery of concept to product development for the user. This research evolves the network of stakeholders to one that provides a support framework for the user through the network of *Associated Stakeholders* for shared usability when using products or services and is displayed in figure 6.

The researcher analysed the depth of knowledge gathered from the life-logging and task observation sessions as a means to underpin and define data. This data was then coded and indexed resulting in conceptual outcomes.

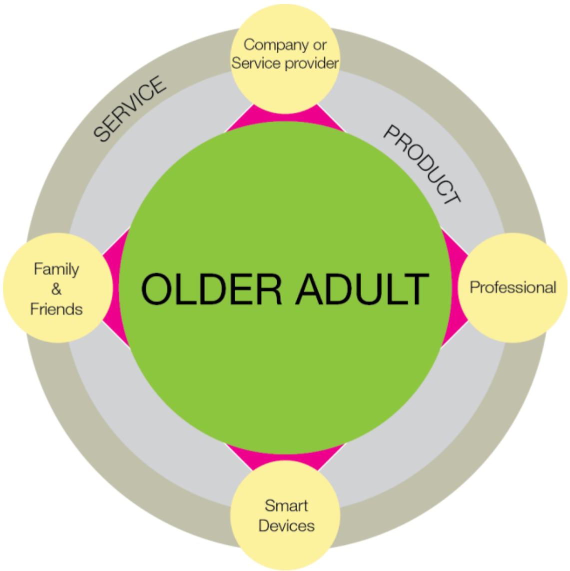


Figure 6. Network of Associated Stakeholders -Shared Usability – (Authors own).

Using a thematic coding approach (- Robson, C; 2011 Ed.) the data collected was collated and placed into themes as per the eight areas of enquiry. After which each theme were assigned labels. The labels were created and directed by the participant's responses and narrative during the life logging and task observation sessions.

Part Three - Research outcomes

The development of the concept of shared usability as a valid mechanism to the design process is evident by the body of research undertaken; in addition it has supported new knowledge outcomes:

The three new knowledge outcomes are as follows:

1. Fieldwork detailed enquiry into Older Adults day to day life experience
2. A definition of Shared Usability.
3. Product Concepts that display Shared Usability benefits to user experience.

Fieldwork detailed enquiry into Older Adults day to day life experience
The ethnographic approach to observing and understanding older adult participants in natural settings provided a deep source of insight and data in eight areas of daily life. The outcomes of fieldwork assisted in identifying and understanding potential product areas that could support a tangible understanding of shared usability.

A definition of shared usability
White described shared usability as a concept for independence (White, P.J., 2012). The purpose of this research was to enquire further into shared usability and to offer design examples from this enquiry. This was achieved by conducting fieldwork with Older Adult participants and other stakeholder's.

The research conducted, highlighted the potential benefits of shared usability in the design of products and services for Older Adults. This research also clearly highlighted the benefits of the engagement of User with Associated Stakeholders in product or services use. The User and Associated Stakeholder network have also been defined in this research offering understanding of the potential relationships that can support shared usability.

The definition of shared usability created from this research is as follows:

> *Mutual agreement between the User and Associated Stakeholders on the level of management or interaction required with a product or service as an objective to achieve positive usability.*

Product Concepts that display shared usability benefits to user experience.
The eight areas of enquiry pursued during fieldwork were comprehensively explored as a means to identify unmet needs in products and services for Older Adults. This offered the researcher many areas to pursue design conceptual development. The fieldwork data gathered was triangulated determining the area of 'Access' as the most appropriate area to progress product development conceptualise within. The conceptual stage involved further feedback sessions informally between the researcher, Older Adults and Associated Stakeholders as a means to determine concept and product outcome.

During the fieldwork, a number of the participants had discussed problems regarding their home heating systems, some of them sharing how they often do not 'set' or automatically time their central heating using the timer- *particularly* mechanical timers.

Figure 7. Sample mechanical timer as per a participant's home.

A number of reasons were offered:

- Some participants preferred to know the cost implication, and preferred to turn it on and off manually as required
- 'Pins' that you raise or lower to set the time were too awkward to manage with fingers
- The location of the timer was poorly lit
- The small print of the numbers is difficult to see and accurately set the time
- Some participants felt it was more challenging to set the timer
- The location of the timer is usually under a press or cupboards and often located in the 'hot press'.

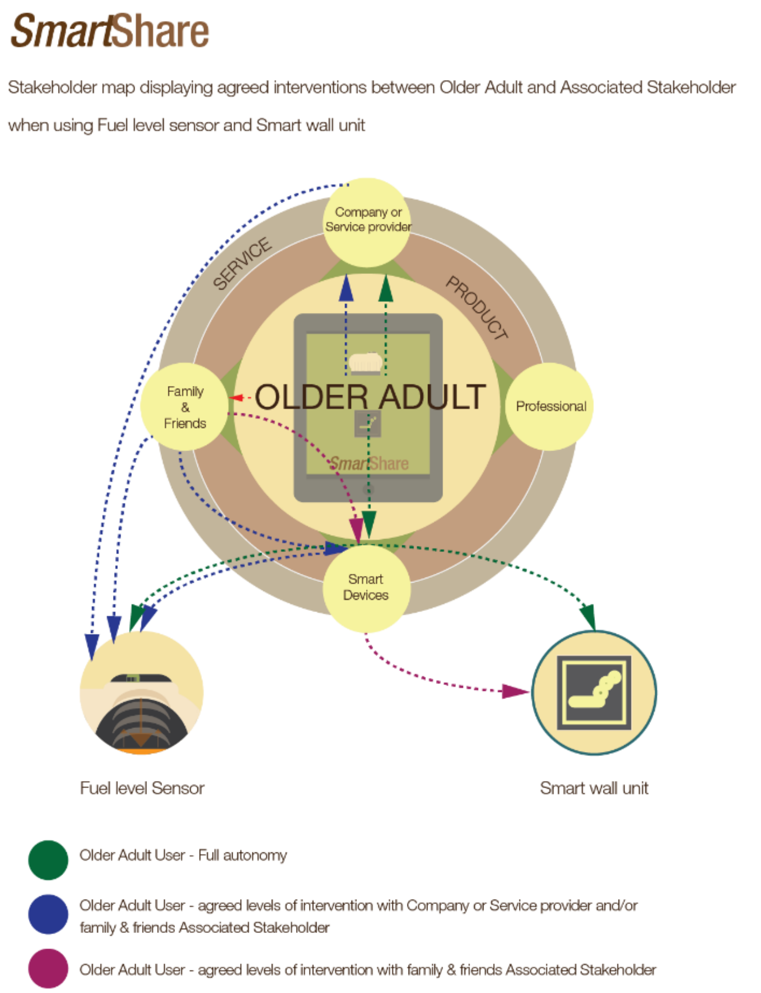


Figure 8. Stakeholder map displaying interventions between Older Adult and Associated Stakeholders.

The product outcome for this research, is one that involves the User (Older Adult) plus Associated Stakeholders (these can be family members, service providers, other companies). It is a retrofit device that is operated on a smart device, tablet or computer. It can be accessed by the user (Older Adult) or managed levels agreed between the Older Adult and Associated stakeholders. There is a second product need identified with the fuel supply and provision to the home, again the concept here is managed through a wireless network and agreed between the Older Adult and perhaps the utility company or service provider. The third area is the system of the ‘App’ supporting the product use and management. This can have further services or features added to as they get developed. This potentially could provide a home with services such as lighting, security, access in addition to the heating and fuel management concepts as an overall home service management system. shared usability stakeholder map is shown on Figure 8 in relation to usability access from user and associated stakeholders.

The conceptual product outcome of the *Smart*Share system supports the application of shared usability to the design process. This is achieved through iteration and collaboration of the following:

- User Centred Design
- Principles of Universal Design
- Design for all approach

User Centred Design considers the needs of a user when applied to the process of design. It requires defining unmet needs as identified by the user. The designer’s responsibility is to deliver a product or service that is intended to fulfil these needs. However, the limitations of User Centred Design can be restrictive when Shared Usability is applied because of the requirement to consider the network of Associated Stakeholders to support the autonomy of the user – the Older Adult.

The principles of Universal Design were beneficial to the consideration of more than one user as a means to a design outcome; it supports also the consideration of human abilities and function when considering product or service system development. This was beneficial particularly to the impact of limiting function and mobility associated with ageing. However, where this faltered was the need to expand and Associated Stakeholders as supporters to the autonomy and independence of the Older Adult using products or services.

“Design for all relies on the involvement of potential users, where this means not only the end users, but all those involved in the design, development, production and marketing processes.” (Krauss 2011, p. 13.2).

The following series of images (Figures 9, 10 and 11) display iteration and development to the relationship of the Design philosophies discussed, and implemented during the conceptual phases as a means to promote the value of *shared usability design*.

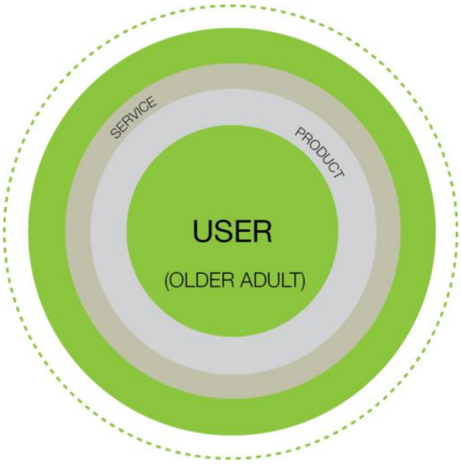


Figure 9. User Centred Design.

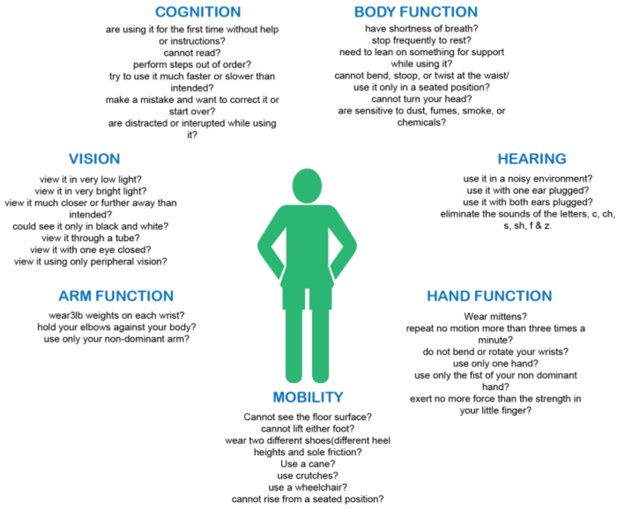


Figure 10. Spectrum of Human Abilities as per Universal Design file.

Conclusions

This research revealed that shared usability was previously an undefined existing activity that Older Adults and Associated Stakeholders engaged in. The research undertaken offers a definition of shared usability; which supports the requirements capture to consider more than one user engaging in the use of products or services. This research focussed on the Older Adult as the 'User' however the promotion of shared usability could offer enabling and empowerment to all users.

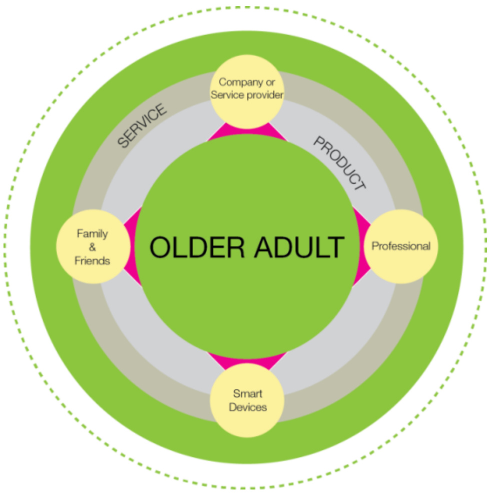


Figure 11. Shared Usability Design.

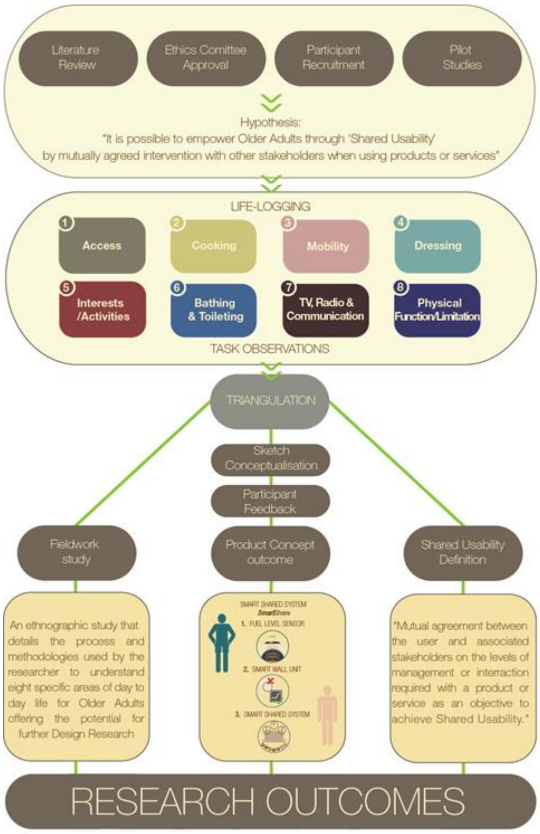


Figure 2 'story' and overview of this research

Future research could explore areas such as Older Adults and dementia, being supported by Associated Stakeholders as a means to prolong independence. Another example that considers shared usability and 'Users' other than Older Adults could be the area of play and recreation for children that would allow the child explore and be curious, whilst also being supported by the Associated stakeholders in their lives – Parents, Guardians, Educators etc. This could be an area that collectively could support the area of healthy eating and obesity or outdoor activities as examples. As a record of the work conducted and completed during this research, Figure 12 highlights the 'story' and overview of this research. It begins with the assessment of what was required in order to develop the hypothesis, and fieldwork strategy. The Life-logging sessions and task observations culminated to a stage of triangulating the gathered data and knowledge as a means to deliver new knowledge outcomes that conclude with product, fieldwork and shared usability outcomes. Shared usability was defined as an outcome to the research. This became a mechanism to support the process of design. The definition proposes that shared usability facilitates a 'User' and a network of Associated Stakeholders to manage and agree levels of interaction and usability when using products or services. Furthermore, it provides autonomy to the User enabling them to remain empowered as a result of initiating levels of usability with the Associated Stakeholders.

References

Atkinson, P., Hammersley, M. (2007). *Ethnography Principles in practice 3rd Edition*. London. Routledge (Taylor & Francis group).

Brown, T. (2009). *Change by Design - How design thinking transforms organisations and inspires innovation*. United States of America, Harper Collins Publishers.

Button, G. (2000). The ethnographic tradition and design. *Design Studies*, *21*, 319-332.

Collins, H. (2010). *Creative Research*. Lausanne, Switzerland., AVA Publishing SA.

Demirbilek, O. (1999). *Involving the elderly in the design process: A participatory design model for usability, safety and attractiveness*. Ph.D, Bilkent University.

Denzin, N.K., & Lincoln, Y.S. (2005). *The SAGE handbook of qualitative research*, Thousand Oaks, Sage Publications.

Dreyfuss, H. (2012 Ed). *Designing for People*, United States of America. Allworth Press.

Farage, M.A., Miller, K.W., Ajayi, F. & Hutchins, D. (2012). *Design Principles to accommodate older adults*. Global Journal of health Science, vol 4, issue 2.

ISO. (2002). ISO/TR 16982:2002(E). *Ergonomics of human-system interaction -- Usability methods supporting human-centred design*. Switzerland. International Organisation for Standardisation.

Krauss, I. (2011). *Universal Design handbook* - 2nd edition, United States of America, McGraw-Hill Companies.

Krippendorff, K., & Butter, R. (2008). Semantics: meanings and contexts of artifacts. In Schifferstein, H,N,J., & Hekkert, P.,eds. Product Experience, Amsterdam, The Netherlands, Elsevier. P. 353.

Norman, D.A. (2002). *The design of everyday things*, Basic books.

Papanek, V. (1985). *Design for the real world - Human ecology and social change.* Thames & Hudson.

Parker, M.G., & Thorslund, M. (1991). The use of technical aids among community-based elderly. *The American Journal of Occupational Therapy, 45*, 712-718.

Robson, C. (2011). *Real World Research: A Resource for Users of Social Research Methods in Applied Settings*. UK. John Wiley & Sons.

Story, M.F., Mueller, J.L., & Mace, R.L. (1998). The Universal Design File: *Designing for people of all ages and abilities*. North Carolina State University. United States of America. The Center for Universal Design

United Nations (2002). *Political Declaration and Madrid International Plan of Action on Ageing*. [http://undesadspd.org/Ageing/Resources/MadridInternationalPlanofActionon Ageing.aspx [accessed 10 June 2014]

White, P,J. (2012). *Designer as Ethnographer: A study of domestic Cooking and Heating Product Design for Irish Older Adults*. Ph.D, National University of Ireland, Maynooth.

World Health Organisation. (2001). *International Classification of Functioning, Disability and Health*. Geneva, Switzerland: World Health Organisation.

Human-centred development of automatically accommodating contact lenses

Max Bernhagen, Patrick Roßner, & Angelika C. Bullinger
Technische Universität Chemnitz, Germany

Abstract

Every person is affected by a decreasing accommodation width. On average, people will not notice this until their fourth to fifth decade of life. If the accommodation width drops below three dioptres, this is called presbyopia. As a result of this decrease, people need a treatment for good vision in the near field. Therapeutic approaches like laser surgery or intraocular lenses are invasive and therefore may involve complications. Common reversible and non-invasive alternatives are glasses and contact lenses. Within this paper, the human-centred development of a contact lens that automatically adapts the refractive power of the lens to the distance of a faced object is presented. This innovative approach will compensate the loss of accommodation width and is going to integrate sensors, actuators and energy supply within the lens. A mixed-methods research design has been applied to identify user requirements as well as requirements by physiological aging. Deliverables from the evaluation of three focus groups, six in-depth interviews and a follow up survey indicate that experts acknowledge such a system as highly desirable. However, some hurdles are expected, e.g. pricing models, acceptance of older people and the competences to adapt such a system to a patient's eye.

Introduction

The ability to change the refractive power of the human eye to see objects in different distances clear is called accommodation (Baumeister & Kohnen, 2008). A specific value for assessing the ability of the eye to accommodate is the width of accommodation. This represents the ability to focus objects in near distances. A relaxed eye focuses objects in infinity. By a contraction of the ciliarian muscle, the lens deforms and the refractive power changes. However the eye's ability to change the refractive power decreases as people get older (Duane, 1912). Young infants have a width of accommodation of about 20 dioptres. As people get older, the width of accommodation shows an asymptotic expansion to the axis of abscissae and levels off at approximately one dioptre. People having an accommodation width below three dioptres are ranked as presbyopics. On average, people pass this mark during their fourth decade of life. Having an accommodation width below three doptres implies that one may not be able to see objects clear within near distances (Nagel, 2012). An accommodation width of three dioptres for an emmoetropia's eye

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

equals the distance to focus on an object to at least 33 centimetres. Therefore daily activities, e.g. reading the newspaper, finding the expiry date or recognizing the time on a wristwatch are influenced. This can be reduced by deploying treatment alternatives, for example glasses, contact lenses or operative procedures. Common standards for non-invasive treatments are regular and varifocal glasses as well as monovision and multifocal contact lenses. Invasive methods for the recovery of the accommodation width are intraocular lenses, laser surgery and intracorneal implants. Innovative and emerging treatment opportunities use miniaturised electronic components within existing technologies. For example Nagel (2012) presents a concept for active intraocular lenses which automatically adapt the refractive power. However, the necessary invasive surgery causes a risk for side effects and is not easily repeatable. On the contrary, active contact lenses are non-invasive and consequently one can easily exchange defect lenses or stop using them, if side effects occur.

The suggested automatically accommodating contact lens as a part of active contact lenses measures the distance to the focalised object by detecting pupil characteristics. Therefore miniature sensor components are needed. Furthermore actuators to change the refractive power of the lens, an internal communication infrastructure as well as an energy source need to be implemented into the contact lens. Besides technical aspects, potential end-users and their requirements have to be considered, because they will use and interact with the product. In order to achieve a holistic view of the development of such a product, this paper illustrates how the human-centred development process has been used (DIN EN ISO 9241 part 210, 2010). The aim of this approach is to integrate end-users and their knowledge in the development process. Therefore analysis of the context of use, user requirements, system design and system evaluation were performed. This paper focuses on the last three stages.

Methodology

Due to the complex and innovative character of the subject a mixed-methods research approach was utilised (Creswell, 2009). In first place qualitative data of contact lens experts were gathered with the help of three focus groups and six in-depth interviews to conduct participatory development of ideas (Möslein et al., 2010). These methods were chosen because of the less explored and documented subject active contact lenses (Gläser & Laudel, 2010). Based on the results an online survey was created which should assess the previously derived assumptions of user requirements and user acceptance. The method focus group has been chosen because it is a group discussion and enables the researcher to gather detailed data and perform a flexible discussion without focussing on a determined structure. One major advantage is the direct interaction with the participants. Thereby the moderator is able to guide the participant through the planed phases and may ask follow-up questions or clarify responses (Langford, McDonagh, 2003). Each stakeholder that has not been able to participate in the focus groups was given the opportunity to take part via a guided interview. This research method is highly suitable for the exploration and deepening of new research fields as well as the recognition of relevant questions and problems by discussions with the interview partner. Guided interviews should follow a structured guideline to ensure comparability of the gathered data (Mayer, 2008). Based on the focus groups and interview results an online survey was conducted. Thereby participants complete a questionnaire online via internet or send a filled questionnaire via E-Mail to the inquiry leader (Jackob et al., 2009). Advantages of online surveys are low collection costs, anonymity of participants, variable ways of depiction and the large scatter range. On the other hand disadvantages of online surveys exist like attendance requirements, sampling problems, social desirability, ambiguous question formulation or answer categories and the lack of control over the participation situation (Pötschke & Simonson, 2001).

In preparation of the focus groups and the online survey, characteristics of contact lenses were identified by the help of a literature review. In order to use them within the online survey these were supplemented by additional characteristics quoted by the participants of the focus groups. Table 1 provides an overview of the findings of the literature review.

Table 1. Characteristics of contact lenses

Characteristic feature	Reference
Limited daily wearing duration	(Schnell & Khaireddin, 2013)
Soft lenses may stick together	(Schnell & Khaireddin, 2013)
Getting out of place on the eye	(Schelle, 2000)
Allergy to contact lens materials or cleaning agents	(Blümle et al., 2013)
Necessary cleaning effort	(Janiak, 2001)
Due to the reduced oxygen supply contact lenses have an overall limited wearing duration	(Koniszewski, 1982)
Feeling of a foreign substance on the eye	(Hartmann & Goertz, 2013)
Try eyes due to contact lens wearing	(Blümle et al., 2013)
Disruptive sediment deposits on the lens	(Janiak, 2001)

Based on qualitative research results of the focus groups and in-depth interviews a questionnaire has been developed. The motivation for this was to assess the characteristics stated by experts and found in literature as well as to analyse whether differences between users of contact lenses and participants who do not use them can be carried out. In addition design alternatives of contact lens systems need to be evaluated by the end-users. Thus the questionnaire was separated in four parts. Within the first, participants had to mark whether they use glasses, contact lenses or other vision aids. Additionally they were ought to rate their usage characteristics and the frequency they go to ophthalmologists. The second part dealt with ratings of most typical characteristics of vision aids. Here, participants were ought to rate their opinion on five point Likert scales. The penultimate part of the questionnaire contained the evaluation of the three design alternatives. For this, the *Usability Experience Questionnaire* (Laugwitz et al., 2008) was utilised as a standardised measurement method. By the help of this questionnaire one can achieve information regarding hedonic and pragmatic quality dimensions using the scales attractiveness, perspicuity, efficiency, dependability, stimulation and novelty. Stimulation and novelty refer to the hedonic quality and perspicuity, efficiency as well as dependability to the pragmatic quality (Rauschenberger et al., 2012). The scale attractiveness is a valence dimension which shows a positive or negative attitude towards a certain product. However the dimensions perspicuity, efficiency and dependability could not be utilised because of the early development stage of the concept.

Results

In-depth interviews and focus groups

All in all the analysis of three focus groups provided 291 text extracts which were linked to 3 main categories and 11 sub categories. In total, 14 people (9 females and 5 males) contributed. All of them were experienced opticians with daily customer contact. In addition to this, 201 text extracts for 5 main categories and 20 sub categories were found in in-depth interviews. Here, three managers of contact lens manufacturers, one ophthalmologist and two managers of contact lens institutes were interviewed. In the following section a consolidation of the results regarding active contact lenses and customer perspective will be given.

In regard to active contact lenses hurdles and risks are in contrast to several odds. First all the participants stated that this new technology might revolutionise the contact lens market, if the framework conditions were considered. Mayor advantages are an improved visual field in comparison to current multifocal lenses due to the repeal of optical zones. In contrast to varifocal glasses the active contact lens is less obtrusive and the limited viewing angle due to the boarder of the glass is not existent. Most importantly for physiological considerations is the intended material of the lens. The material has to ensure that enough oxygen reaches the cornea, has to fit to different lachrymal liquid compositions and should not have any incompatibility of the electronical components. Especially participants of the in-depth interviews demand that the active contact lenses have to be fitted individually to the customer's eye. Otherwise a long wearing period is hardly to be achieved because of a higher risk for pathological changes in the eye. Regarding the integrated components participants stated that one should be aware of the energy supply. Here the system has to work as long as possible without recharging it. A final risk is the price of the lenses. Due to the early development state no assessment of the price is possible. Participants of the focus groups and in-depth interviews say that the price has to be attractive. Therefore, alternative pricing models like a monthly or yearly rate may reduce the perceived costs. Regarding the handling of active contact lenses no differences should be expected, because the handling itself will not differ in comparison to regular contact lenses. One advantage is the reduced time to get used to the lens because the new contact lens automatically adapts the corrective power. Thus the user does not need move his head and eyes in order to use optical zones on a contact lens or glass.

Next, the results of the discussion about customer perspective of the use of contact lenses are of interest. Participants stated that people who are used to contact lenses have no problem using them on a regular basis. Even elderly people are able to put the lenses in and out as well as clean them. A group that drops-out using contact lenses for the correction of presbyopia consists of people who always wear lenses which corrected myopia. When they have to decide whether they want to wear new contact lenses for the reduction of the bites of presbyopia they do not accept new disadvantages, for example light wells or a reduced contrast. Therefore they continue to wear contact lenses for myopia and additionally wear glasses for near distances. Furthermore, participants stated that usually young people in the age between 16 and 30 years are interested in contact lenses. They typically want a vision aid which does not influence their natural appearance. Pathological effects inhibiting wearing glasses are rare reasons. It was also said that because of the trend towards a more active lifestyle and extraordinary hobbies many people need contact lenses for their free time activities. Experts stated that elderly people who already wear contact lenses are more willing to wear contact lenses for the correction of presbyopia. Only on rare occasions people want to try contact lenses for the treatment of presbyopia if never wore contact lenses before. Independent of the correction type people want to regain their natural eye sight. This means that they want a high refractive power, no optical zones as well as a comfortable and low obtrusive vision aid. People who do not want any device for vision treatment mostly use laser surgery. Finally the distribution of contact lenses in Germany may cause problems because customers are able to avoid the consultation of an expert. By buying contact lenses on the internet or drug stores one cannot speak to a contact lens expert. Thus important information about characteristics of contact lenses, necessary cleaning instructions or signals of pathological changes within the eye may not be addressed. As a result, experts state that the distribution of mass produced contact lenses on the internet and unauthorised shops should be avoided in order to secure a good consultation and to have the opportunity to examine the patient's eye to prevent pathological changes.

Online survey

The online survey has been conducted from May 15th to May 31th. In total 174 people participated of which 43 percent (75) were females and 57 percent (99) males. The mean age was 25.80 years (*MD*= 24; *SD*= 7.13; range: 19 – 75). Furthermore participants were asked about existing vision impairments and their possession of vision aids. Results are shown in table 2.

Table 2. Typology of the participants

	Type of vision impairment	**Possession of vision aids**	
		Glasses	**Contact lens**
Myopia	53.70%	51.28%	34.84%
Hyperopia	6.17%	8.39%	2.58%
Presbyopia	3.71%	5.77%	0.65%
Emmetropia	36.42%	-	-

Approximately 60 percent of the participants state that they have vision impairment. In particular myopia is a common impairment which affects more than half of the sample. These are persons who are not able to see objects clearly in near distances. In contrast 6.17 percent of the participants have problems seeing objects in far distances. A smaller part stated that they suffer from presbyopia which is an age related impairment. Within investigations of samples of elderly people this value may be higher. Additionally approximately one third of the sample states that they have no vision impairments (emmetropics). Figures for the possession of vision aids include participants who use these on a regular as well as an irregular basis. Undoubtedly most of the sample utilise glasses whereas the correction of myopia is of major interest. The recorded values match with published values from the Institut für Demoskopie Allensbach (2014). They figured out that 63.5 percent of the German population is wearing glasses.

Furthermore a contact lens penetration of approximately 35 percent has been stated by the participants. This value strongly diverges from the values published within literature. Allensbach (2011, 2014) for example collected data for Germany (penetration: 2011 – 5.3%, 2014- 5.2%) and the eastern part of Germany (penetration: 2011 – 3.6%, 2014- 2.8%). A reason for the variation of the contact lens penetration may be due to the non-representative sample. The inquired participants were relatively young and had a high education in comparison to the average German population. Results from a study conducted by Nagl and Braun (2012) seem to emphasize an age affect, because the recorded a penetration of contact lenses of people between the age 16 and 25 of 31.6 percent. Within the age 26 to 50 the penetration decreases to 18.4 percent.

As discussed in section 3 the online survey included three design alternatives for contact lenses. People had to assess the lenses depicted in figure 1.

Contact lens 1 (normal)	
Contact lens 2 (obtrusive)	
Contact lens 3 (unobtrusive)	

Figure 1. Design alternatives for automatically accommodating contact lenses

In order to set a baseline a standard contact lens has been chosen as one alternative. It is a standard lens which is well known and rarely visible. Within an expert workshop, several design alternatives were discussed and narrowed to two. These two designs base on technological considerations. Contact lens two represents an obtrusive version which may be realizable within early product development stages. Here the colour is not adaptable to the eye's colour. Contact lens three is mostly unobtrusive because of especially covered electronical components. This is due to the colouring of the electronical components as well as the utilization of semi-transparent materials. However experts state that the central board which contains most of the sensors and actuators cannot be adapted to the eye colour. The results of the assessment of the three design alternatives are shown in figure 2.

Figure 2. Assessment of design alternatives

In the User Experience Questionnaire, a limit value defining the boarder to reliable tendencies is given. This value is 0.8 and depicted by a red line. First participants judged the attractiveness different. The obtrusive contact lens is assessed neutrally with a non-reliable negative tendency (*M* = -0.679). In contrast contact lens one (*M* = 1.263) and three (*M* = 0.912) are attractive to the participants with reliable tendencies. The overall rating of contact lens stimulation is positive. Although the high attractiveness for the classical lens (*M* = 0.497) participants assess the stimulation as neutral with a trend towards positive. Contact lens two (*M* = 0.784) and three (0.988) transcend the limit value and are assessed as stimulating. Novelty of lens one (*M* = -0.115) has a neutral ranking with a negative tendency. However participants assess active lenses as novel and innovative. Especially the obtrusive lens (*M* = 1.373) received a strong positive rating followed by the rating 0.943 for contact lens three. Over all scales the unobtrusive contact lens receives reliable positive ratings by the participants. For assessing the results the Cronbach's alpha values for each contact lens were calculated. First the correlations of the items for the regular contact lens resulted in α = .78. Furthermore Cronbach's alpha value for contact lens two is α = .85 and for contact lens three α = .82. Consequently one can say that reliable scales were used. Especially due to the low number of items asked these values proof the validity of the questionnaire (Field, 2009).

In addition to design considerations the opinion of participants on characteristics of vision aids were obtained. Therefore questions about characteristics determined by qualitative data analysis as well as literature review were asked. In order to figure out differences due to possession of vision aids the sample was divided in three groups. People who never owned a vision aid, people using glasses and participants who are experienced in wearing glasses and contact lenses were chosen as an independent variable. A further intended group are people using only contact lens. In the sample only one participant fulfilled these criteria. Therefore this group was excluded from consideration. Diagram 3 presents the results inclusive markers for significant differences. In order to obtain statistical evidence an ANOVA was executed.

Six questions show significant differences within the groups. Of special interest are differences between contact lens users and participants who do not wear them. Here the assessments regarding care effort, feeling of foreign substances and inserting as well as removing the lens are rated significantly different. Each negative characteristic is assessed less worse by participants who have experience in using contact lenses. This implies that they do not assess these things as a problem and these factors may be biased opinions for other groups. Also this tendency can be found for the item concerning the containment of planning of daily activities due to contact lenses. However the difference here is less distinct. Furthermore one may see significant differences for the first and fourth item. These items refer to negative characteristics of glasses. Similar to the assessment of contact lens characteristics, people who have possession of glasses rate the items as less negative. Even though they agree on these two negative characteristics they rate it better than the other groups.

In general participants mostly agree to the statements that a good vision for close and far distance would improve their lives as well as it is important to have an unlimited field of view. Furthermore they agree to the statements that a vision aid should not influence the personal appearance and that limitations in hobbies due to the visions aids are not acceptable.

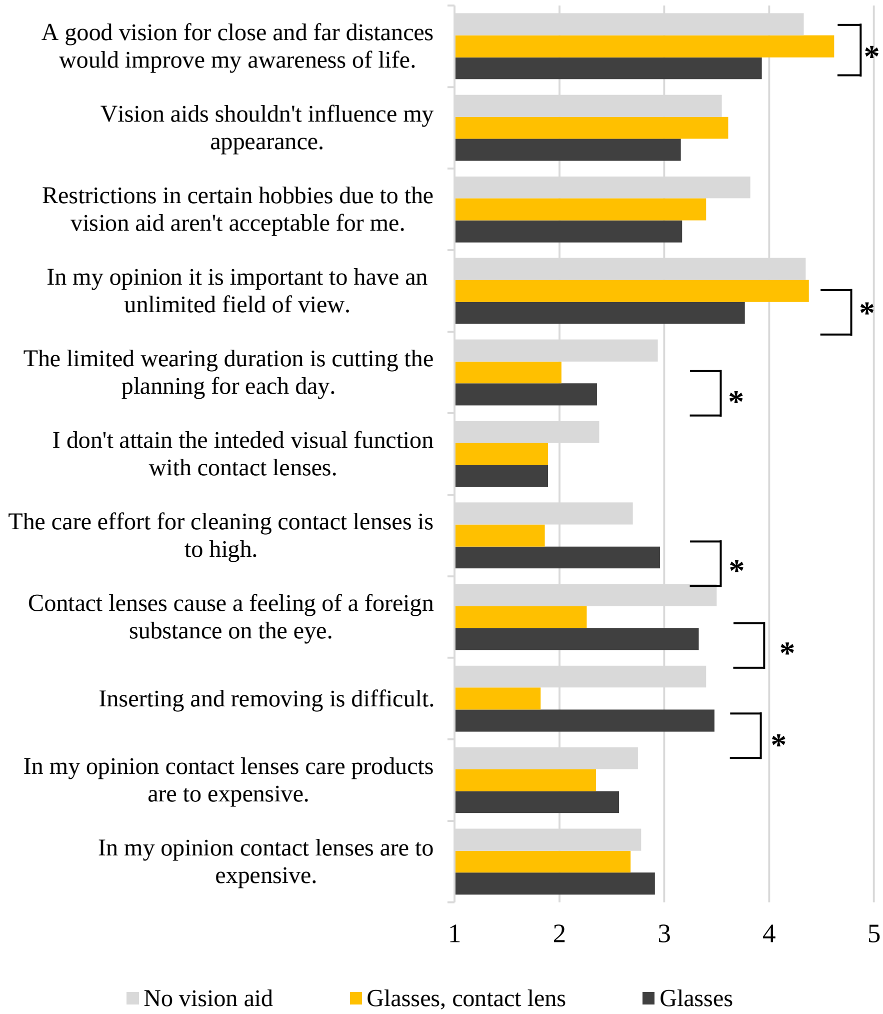


*Figure 3. Comparison of user assessment with distinction between vision aid possession (**1**- I don't agree; **5** – I absolutely agree, * p < 0.05)*

Summary and discussion

In conclusion the conducted research provides a basis for the technical development of an automatically accommodating contact lens, assessed by potential end-users. Especially the fact that about one third of the survey participants already is experienced in wearing contact lenses underlines that they are a meaningful alternative for the treatment of vision impairments. Considering the expert statement that almost all of the users of contact lenses for the treatment of presbyopia are people who are experienced wearing contact lenses, the today's young people should be responsive to the novel lenses. Referring to the assessment of design alternatives, participants rate the innovative lenses as novel and stimulating. However the design should be developed with caution because obtrusive solutions are assessed as less attractive. The aim should be to develop an unobtrusive contact lens which is rated as attractive with only a slight difference to regular contact lenses.

Participants also indicated that the current contact lenses are not too expensive. Therefore they may be an alternative for automatically accommodating lenses. Here novel pricing models should be considered in order to lower the subjective perceived costs of the innovative contact lenses and convince end-users that the value added is worth the additional costs. Financial subventions by health insurances may support the distribution as well. Furthermore all advantageous characteristics of automatically accommodating contact lenses were positively rated by the participants. This shows that people demand a vision aid which should not constrict their free time activities does not limit the field of view and does not influence the appearance. All these factors can only be fulfilled by contact lenses or invasive surgeries. In addition people experienced in wearing contact lenses rated most of the negative characteristics of contact lenses as significant less meaningful as people who are unexperienced. This indicates that prejudices against contact lenses are present and need to be reduced. A reason for the deviations may be memories of older people who used contact lenses decades ago. At that time the lenses were larger and had a lower wearing comfort. One further requirement stated by experts is a regulated distribution of automatically accommodating contact lenses. This ensures on the one hand that experts can advise their patients in contact lens care, advantages and disadvantages and possible eye disease due to wearing contact lenses. One the other hand experts can fit the lenses to each individual’s eye. This results in a better compatibility of contact lenses.

To summarise the results of the mixed methods research approach, rich insights into user requirements, system design as well as organisational matters were carried out. Furthermore the results emphasize a high potential for the technology of automatically accommodating contact lenses. In following steps, the evaluation of early prototype designs as well the assessment of accommodation characteristics by the help of virtual simulation has to be conducted.

Acknowledgement

The joint research and development project “SmartLens” was funded by the German Ministry of Education and Research. The research association was conducted by Fraunhofer Institute for Electronic Nano Systems and consists furthermore of the Technische Universität Chemnitz, Professorship for Ergonomics and Innovation Management.

References

Institut für Demoskopie (IfD) Allensbach (2011). **Allensbach-Brillenstudie 2011** „Sehbewusstsein der Deutschen“ im Auftrag des Kuratorium Gutes Sehen (KGS).

Institut für Demoskopie (IfD) Allensbach (2014). Allensbach-Brillenstudie 2014/2015 „Sehbewusstsein der Deutschen“ im Auftrag des Kuratorium Gutes Sehen (KGS).

Baumeister, M., & Kohnen, T. (2008). Akkommodation und Presbyopie, Teil 1: Physiologie der Akkommodation und Entwicklung der Presbyopie. *Der Ophthalmologe*, *6/2008*, 597-610.

Blümle, S., Kaercher, T., & Khaireddin, R. (2013). Richtige Kontaktlinsenhygiene. *Der Ophthalmologe, 6/2013*, 522-527.

Creswell, J.W. (2009). *Research Design. Qualitative, Quantitative, and Mixed Methods Approaches*. Licoln: Sage Publications.

DIN EN ISO 9241 Teil 210 (2010). Ergonomie in der Mensch-System-Interaktion – Teil 210: Prozess zur Gestaltung gebrauchstauglicher interaktiver Systeme.

Duane, A. (1912). Normal values of the accommodation at all ages. *Journal of the American Medical Association, 59*, 1010-1012.

Field, A. (2009). *Discovering Statistics unsing SPSS*, Third Edition. London: Sage Publications.

Gläser, J., & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse*. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hartmann, B., & Goertz, W. (2013). *Arbeitsplatz Augenpraxis*. Berlin: Springer.

Jackob, N., Schoen, H., & Zerback, T. (2009). *Sozialforschung im Internet: Methodologie und Praxis der Online-Befragung*. Wiesbaden: VS Verlag für Sozialwissenschaften, GWV Fachverlage GmbH.

Janiak, C. (2001). Damit der Durchblick erhalten bleibt: Proteinentfernung von Kontaktlinsen. *Chemie in unserer Zeit*, 6/2001, 348-354.

Langford J., & McDonagh, D. (2003). *Focus Groups: Supporting effective product development*, London: Taylor & Francis.

Laugwitz, B., Held, T., & Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. In A. Holzinger, *HCI and usability for education and work, Proceedings of the 4th symposium of the workgroup human–computer interaction and usability engineering of the Austrian Computer Society* (pp. 63-76). Berlin: Springer.

Koniszewski, G. (1982). Korneale Neovaskularisation nach kurzzeitigem Tragen weicher Kontaktlinsen. *Klinische Monatsblätter für Augenheilkunde*, 12/1982, 467-470.

Mayer, H.O. (2008). *Interview und schriftliche Befragung – Entwicklung, Durchführung, Auswertung*, 4. Auflage. München: Oldenburg Wissenschaftsverlag.

Möslein, K.M., Halle, J.B.A., & Bullinger, A.C. (2010), Open Evaluation: IT für das Innovationsmanagement, *Wirtschaftsinformatik & Management*, *5/2010*, 24-33.

Nagl, A., & Braun, J. (2012). Ergebnisse einer Trendumfrage der Hochschule Aalen. *DOZ Optometrie & Fashion*, 09/2012, 58-62.

Nagel, J. (2012). *Neues Konzept für die bedarfsgerechte Energieversorgung des Künstlichen Akkommodationssystems*. Karlsruhe: KIT Scientific Publishing.

Pötschke, M., & Simonson, J. (2001). Online-Erhebungen in der empirischen Sozialforschung: Erfahrungen mit einer Umfrage unter Sozial-, Markt- und Meinungsforschern. *ZA-Information / Zentralarchiv für Empirische Sozialforschung*, 49, 6-27. Köln: SSOAR, GESIS – Leibniz Institute for the Social Sciences.

Rauschenberger, M., Schrepp, M., Cota, M.P., Olschner, S. & Thomaschewski, J. (2013). Efficient measurement of the user experience of interactive products - How to use the User Experence Questionnaire (UEQ).Example: Spanish Language Version. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 2, No. 1, 39-45.

Schelle, H. (2000). *Kontaktlinsen: Neues Sehen – selbst erleben*. Stuttgart: Trias-Verlag.

Schnell, D., & Khaireddin, R. (2013). Hornhautstoffwechsel unter Kontaktlinsen im Sport. *Der Ophthalmologe, 6/2013*, 502-510.

Embodied Virtual Agents and Electronic Bracelet to support independent Travel by People with Cognitive Disabilities

Pedro Ribeiro[1], Anna Michel[1], Christian Ressel[1], Stefan Koenen[1], & Sebastian Schmitz[2]
[1]Rhine-Waal University of Applied Sciences
[2]Hochschule Niederrhein, University of Applied Sciences
Germany

Abstract

Individuals with cognitive impairments have difficulties to independently use public transport. Some studies suggest that independence can be improved through effective orientation instructions and by reducing the travel anxiety. An Embodied Virtual Agent (EVA) as a travel assistant has the potential to provide valuable support. In this paper, we explore EVA capabilities, as a mean to improve the independence of individuals with cognitive disability, within a personal navigation application for mobile devices. We explore EVA ability to provide location-aware instructions that reassure users. Besides, we use an electronic bracelet (EB) equipped with an RFID-Reader to read mounted tags at the bus stop. The EB is able to effectively alarm the user for important EVA's instructions. Furthermore, it allows the user to check if a specific bus stop is correct in the context of the current trip and therefore helping in the decision making. Findings suggest that EVA and EB help to reassure and effectively perform the tasks inherent to a trip.

Introduction

Public transport plays an important role in promoting the social inclusion of people with cognitive disabilities, affecting their access to healthcare, work, education, social relationships and other basic services. The access to public transport is particularly crucial because many people of this vulnerable group typically are not able to use a car. There are, however, transport barriers that affect in different ways people with cognitive disabilities preventing them to independently use public transport. These barriers are for example related with their capacities to understand/remember transport information. The relevance of this topic has been stated most recently as a part of a European Parliament study (Lodovici & Torchio, 2015), which identify people with cognitive disabilities as a group with risk of social exclusion, and ultimately points to the lack of research in the area.

Research on assistive technology, identified needs and solutions for navigation systems that aim to support the bus usage by people with cognitive disabilities. Lemoncello et al. (2010) studied wayfinding performance of participants with

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

cognitive disabilities and found that they perform navigational tasks with greater hesitancy and errors than non-disabled individuals. The authors concluded that assistive devices must provide concrete/explicit orientation instructions complemented with salient landmarks (Davies et al. 2010). They also recommended that navigational assistance must include the capacity to reassure the user in order to reduce his hesitancy/anxiety. Risser et al. (2012) identified several barriers that restrict the independent outdoor mobility, and concluded that the complexity of the environment, in combination with the lack of self-confidence results in uncertainty and fear to travel alone. Livingstone-Lee et al. (2014) examined 159 personal navigation applications (PNAs) and considered that only seven are suitable to assist people with cognitive disabilities. They recommended features for the future transit apps and highlighted that a PNA must be capable of reassuring users.

Smartphones offer a big potential to improve the navigational assistance (Livingstone-Lee et al. 2014). The tendency is that in future most of cognitive disabled persons will have experience in using smartphones. Furthermore developers have nowadays more freedom to create solutions because these devices offer a great computation and storage capability. This is an opportunity to explore the capabilities of an Embodied Virtual Agent (EVA) as a travel assistant that is capable to provide understandable instructions and reassure users. EVAs are virtual beings that interact with users through an expressive virtual body, capable of performing a natural verbal and non-verbal communication. A substantial body of work exists on the EVA functions. Prendinger et al. (2005) explored the emotional contagion effect and concluded that the presence of an empathic EVA can reduce the stress of solving a difficult task. Hone (2006) obtained a similar result maintaining that EVAs displaying sympathy or empathy can reduce the user's frustration. Hone (2006) also concluded that a female EVA is more effective in the emotional influence than a male EVA. Additionally, EVAs are able to influence the motivation and enjoyment to use a technological system (Heerink, et al. 2008). Factors that impact this motivational ability are the social presence, the EVA capacity to smile, express emotions and to display a social behaviour in the technical system (Morandell, et al. 2008; Bickmore, et al. 2009; Heerink, et al. 2010). A very important and unique capacity of the EVA is the non-verbal communication. Body gestures, gaze behaviour, facial and emotional expressions provide extra information to the user. This capacity may (1) affect the decision making (Melo, 2010); (2) enhance the communication, by overriding verbal communication (Krahmer & Swerts, 2007); and (3) maximize the memorization (Buisine & Martin, 2007). These functions suggest that EVAs have the potential to support individuals with cognitive disabilities and help them to overcome barriers to transportation use. An EVA can help users in their travelling activities by employing easy language explanations or reminding the user an important task and ultimately to reassure them.

The EVA's potential can be explored via smartphone. Nevertheless the alarming capacity of a smartphone can be insufficient e.g. vibration or beeping not perceived when smartphone is in a bag. Thus it's important to complement the smartphone with a freehand mean that is able to effectively alarm the user for important EVA's instructions. Currently available smart watches are undoubtedly a freehand device, capable to be synchronized with the smartphone, but far too complicated for users with mental disabilities. Additionally, the displays of these devices are difficult to read under strong sunlight. Therefore, it is considered necessary to build an EB from the scratch that integrates bright LED arrays to overcome the reading problems, a vibration unit to ensure an effective alarm and a Bluetooth LE unit for the communications with the smartphone. It is also considered important to integrate a RFID reader that allows a contactless identification of bus stops, because GPS based localization is not sufficient for example in subway stations. Furthermore GPS based systems often locate users on the wrong side of the street.

The purpose of the current study is to build a PNA and explore the Embodied Virtual Agent and electronic bracelet capacities to provide location-aware instructions that reassure users, and fundamentally help individuals with cognitive disability to independently use public transit.

Method

To investigate the effect of PNA on improving the independent use of public transport by people with cognitive impairments, a within-subjects design was chosen.

Participants

A sample of 7 adults was recruited to participate in this study (2 females and 5 males). Their average age was 31.14 years (*Range*= 24 to 48, *SD*= 8.35). All participants had light to medium mental retardation and epilepsy. Some participants were capable of reading at least words and had impaired speech capabilities. Some participants also present motoric problems. All participants provided written consent for participating in the study. It was made clear to the participants that all study data is confidential.

Materials

The study for testing of the PNA prototype in a real scenario rely on the insights/findings of previous explorations/analysis 1) observation study to identify the barriers in the mobility of the target group, and 2) employing participatory design methods to develop the PNA prototype.

Observation of barriers

Solving mobility problems means to get contact with environmental factors that might affect users. Accordingly, a qualitative field research study was conducted to determine and prioritize barriers in the mobility of individuals with cognitive disabilities. Ten subjects with different cognitive impairments did a leisure trip using public transport. Subjects were observed and the mobility obstacles, that prevented the participants to use public transport, were documented on the basis of the evolved observation sheet. A video-record from two different perspectives was also created for each participant (Figure 1). The data were analysed and the prioritization of the categories was determined by frequency count. The classification of the observed barriers ensued on foundation of the knowledge of Human Factors, who assumed that deviations in exogenous and endogenous factors might misconduct the independent using of public transport and be found at all levels of interaction between individuals and the complex system of public transport (Badke-Schaub, Hofinger & Lauche, 2012; Marquardt, Gades, Robelski & Höger, 2010; Wickens, Lee, Lui & Becker, 2004). Regarding the mobility barriers reasoned in cognitive processes the study indicated determining factors in the lack of temporal and spatial orientation, attention, and decision making. Furthermore, the outcomes show that barriers in the emotional sphere as anxiety, stress, insecurity, and lack of motivation are ponderous. Accordingly, there is a demand to reduce the uncertainty, to modify the decision making and motivation to use public transport.

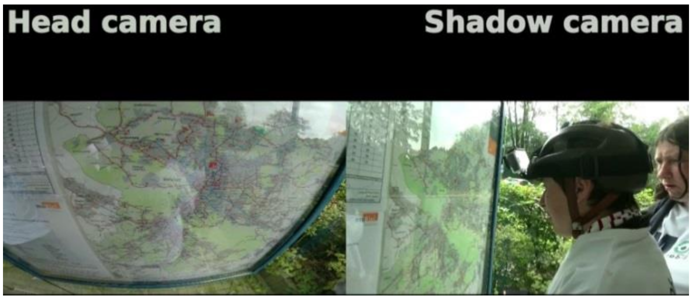


Figure 1. Videoed barriers observation.

Participatory design

In order to discuss the PNA´s mockup a focus group session with five subjects was conducted. The function of the PNA is to provide understandable instructions to users in the various moments of a trip. For the mock-up, it was defined a travelling story with a single bus route that was supported with street view panoramas. Mockup's screens (Figure 2) were presented to the subjects on a laptop and discussed one by one.

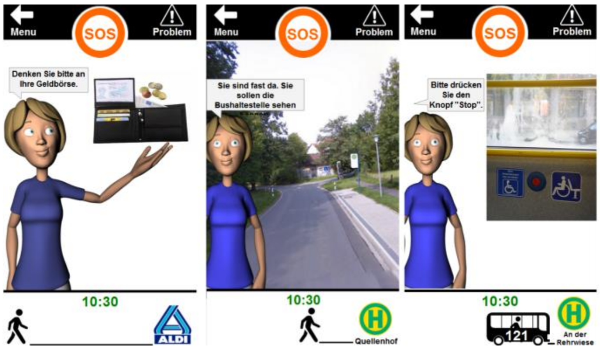


Figure 2. Mockup screens examples.

EVA's messages were in general well understood; the misunderstanding situations occasionally happened because of some unknown words. This fact reinforces the importance of using an easy language. It was also concluded that every informative communication shall be succeeded by instruction from EVA. The contextual image that complements the EVA verbal communication effectively reinforced messages. Subjects also considered the landmark images a good support to the pedestrian navigation. In most of the cases the progress bar (based on concrete elements) was correctly interpreted, participants were able to understand the estimated time arrival and describe the progress bar. Additionally, it was found that the progressive construction of the user interface proved to be a good way to enhance the user interface understanding (Figure 3).

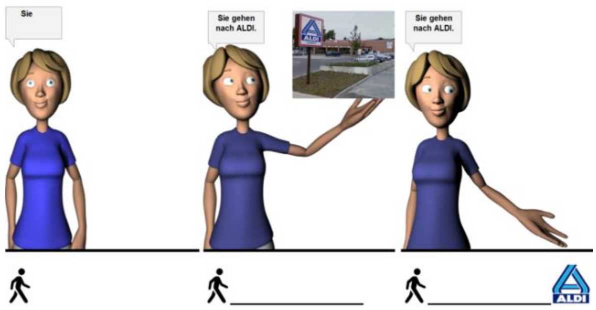


Figure 3. Progress bar construction.

In an exploratory session with 12 subjects, three focus groups were created to promote a discussion about the EVAs and the EB. With respect to the EVAs, it was found that subjects prefer a more visually realistic virtual agent. Participants understood correctly the facial expressions and realized how serious a message can be depending on the expression. Furthermore, subjects did not perceive difference between a perfect lip-sync and a simple lip-sync in which the speech audio starts synchronized with a random movement of the lips, and stops synchronized with the immobilization of lips. Some subjects considered the EVA's voice (text to speech based) unnatural. Finally, the nonverbal behaviour successfully enhanced the verbal communication with a bad audio quality. Regarding the EB, a green or red animation was used to inform the user that he is at the correct or wrong bus stop. This communication has been well understood by the subjects. The combination of a yellow animation and a vibration proved to be a reliable mean to alert user. Furthermore a LED bar and LED ring for displaying an amount of remaining time was shown to the subjects. However the usage of the same colour scheme for different purposes (e.g. red colour for wrong bus and less time) resulted in user confusion.

In a different session participants were invited to check a bus stop. This checking process consists of placing (1) the hand which wears the EB; or (2) the smartphone on a hand sign, resulting on the feedback: "correct bus stop" or "wrong bus stop" (Figure 4). The process was understood but the execution was difficult, because the EB or smartphone should be in contact with the surface, but subjects had no experience with that.

Figure 4. Left: Checking with electronic bracelet; Right: checking with smartphone.

Prototype System

The results obtained in the participatory design were considered for the construction of a functional prototype. This comprised two devices, the smartphone and the EB, and implemented implicit and explicit interaction. The implicit interaction was automatically triggered depending on the context. For instance, the instruction “press the stop button” was triggered near the destination bus stop (Figure 5). Implicit interaction always triggered the alert behaviour on the devices: yellow animation plus vibration on the EB, and an acoustic signal plus vibration on the smartphone. Explicit interaction was possible whenever the user places one of the devices on the hand sign installed in the bus stop (Figure 4), and by pressing the buttons on the smartphone screen (“hear instruction”, “repeat instruction” and “ok”). The devices were programmed to communicate via Bluetooth in order to trigger behaviours on the EB, for all events started on the smartphone and vice versa. The developed EVA was responsible for giving guidance. It was a female EVA with a human appearance able to perform verbal and non-verbal behaviours that were based on motion capture data. A female person was selected to record easy language messages to be used as the voice output of the EVA. Each message was recorded in two versions: 1) normal speed version; 2) slow version to be used whenever the user presses the repeat button.

Figure 5 . Starting from left: alert user; buttons to get the message; subject performing.

The prototype was able to provide support within a workflow: 1) identify the correct bus stop; 2) board the correct bus; 3) disembark at the correct bus stop. Accordingly, the prototype started in a pedestrian navigation context. When the bus stop was near, the EVA said “You are almost there” and the contextual picture of the correct bus stop was displayed. In the Bus stop area, the user was invited to check the bus stop. Then by placing the device on the hand sign the result was provided followed by an instruction (e.g. "wait for the bus"). The result of the check was reinforced by a colour scheme and non-verbal behaviour (e.g. stop gesture; Figure 6).

Figure 6. Starting from the left: EVA informing "correct stop"; EVA informing "wrong stop"; EVA instructing "correct bus stop is on the other side of the street".

While waiting for the bus, the EVA informed the estimated time of arrival and a progress bar was displayed (Figure 7). When the bus arrived, a picture of the bus was displayed and the EVA instructed the user to board. Then just before the destination stop, the EVA instructed the user to press the stop button (Figure 5). When the bus arrived the user was instructed to disembark.

Figure 7. Progress bar.

In this study it was used a Wizard-of-Oz version for the implicit interaction (location/schedule based events). Over the course of the trip the Wizard constantly observed: (1) subject's position; (2) time; and used a tablet computer to trigger behaviours via Bluetooth in the PNA.

Procedure

One month before the study a pre-test was conducted to verify the study method and train the participants to use the prototype. A well-known short route was chosen. Before the pre-test, the procedure was explained to the subjects. After, a HR monitor chest strap was presented and only installed by approval. The heart rate (HR) is one of the most simple and reliable stress and fear indicators (Hoyer et al, 2005; Wijsman, 2014). Finally the prototype usage was explained. The pre-test starts with a walk to the bus stop. In the bus stop area, the participants had checked the bus stop. Once the correct bus stop has been reached, the prototype instructed the participants to do a short drive by bus to a predefined bus stop. The pre-test has shown some procedure and usability problems such as missing the bus, long waiting times and not complying with the schedule. Hence some modifications were made for the study; for instance shortening the walk to bus stop and extending the bus trip. Furthermore, another training session was taken two days before the study.

For the evaluation it was selected a testing area where the participants were not familiar with the route. The duration of each evaluation session was approximately 25 min. The trip started with a walking part (2 min) to the bus stop area. There were two bus stops one on each side of the street (Figure 8). Buses came every 30 min and the waiting time for each participant ranged from 12 to 24 min. The bus trip took approximately 5 min.

Before the study each participant was equipped with an HR monitor chest strap. At the beginning of the study participant was briefed with the study procedure. The important travelling information like bus number or destination bus stop was provided. Then, a smartphone plus EB were handed and instructions on how to use the technology were given.

Figure 8. Origin bus stop area.

Each participant was accompanied by two researchers who lead the walking part until the bus stop area. Researchers just intervened to present the tasks and prevent dangerous situations. During the evaluation session, subjects were invited to solve four different tasks: (1) identify the wrong bus stop (before and after using the prototype) (Figure 9); (2) identify the correct bus stop (before and after using the prototype); (3) board the correct bus; (4) disembark in the correct bus stop, with the question "Is this the correct …?". After each task the question "How sure are you?" were asked. Sessions were videotaped with a chest mounted camera and the observer recorded the success vs. failure and the certainty level on a three-point scale (certain, partly certain, uncertain). Finally a post-questionnaire was conducted, comprising questions about the prototype acceptance and usability. The answers could be given on a three-point scale (agree, disagree, undecided).

Figure 9. Checking bus stop.

Results

SPSS was used to perform exploratory statistical analysis with the collected data to find out the significant differences in HR mean on the wrong and right bus stop before and after using the prototype. In the same way the frequency of the success and certainty was analysed in the four tasks. The HR mean was calculated for each task, using a 30 seconds sample, starting from the task presentation. To test the differences of the HR mean on the wrong and right bus stop before and after using the prototype, a 2 (bus stop) x 2 (pre and post measurement) within subjects and repeated measure ANOVA was run. The HR mean rises minimally at the right bus stop from 95.804 to 96.540. Regarding involving the effect of handling on the wrong vs. correct bus stop, the main effect was not significant (F (1) = .182; p = .688). However, there is a trend to significant contrasts effect (F (1) = 20.427; p = .006) before and after using the prototype. In comparison to HR mean before and after using the prototype the results shows that the HR mean decreases from 100.903 to 91.442.

Table 1 shows the cross-tabulation of success and certainty by task. Before checking the first bus stop with the prototype, three of the participants had success identifying the wrong bus stop, two participants failed and the other two participants did not answered. The well succeeded participants presented different levels of certainty and the two participants who failed affirmed to be sure. Furthermore, six participants succeeded in using the prototype and successfully accomplished the first task. One participant did not check the bus stop stating that the correct bus stop is on the other side of the street. From the six well succeeded participants, three affirmed to be sure and two present mid-level of certainty. Regarding the second task, before checking with the prototype, all the participants had success identifying the correct bus stop. Five participants affirmed to be sure and two showed mid-level of certainty. Furthermore, all the participants were able to use the prototype and successfully fulfilled the task. Due to a design limitation, the certainty level was not recorded after checking the correct bus stop with the prototype. In the third task all the participants identified the correct bus to board. Six participants were sure about their decision and one was unsure. Finally, all the participants disembarked the bus in the correct bus stop. In this last task six participants were sure and one participant shows mid-level of certainty.

Table 1. Cross-tabulation of success and certainty by task

	Identify wrong bus stop before using PNA	*Identify wrong bus stop after using PNA*	*Identify correct bus stop before using PNA*	*Identify correct bus stop after using PNA*	*Board correct bus*	*Disembark in the correct bus stop*
Success	3 (42.9%)	6 (85.7%)	7 (100.0%)	7 (100.0%)	7 (100.0%)	7 (100.0%)
Certainty	3 (42.9%)	3 (42.9%)	5 (71.4%)	-	6 (85.7%)	6 (85.7%)

Table 2 shows the frequency table of the acceptance questionnaire. The questions related with the performance expectancy and anxiety obtained a unanimous positive answer. Learning to use the PNA was considered difficult by one of the participants. Some participants reveal indecision or did not answer questions related with attitude toward using technology and facilitating conditions. Finally all participants considered it a good idea to use the PNA, and a majority showed intention to use the PNA in the future.

Table 2. Frequency table showing the acceptance questionnaire responses

	Agree	*Disagree*	*Undecided*
PNA is useful in my daily life	7	0	0
PNA enables me to accomplish a trip	7	0	0
PNA is easy to use	6	0	0
Learning to use the PNA is easy	6	1	0
Using the PNA is a good idea	7	0	0
PNA makes life more interesting	4	0	2
I like living with the PNA	4	0	1
I am able to use the PNA	5	0	2
I have enough knowledge to use the PNA	4	0	3
I feel apprehensive about using the PNA	0	7	0
It scares me to push the wrong button.	0	7	0
The PNA is intimidating to me	0	7	0
I want to use the PNA in the future	5	2	0

Table 3. Frequency table showing the usability questionnaire responses

	Agree	*Disagree*	*Undecided*
EVA helps me to accomplish a trip.	7	0	0
EB helps me to accomplish a trip.	7	0	0
Pictures on the screen were easy to understand.	7	0	0
I always knew which button to press.	5	0	2
It was easy to press the buttons.	7	0	0
I always understood the elements in the screen.	7	0	0
I always understood what the EVA said.	7	0	0
The EVA's voice was loud enough for me.	6	1	0
I found the EVA friendly.	7	0	0
The alarm (vibration and light) bothered me.	0	7	0
I know when the EVA has a message for me.	6	1	0

Table 3 shows the frequency table of the usability questionnaire. The EVA and EB were unanimously considered helpful. One participant considered that the volume level of the EVA voice must be increased. Other critical user interfaces were not considered problematic and the alarm event was well accepted.

Discussion

Findings suggest that the PNA’s prototype helps to improve the independence of individuals with cognitive disability, particularly to reassure and effectively perform the tasks inherent to a trip. In fact there is trend to significant effect before and after using the prototype that suggests absence of stressors, such as fear, after checking a bus stop with the prototype. Hereupon, it is considered important, as future work, to conduct a study by employing both experimental and control groups in order to compare differences in the physiological parameters of fear. Furthermore, testing and evaluating physiological parameters of fear and subjective fear assignment before and after using the PNA can increase the reliability of evaluating the certainty parameter. Additionally, it is important to analyse other situations where participants feel unsure.

Results from the questionnaire suggest that users do not feel anxious while using PNA. They consider it a useful tool that enables them to accomplish a trip. Participants rate the use of the PNA and learning to operate the PNA as easy. Some participants state to have enough capacities to use PNA. Connecting these results with the free comments suggest that more training is mandatory. Training can influence the confidence and ability in using the PNA and increase the desire for prospectively using the PNA. Regarding the usability of the prototype, the results indicate an unambiguous positive appraisal of PNA with a few exceptions in 1) sound 2) alarm perception and 3) which button to press. The last two points can be achieved via training.

Generalization of the aforementioned findings is difficult. For example, the researchers’ presence certainly affected the subjects’ behaviour (e.g. subjects looking to the researchers expecting answers or help). The small size of the group is also an important limitation and a larger sample is mandatory for further tests. In fact the sample size was also affected by the withdrawal (due to illness) of some participants. Finally, further investigations should be carried out to relate stress level with the execution efficiency. Following this idea it would be interesting to integrate a stress sensor in the EB that could influence the EVA´s behaviour depending on the context and stress level. This research has practical value, and if validated it might enhance the quality of life of cognitively impaired people, and would facilitate mobility inclusion for this vulnerable group of public transport users (Sherman & Sherman, 2013).

Acknowledgement

This research was conducted within the "Mobile - Mobil im Leben" project funded by the Federal Ministry for Economic Affairs and Energy as part of the programme: ’“Von Tür zu Tür” – Eine Mobilitätsinitiative für den Öffentlichen Personenverkehr der Zukunft’.

References

Badke-Schaub, P., Hofinger, G., & Lauche, K. (2012). Human Factors. In P. Badke-Schaub, G. Hofinger, and K. Lauche (Eds.), *Human Factors. Psychologie sicheren Handelns in Risikobranchen* (pp. 4-20). Berlin Heidelberg: Springer.

Bickmore, T.W., Pfeifer, L.M., & Jack, B.W. (2009). Taking the time to care: Empowering low health literacy hospital patients with virtual nurse agents. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1265-1274). New York: ACM.

Buisine, S., & Martin, J.C. (2007). The effects of speech-gesture cooperation in animated agents' behavior in multimedia presentations. *Interacting with Computers, 19*, 484-493.

Davies, D.K., Stock, S.E., Holloway, S., & Wehmeyer, M.L. (2010). Evaluating a GPS-based transportation device to support independent bus travel by people with intellectual disability. *Intellectual and developmental disabilities*, *48*, 454-463.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2008). The influence of social presence on enjoyment and intention to use of a robot and screen agent by elderly users. In *The 17th IEEE international symposium on robot and human interactive communication* (pp. 695-700). Piscataway: IEEE.

Heerink, M., Kröse, B., Evers, V., & Wielinga, B. (2010). Assessing acceptance of assistive social agent technology by older adults: The almere model. *International Journal of Social Robotics*, *2*, 361-375.

Hone, K. (2006). Empathic agents to reduce user frustration: The effects of varying agent characteristics. *Interacting with Computers*, *18*, 227-245.

Hoyer, J., Helbig, S., & Margraf, J. (2005). *Diagnostik der Angststörungen*. Hogrefe Verlag.

Krahmer, E.J., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of memory and Language, 57*, 396-414.

Lemoncello, R., Sohlberg, M.M., & Fickas, S. (2010). When directions fail: Investigation of getting lost behaviour in adults with acquired brain injury. *Brain Injury*, *24*, 550-559.

Livingstone-Lee, S.A., Skelton, R.W., & Livingston, N. (2014). Transit Apps for People With Brain Injury and Other Cognitive Disabilities: The State of the Art. *Assistive Technology*, *26*, 209-218.

Lodovici, M.S., Torchio, N. (2015). *Social Inclusion in EU Public Transport*. Brussels: European Union.

Marquardt, N., Gades, R., Robelski, S., & Höger, R. (2010). Determinanten menschlicher Fehler in Risikounternehmen. In R. Trimpop, G. Gericke & J. Lau (Eds.), *Psychologie der Arbeitssicherheit und Gesundheit: Sicher bei der Arbeit und unterwegs wirksame Ansätze und neue Wege* (16. Workshop). Kröning: Asanger

Melo, C., Carnevale, P., & Gratch, J. (2010). The influence of emotions in embodied agents on human decision-making. In J. Allbeck, N. Badler, T. Bickmore, C. Pelachaud, and A. Safonova (Eds.), *Intelligent virtual agents* (pp. 357-370). Berlin: Springer.

Morandell, M., Hochgatterer, A., Fagel, S., & Wassertheuer, S. (2008). Avatars in assistive homes for the elderly. In M. Morandell, A. Hochgatterer, S. Fagel, and S. Wassertheurer (Eds.), *HCI and Usability for Education and Work* (pp. 391-402). Berlin:Springer.

Prendinger, H., Mori, J., & Ishizuka, M. (2005). Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. *International Journal of Human Computer Studies*, *62*, 231-245.

Risser, R., Iwarsson, S., & Ståhl, A. (2012). How do people with cognitive functional limitations post-stroke manage the use of buses in local public transport? *Transportation research part F: traffic psychology and behaviour*, *15*, 111-118.

Sherman, J. & Sherman, S. (2013). Preventing Mobility Barriers to Inclusion for People With Intellectual Disabilities. *Journal of Policy and Practice in Intellectual Disabilities*, *10*(*4*), 271-276.

Wickens, Ch.D., Lee, J., Lui, Y. & Becker, S.G. (2004). *An Introduction to Human Factors Engineering.* New Jersey: Pearson Education.

Wijsman, J.L.P. (2014). *Sensing stress: stress detection from physiological variables in controlled and uncontrolled conditions*. University of Twente: Twente University Press.

An idiographic study into the physiology and self-reported mental workload of learning to drive a car

Matthijs L. Noordzij[1], Serena M. Dorrestijn[1], & Irma A. van den Berg[2]
[1]University of Twente
[2] Irma van den Berg Tekst- en verkeersproducties
The Netherlands

Abstract

A driving instructor has to monitor the performance and state (e.g. mental work load) of the pupil who is learning to drive. However, the instructor is also responsible for road safety. Therefore, it might be beneficial when additional monitoring technology would be available to the driving instructor. Fluctuations in skin conductance are indicative of changes in the autonomic nervous system and have been operationalized as changes in stress or mental workload. For the present idiographic study six pupils were followed during their driving training, while measuring their self-reported (and by their driving instructor) workload and their skin conductance levels (with a wrist-worn bio-sensor). The quality of the physiological measurements was acceptable in most cases. Most students showed their highest physiological values 3-7 lessons before their final exam. The driving instructor was good at predicting the self-reported workload of her pupils. Importantly, there was no correlation between physiological fluctuations and fluctuations in self-reported workload. This makes skin conductance measurement unsuitable to replace subjective workload assessments. The physiological data did provide first evidence that a modular driving instruction methodology, with several partial exams, seems to prevent extremely high physiological activity during the final exam.

Introduction

A driving instructor has to closely monitor elements in the environment and in the car related to road safety, as well as the learning progress and capacity of the pupil. These tasks are performed up to 8 hours a day and can be considered demanding monitoring tasks where the instructor has to stay vigilant. It is well known that performance on vigilance and monitoring tasks can strongly diminish after relatively short times (under 30 minutes) (Warm, Parasuraman, & Matthews, 2008). At present, wearable (i.e. applicable on the body) technology methods are becoming available to continuously monitor physiological measures (e.g. heart rate), which might be helpful for a driving instructor to keep insight into the physical or mental state of the pupil. However, it is unclear whether these measures can, to some extent, replace the personal assessment of an instructor or if these measures are actually complementary and might provide previously unavailable, continuous insight into a student's state. The present study examines the patterns of self-reported workload

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

and physiological measurement from a new type of wrist-worn bio sensor which measures electrodermal activity (EDA) during real life driving lessons.

A construct that is often used in Human Factors is mental workload. As a first approximation mental workload can be described as academic jargon for 'mental effort'. It is situated within the information processing model of human cognition (Vidulich et al., 2010) and is defined by Parasuraman et al. (2008) as "*the function relating the mental resources demanded by a task and those resources available to be supplied by the human operator*" (pp. 145–146). For decades researchers have been studying the mental workload of drivers. It is beyond the scope of this paper to exhaustively introduce theories on mental workload and how fluctuations of mental workload occur in different driving situations (for excellent reviews see for example De Waard, 1996, and Vidulich et al., 2010). It is clear that learning to drive is a complex task with many different demands, which are placed on the novice driver at the same time. During driving unexpected events that have to be dealt with immediately occur often (e.g. traffic jams, road blocks and detours). Taken together it is clear that learning to drive relies heavily on perceptual and cognitive abilities, and that risks of cognitive overload are greater for student and novice drivers (Ross et al., 2012). Therefore, measures of workload during the crucial phase of learning to drive a car could be very welcome to both driving instructor and student.

A simple, but reliable and valid (Verwey & Veltman, 1996) way to measure self-reported workload, is via the Rating Scale of Mental Effort (RSME) (Zijlstra, 1993). This unidimensional scale asks participants to rate the effort they expanded on a task they have just carried out. Our study is the first, to our knowledge, that follows student drivers in a longitudinal manner through several driving lessons and their exam, and measures RSME values for lessons. This makes it difficult to make precise predictions on how student drivers and their instructors will report on the mental effort of the lessons. Our first exploratory research question pertains to the type of patterns we will find for self-reported mental workload during driving lessons. Although precise predictions are difficult we would like to make two suggestions based on (1) literature on novice drivers and (2) the type of driving training our participants receive. First, it is well known that over time people automate the visuo-motor control aspects of driving (e.g. shifting gear), as well as habituate to certain driving situations and scenario's (e.g. Verwey, 2000). From this it could be expected that students will show declining levels of self reported mental workload over lessons. Second, the present study follows student drivers who are being instructed according to a modular driving instruction methodology ('Rijopleiding in Stappen'). A traditional, non-modular system (still most common in the Netherlands) would have a final exam in which effort and stress levels are expected to be very high. In contrast, the modular system tries to shift this moment with several partial exams. When approximately 75% of the lessons are finished most of the required skills can be shown and passed in a final partial exam. The actual final exam is presented more as a formality, which all students are expected to pass. From this it could be predicted that students might show their highest mental effort 3-7 lessons prior to the final exam. Next to self-reported measures this study also includes EDA as a possible physiological index of workload.

EDA refers to all electrical phenomena in the skin, most often expressed in skin conductance (SC) units (Boucsein, 2012), which are typically measured at the fingers or the hand palm. The SC signal can be described as small, short waves (Skin Conductance Responses (SCRs)) riding on a larger wave (the Skin Conductance Level (SCL)). Fluctuations in the occurrence and amplitude of SCRs are caused by activity of the eccrine sweat glands, which are innervated purely by the sympathetic nervous system. Typical values (mostly from lab studies) are between 0 to 20 SCRs per minute and an average SCL of 2 μSiemens (μS) (Dawson, Schell, & Filion, 2007).

EDA has been systematically applied in traffic research (Boucsein, 2012). It is often mentioned as one of the methods to do online physiological approximations of workload, but consensus is that it is a rather general measure of arousal or stress (e.g. Healey & Picard, 2005). Recently, it has become possible to measure EDA at the wrist in an ambulatory non-intrusive fashion with no wires, thereby minimizing motion artifacts (Poh, Swenson, & Picard, 2010). The second research question is what the pattern of skin conductance parameters over lessons will be from a wrist–worn EDA sensor. Similar to the RSME values we will explore the data over time for the presence of consistent patterns.

Next to the central research questions concerning the possible RSME and skin conductance patters during driving lessons three related side issues will also be explored. (1) The standard location to measure EDA is the fingertips or the hand palm. For this study we will measure EDA at the wrist. Although this study is by no means a validation study for the wrist-worn sensor we will be able to find out whether relevant skin conductance parameters (i.e. number and amplitude of SCRs and the SCL) can be extracted from data from a wrist-worn bio sensor during driving lessons. (2) In addition we can test whether any of the skin conductance parameters correlate with self-reported RSME values. (3) Finally the driving instructor will also assess how mentally effortful a lesson was for her student. A previous study had suggested that self-report and report by a driving instructor had medium correlations (Victoir et al., 2005). We will try to replicate this finding in our study.

To explore the research questions an idiographic approach (i.e. the science of individualized measurements) is employed (Molenaar, 2004). Instead of measuring data for a great number of student drivers for one lesson, a limited number of students is followed during a number of lessons (including their exam). There are several reasons why such an idiographic approach is appropriate for our research questions. The individual differences in relation to fluctuations in both self-reported measures and physiology are large, but individuals might have clear patterns over time. Moreover, our research is exploratory and would benefit from extensive individual data, which subsequently could inform more hypotheses driven nomothetic studies (e.g. studies in which group means are compared such as a randomized controlled trial). Finally, as stated by Picard (2009, p. 3580), and potentially very relevant for individuals looking for insight in the physical and mental effort during driving lessons: '*with an individualized data-intensive approach based on measurement in a person's natural environment it is not just the* *science that benefits: each participant can now benefit with information specific to his or her needs and situation*'.

Method

Participants

Six students (four female and two male aged between 17 and 22 years) from driving school Irma van den Berg (www.irmavandenberg.nl) were recruited via convenience sampling.

Materials

Q sensor
The Affectiva Q™ sensor is a wrist worn, watch-like sensor which measures EDA with 1cm diameter Ag-AgCl dry electrodes at the ventral side of the wrist. In addition to EDA, actigraphy and skin surface temperature data are also logged at a sampling rate of 32Hz.

Rating Scale of Mental Effort (RSME)
The RSME (Zijlstra, 1993) is a unidimensional, reliable instrument to measure self-reported mental workload. The RSME is a 15 cm long line (150-points) with markings at every centimetre, and nine anchor points with verbal labels going from "absolutely no effort" (around the 0 point) to "extreme effort" (around the 112^{th} point on the scale). Answers can be given by marking the line at a point that fits the experienced workload.

Design & Procedure

All students (or their legal caretakers for students younger than 18 years) signed an informed consent. All participants were explicitly told that the measurements would have no influence on their driving lessons and that they were free to remove the bio sensor at any time. In addition, a female driving instructor (57 Years old, 27 years' experience as a driving instructor) acted as instructor during all the lessons.

As mentioned above, the present field study has an idiographic and observational approach and design. A limited number of driving students were followed intensively during (parts of their) driving lessons (including their driving exam) in the same Mitsubishi Outlander with a manual gear box. The driving lessons occurred as they would have normally, except for the following differences: 1) at the start and at the end of the lesson both the student and the instructor filled in the RSME. The student was asked to fill in the RSME at the start to indicate how effortful the task was he or she had been doing before the driving lesson began (not used for this present paper), and to indicate on the RSME at the end of the how effortful the driving lesson as a whole had been. The instructor was asked to fill in the RSME estimating the effort of the student during the driving lesson (thus not her own effort). The student and instructor were blind to each other's rating. 2) At the start of the lesson the Q sensor was attached to the left wrist by the student with the instruction to get a tight, but comfortable fit (using an easy Velcro band).

The left wrist was chosen to minimize the influence of movement, which was deemed more likely on the right wrist due to movements related to (manual) gear changes. The left hand is mainly involved in smooth and slow turning motions probably making the influence of motions on the physiological measurements rather negligible (Kappeler-Setz et al., 2013). An exception might be special manoeuvres which require a quicker steering. Below (under the heading 'Motion Artefacts') a description is provided on how the actigraphic data was correlated to the EDA data to check whether motion fluctuations are associated with EDA parameter fluctuations.

No pressure was put on the instructor to use the Q sensor during any given lesson or the exam. Furthermore, there was no experimenter present in the car during the lesson or the exam, making the driving situation identical to a normal driving lesson or exam.

Data analysis

Skin Conductance parameters
EDA data were down sampled to 16Hz, and pre-processed with a Continuous Decomposition Analysis (CDA) as implemented in Ledalab (Benedek & Kaernbach, 2010), which requires MATLAB (www.mathworks.com). From the CDA an estimate of the skin conductance level (SCL) was acquired. The phasic activity coming from classical Through-to-Peak analysis was reported (threshold for an SCR amplitude was set at .03 µS) (Boucsein, 2012).

As recommended (Boucsein, 2012), visual checks were performed on plots of skin conductance data to identify failed measurements, "non-responding" (indicated by an absence of SCRs in a given measurement), and incorrect classification of SCRs. Data from these problematic measurements were removed from further analysis.

The SCL and SCR parameters were expressed in a number of variables both at the **minute** level and at the **lesson** level. For every individual minute the mean SCL, number of SCRs and total amplitude of SCRs were calculated. In addition, mean SCL, mean number of SCRs per minute, and mean total amplitude per minute were calculated for every lesson. By also aggregating the skin conductance parameters per lesson, the physiological and self-reported RSME scores could be correlated, because the latter only concerns a complete lesson. Only the RSME score provided at the end of the lesson was taken into consideration.

Motion artefacts
To check whether fluctuations in motion were associated with fluctuations in skin conductance parameters, the total magnitude of the vector of acceleration (total acceleration per minute (t_a)) was calculated from the actigraphic data via $t_a = \sqrt{a_x^2 + a_y^2 + a_z^2}$. Subsequently, t_a was correlated with the skin conductance variables (per minute).

Results

Quality of the EDA data and motion artefacts

We first checked whether the wrist worn sensor delivered usable data and whether the EDA parameters were associated with motion fluctuations. The Q sensor was applied during 99 lessons of which 86 (85%) resulted in EDA measurements that were suitable for subsequent analyses (Figure 1A). Total EDA measurement time was 7988 minutes, and the mean lesson time was 86 (*SD*. = 17.5) minutes with a range between 30 and 147 minutes. In Figure 1 EDA measurements are shown for two driving lessons (panel A and Panel B visualize the EDA data for two complete lessons). Most lessons were associated with typical EDA patterns (Figure 1A). Some lessons had failed measurements with no clear changes in skin conductance (Figure 1B, 15% of the data).

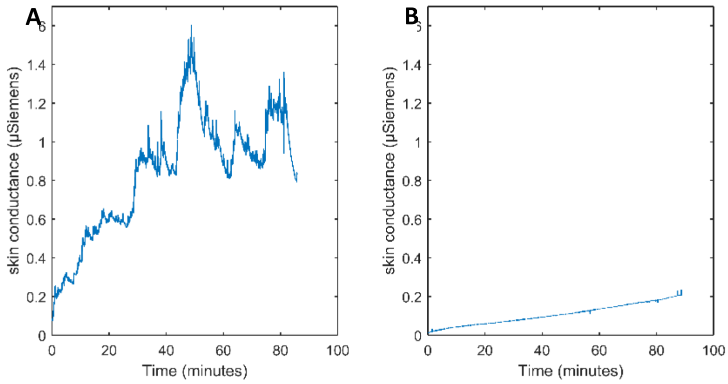


Figure 1. A) An example of a typical successful measurement during a driving lesson depicting both variation in SCL and SCRs. B) A failed measurement due to "non-responding" of the participant. On the x-axis is Time (minutes), and on the y-axis is skin conductance level (μSiemens).

There were no correlations between fluctuations in the total acceleration (t_a) and skin conductance variables calculated per minute (see Table 1). In contrast, as can be expected, the skin conductance variables showed strong to very strong correlations to one another.

*Table 1. Correlations for acceleration and skin conductance variables per minute (all p's < 0.001 marked with *, n (minutes) = 7277)*

	Acceleration	*Number of SCR*	*Amplitude (SCR)*	*SCL*
Acceleration	-	-0.02	0.004	0.07*
Number SCR	-0.02	-	0.86*	0.67*
Amplitude SCR	0.004	0.86*	-	0.57*
SCL	0.07*	0.67*	0.57*	-

RSME values and Skin conductance parameters: Possible patterns over lessons

Figure 2 gives insight in how skin conductance parameters (Fig. 2 A and 2B) and RSME values (Figure 2C) fluctuate over lessons and between individuals. Our two research questions concerned patterns in RSME or SC values. There were strong variations in RSME values between lessons and students (see Figure 2C). Overall there was a positive correlation between lesson number and RSME values, $r(93) = .37$, $p < .001$, yet, this correlation differed strongly between students, from $r = -.68$ to .42. For almost all participants (one participant had a correlation of -.1) there were weak, to medium, positive correlations between lesson number and the skin conductance parameters, all r's between .05 and .48, all p's < .001.

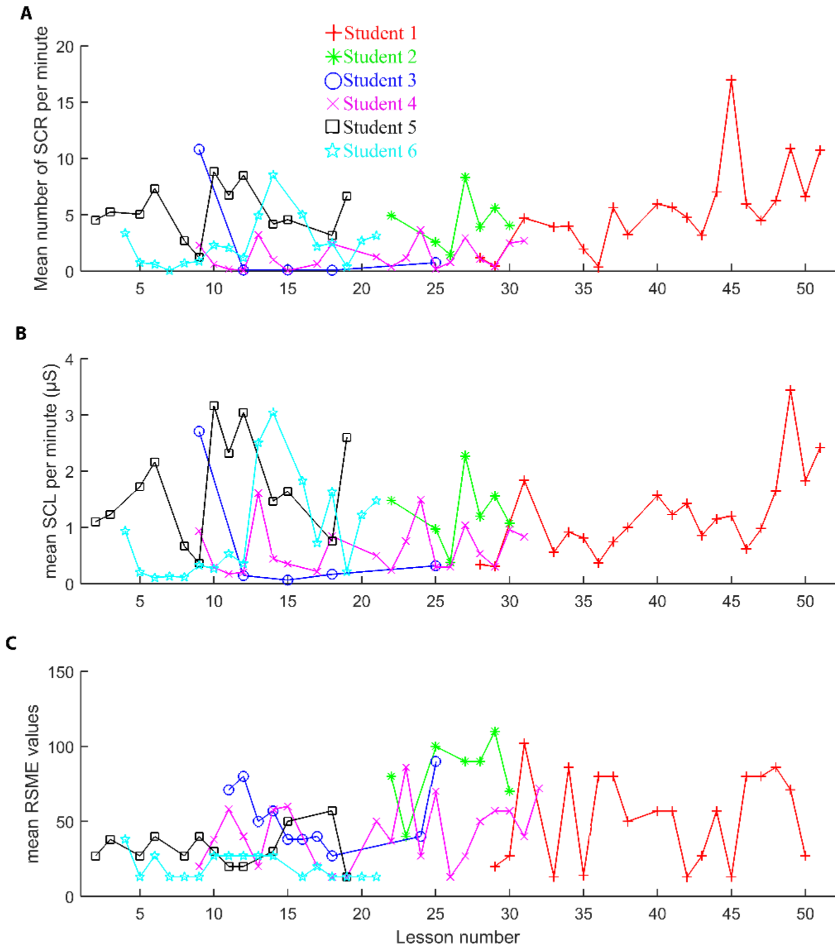


Figure 2. Variations of mean number of SCRs(A), mean SCL (μSiemens) (B), and mean RSME values per lesson (C), per student. The data for the mean amplitude of SCRs are not depicted because it follows a very similar pattern to the number of SCRs data (see also the high correlations in Table 1).

Participants 1, 2, 4 and 5 showed their highest skin conductance values in the 3-7 lessons prior to their driving exam. Importantly, five out of six students passed their driving exam (student 6 stopped taking driving lessons altogether). Student 3 was no longer followed with the Q sensor after lesson 18 because of very low SCR values (virtually non-responding), and only the final exam was measured (so for 4 participants we had a complete data set).

We also tried to replicate earlier research that found medium correlations between self-reported workload by the student driver and the driving instructor. The evaluations by the instructor of the experienced workload of her students in our study correlated positively with the evaluations of the students themselves, $r(70) = .8$, $p < .001$ (range = .14 to .89, see Table 2).

Table 2. Descriptive statistics for RSME values as reported by students themselves and as assessed by the instructor. The number of lessons (n) with valid RSME values is indicated after the student number. The RSME exam column indicates what value students gave for their driving exam (student 1 did not rate the exam and student 6 stopped taking driving lessons).

Student number	*Range student*	*Mean student (sd.)*	*RSME exam*	*Range instructor*	*Mean instructor (sd.)*
1 (n = 20)	13 - 102	52 (30.1)	-	13 - 108	59 (36.4)
2 (n = 7)	40 -110	83 (22.9)	70	27 - 118	86 (32.5)
3 (n =9)	27 - 90	49 (17.3)	90	50 -101	66 (21.5)
4 (n = 21)	13 - 86	41 (20.7)	72	13 - 101	43 (25.6)
5 (n =12)	13 - 57	34 (11.5)	13	10 - 71	40 (22.3)
6 (n =16)	13 - 38	20 (8.2)	-	10 - 57	37 (16.1)

Finally, the possible correlation between the reported RSME values and the physiological patterns were explored. Overall the fluctuations per lesson in self reported RSME values by the student and skin conductance parameters did not correlate, all r's(85) between .01 and .05, all p's > .66. In addition, the RSME values indicated by the instructor also showed no correlation with the skin conductance parameters of the student, all r's(65) between -.13 and -.03, all p's > .3. This total lack of correlations was also found when only the last ten minutes of SC data (the minutes close to the RSME measurement moment) of every lesson were taken into account, all r's(85) between -.09 and -.05, all p's > .43. For one individual student (number 4) a correlation between mean number of SCR per minute and RSME values was significant. However, this correlation was negative, $r(22) = -.49$, $p = .03$.

Discussion

This idiographic study set out to explore physiological and self-reported indices of workload of young people learning to drive. The main results were that (1) relevant skin conductance parameters (with biologically plausible values) could often be extracted successfully from EDA signals acquired from a wrist-worn sensor during a real life driving lesson or exam. (2) Correlations showed that driving students' self-reported workload and skin conductance parameters increased, instead of decreased as the driving lessons progressed. However, for the self-reported workload this was highly student dependent with some students showing a negative correlation. In addition, visual inspection seemed to suggest that students reached their (mean) peak physiological values 3-7 lessons before the driving exam after which their values started to decrease. (3) There was absolutely no correlation whatsoever between self-reported (both by the student and the instructor) workload and skin conductance parameters. Below further elaboration will be provided on these findings in the light of the feasibility of ambulatory EDA measurements, Human Factors theory, and practical recommendations for the integration of EDA measurements during driving lessons.

EDA measurements at the wrist with dry electrodes seem technically possible and easy in a real-world driving lesson scenario. At present, the state of the art concerning EDA measurements comprises of measurements at fingers or in the palm with electrode gel (Boucsein et al., 2012). Our data showed that in 85% of the cases the EDA data patterns were according to what can be expected with clearly identifiable SCRs and very little motion artefacts. There were individuals that had quite a few non-responding sessions, but it must be noted that even when following the state of the art there will always be participants who exhibit a total lack of SCRs (Boucsein, 2012). The range of the number of SCRs per minute, and the average SCL were in a biologically plausible range (between 0-20 per minute, and near 2 µS, Dawson et al. (2007)). In addition, fluctuations in movement as captured by the accelerometer did not correlate at all with fluctuations in skin conductance parameters, providing evidence that motion artefacts were not a major problem in this case. The EDA signals in this study were not directly compared to a measurement at the fingers. However, other studies have done this comparison in a lab setting and showed that the wrist measurements are very comparable to the finger measurements (Poh et al., 2010). The present study adds to this lab validation a field validation of EDA wrist measurement in a particular context: Learning (how) to drive in a naturalistic setting.

Skin conductance parameters and RSME values (Zijlstra, 1993) did not steadily decline over the course of various driving lessons. The physiological values actually increased until shortly before the driving exam, after which they seemed to decrease. Decades of research into (cognitive) workload and habituation during driving lead to the idea that the repetition of an activity could lead to initial high values associated with a very explicit knowledge driven phase, after which more automatic cognitive-visuo-motor associations would form leading to less workload (e.g. Collet et al., 2009; De Waard, 1996; Dijksterhuis et al., 2011; Verwey, 2000). The present study suggests that the knowledge obtained from (novice) drivers might not be completely generalizable to student drivers. Up to 3-7 lessons prior to their (successful) driving exams skin conductance parameters increased. Only then did they start to decrease. This pattern in the skin conductance parameters fits with the basic assumption from the modular driving methodology the students were following, where the most important (partial) test is taken when 75% of the lessons have been completed. The goal is to make the final exam less stressful and more of a formality. The physiological data provide evidence that this is indeed working. The RSME values were less clear with some students still showing the highest value for their final exam. Obviously, this value of the modular methodology needs to be verified with a larger, and importantly a more diverse sample of students passing, but also failing their exam, while following different driving instruction methodologies. However, it is exactly this kind of interesting pattern one can identify within an idiographic study, and subsequently test in larger (and more expensive) idiographic and also (perhaps at a later time) nomothetic studies.

It was found that the instructor was very good at predicting the RSME values of the participant. This is in accordance with previously reported correlations (Victoir et al., 2005) between instructor and student, although the strength of the correlation was much higher in our study (.8 vs. .4). Given that this particular driving school has a very high success rate, it would be interesting to examine the association between the instructor and the student in relation to student success in future studies.

There was no relation between fluctuations in self-reported workload and skin conductance parameters. Therefore, the present results fit within a theoretical position of emotion research which holds that self-reported measures are associated with and predictive of cognitive schemata, personality dimensions and recall biases, and not with physiological activation (Myrtek et al., 2005; Robinson & Clore, 2002). One could argue that the lack of a correlation is caused by the fact that we averaged the physiology over the whole lesson to compare it to the RSME value. However, only taking the average of the last minutes of the lesson and compare these to the RSME values did nothing to improve the correlation. In Human Factors literature this dissociation between self-reported workload and physiology has also been reported. It has been suggested that these two types of measures are indicative of different mechanisms and processes underlying fluctuations in mental workload (Johannes & Gaillard, 2014; Yeh & Wickens, 1988). Finally, as mentioned in the introduction, EDA as a physiological measure is not assumed to be selective for workload or mental effort (Boucsein, 2012). Therefore, the lack of a correlation for all skin conductance parameters to the RSME value, corroborate in a longitudinal ambulatory context, that EDA is unsuitable to replace specific subjective workload assessments (see also Seitz et al. (2013) for a similar result in experienced truck drivers).

Future research could focus more on the particular strengths of physiological measurements. Unlike self-report, physiological measurements are online, and continuous, providing insight into moment-by-moment fluctuations in arousal. It would be interesting to find out whether EDA can be a valid indicator of “peak” moments (De Waard, 1996), when arousal has become too high for optimal, or at least, acceptable driving and learning behavior. As mentioned above, different driving instruction methodologies should be systematically compared in terms of the physiological arousal that is associated with the various exam moments. It could be that more modular approaches (with several partial exams) are better at removing extreme stress from a final exam moment, and provide a more valid test of the actual driving skills.

This study provides some basis for recommendations to driving instructors. First, the RSME values could be collected quite easily after each lesson, and can be taken as a rough estimate on what kind of learning trajectory the student are facing according to their own experience. In this study the driving instructor was overall very good in assessing the subjective state of her student (a correlation of .8 was found). However, this was not true for every student, and might be caused by specific qualities in this particular instructor. All instructors therefore could consider asking for the RSME value after each lesson to get the self-reported mental effort directly from the student. This would provide the instructor with direct insight into how much effort the student is putting into the lesson. Second, ambulatory EDA measurements are not easily available to the professional market. However, it seems this is about to change (see for example the "Embrace" hardware from Empatica, www.empatica.com/product-embrace). Instructors should be careful in using this hardware and interpreting the results as specifically indicative of mental effort. In general, EDA measurements should not be interpreted in a very specific way, but more as an additional information source (at present not easily available) about the arousal of students. At present, self-reported RSME ratings would give a better idea about the mental workload of student drivers.

Conclusion

The present study showed that closely following the physiology and self-reported workload of student drivers over time provides new and fresh insights. More research is needed on the questions whether physiological measurements would be valuable to use during driver training, and how wearable technology might play a role in this process.

References

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*, 647–658.

Boucsein, W. (2012). *Electrodermal Activity* (2nd ed.). New York, USA: Springer.

Boucsein, W., Fowles, D.C., Grimnes, S., Ben-Shakhar, G., Roth, W.T., Dawson, M.E., & Filion, D.L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology, 49*, 1017–1034.

Collet, C., Clarion, A., Morel, M., Chapon, A., & Petit, C. (2009). Physiological and behavioural changes associated to the management of secondary tasks while driving. *Applied Ergonomics, 40*, 1041–6.

Dawson, M.E., Schell, A.M., & Filion, D.L. (2007). The electrodermal system. In J.T. Cacioppo, L.G. Tassinary, and G.G. Berntson (Eds.), *Handbook of psychophysiology* (3rd ed., pp. 159–181). New York: Cambridge University Press.

De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. PhD Thesis, Rijksuniversiteit Groningen, The Netherlands.

Dijksterhuis, C., Brookhuis, K.A., & De Waard, D. (2011). Effects of steering demand on lane keeping behaviour, self-reports, and physiology. A simulator study. *Accident Analysis and Prevention, 43*, 1074–1081.

Healey, J.A., & Picard, R.W. (2005). Detecting stress during real-world driving tasks using physiological sensors. *IEEE Transactions on Intelligent Transportation Systems, 6*, 156–166.

Johannes, B., & Gaillard, A.W.K. (2014). A methodology to compensate for individual differences in psychophysiological assessment. *Biological Psychology*, *96*, 77–85.

Kappeler-Setz, C., Gravenhorst, F., Schumm, J., Arnrich, B., & Tröster, G. (2013). Towards long term monitoring of electrodermal activity in daily life. *Personal and Ubiquitous Computing*, *17*, 261–271.

Molenaar, P.C.M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*, 201–218.

Myrtek, M., Aschenbrenner, E., & Brügner, G. (2005). Emotions in everyday life: An ambulatory monitoring study with female students. *Biological Psychology*, *68*, 237–255.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2008). Situation Awareness, Mental Workload, and Trust in Automation: Viable, Empirically Supported Cognitive Engineering Constructs. *Journal of Cognitive Engineering and Decision Making*, *2*, 140–160.

Picard, R.W. (2009). Future affective technology for autism and emotion communication. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 3575–3584.

Poh, M.-Z., Swenson, N.C., & Picard, R.W. (2010). A Wearable Sensor for Unobtrusive, Long-Term Assessment of Electrodermal Activity. *IEEE Transactions on Biomedical Engineering, 57*, 1243–1252.

Robinson, M.D., & Clore, G.L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*, 934–960.

Ross, V., Jongen M.M., E., Brijs, T., Ruiter, R., Brijs, K., Komlos, M., & Wets, G. (2012). The Relation between Executive Functioning and Risky Driving in Young Novice Drivers. *Applied Neuropsychology: Adult*, *22*, 61-72.

Seitz M., Daun T.J., Zimmerman, A., & Lienkamp, M. (2013). Measurement of electrodermal activity to evaluate the impact of environmental complexity on driver workload. In *Proceedings of the FISITA 2012 World Automotive Congress, Lecture Notes in Electrical Engineering, Volume 200* (pp. 245–256).

Verwey, W.B. (2000). On-line driver workload estimation. Effects of road situation and age on secondary task measures. *Ergonomics*, *43*, 187–209.

Verwey, W.B., & Veltman, H.A. (1996). Detecting Short Periods of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. *Journal of Experimental Psychology: Applied*, *2*, 270–285.

Victoir, A., Eertmans, A., Van Den Bergh, O., & Van Den Broucke, S. (2005). Learning to drive safely: Social-cognitive responses are predictive of performance rated by novice drivers and their instructors. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*, 59–74.

Vidulich, M.A., Base, W.A.F., & Tsang, P.S. (2010). Mental workload and situation awareness. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (4th ed., pp. 243–273). New Jersey: John Wiley & Sons Inc.

Warm, J.S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors*, *50*, 433–441.

Yeh, Y.-Y., & Wickens, C.D. (1988). Dissociation of Performance and Subjective Measures of Workload. *Human Factors*, *30*, 111–120.

Zijlstra, F. (1993). *Efficiency in Work Behaviour: A Design Approach for Modern Tools*. PhD Thesis, Technische Universiteit Delft, The Netherlands.

Emotions in decision-making processes of Air Traffic Management stakeholders

Maria Freese[1], Alexander Jakobi[2], & Meike Jipp[1]
[1]German Aerospace Center
[2]Fernuniversität Hagen
Germany

Abstract

The aim of this study is to examine the most relevant emotions for efficient decision-making of Air Traffic Management (ATM) stakeholders. This is important, since ATM-stakeholders (airport operators, airlines, air traffic controllers, ground-handlers) have to cooperate in order to guarantee the most efficient air traffic. However, divergent goals and interests can affect the solution that is optimal for the European air traffic system. Moreover, there are factors like workload and situation awareness, which have an influence on decisions. Emotions also affect solutions positively or negatively in decision-making processes, but have hardly been considered in ATM. So as to discover, which emotions are relevant in the stakeholders' working life, an online-questionnaire was developed. This survey consists of a modified version of the Job-Affective Well-Being Scale and was distributed to stakeholders of European airports (N_{airports}=23). The paper will discuss results, in particular that there are two key-emotions (joy and frustration), which are supposed to have an impact on the tested stakeholders' decision-making. Future, potential applications of this research include the development of an assistance system, which reacts adaptively the users' emotional requirements. Adaptive Systems might serve as a co-worker and support the human, especially during decision-making processes in teams to guarantee the most efficient benefit.

Introduction

Emotions are everywhere. They influence our daily life; they are important for our thoughts and actions. Without emotions a modern community would not hold up. However, there still needs to be a lot of research to find out about the emotions` complexity and its connection to decision making. Especially in the context of Air Traffic Management (ATM) it is necessary to focus on emotions during decision making, since they are an important and essential part of decision-making (Kiefer, 2002). Yet they have mostly been disregarded in the context of ATM. Like this the expected increase in the number of passenger movements at an airport could be managed more easily (International Air Transport Association (IATA), 2014). To guarantee the most efficient air traffic, different stakeholder groups (e.g., air traffic controllers, airlines, airport operators and ground handlers) working at an airport

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

have to cooperate. But there are challenges like divergent goals and different interests of each stakeholder group. Therefore the focus has to set at the collaboration of different involved stakeholders groups at an airport. But not only divergent goals, but also emotional influence factors like anger or frustration are supposed to have an influence on the decision-making process of different stakeholders working at an airport. The purpose of the paper at hand is to demonstrate the importance of focussing the decision-making processes and consequently the influence of emotions during these processes.

Theoretical Framework

In the following theoretical aspects of ATM as well as the process of decision-making are presented. Moreover the factor emotion as one of the influence factors on decision-making is discussed.

Concepts of Teamwork in Air Traffic Management

Many different stakeholder groups are at work in an airport. These are:

- Air Traffic Controllers
- Airlines
- Airport Operators
- Ground-Handlers.

To ensure an efficient, safe and ecological correct mobility, new concepts have the aim to optimise the work of the different kinds of stakeholders (Qudrat-Ullah et al., 2007; Shetty, 2008; Harris & Stanton, 2010). To this end, the process of decision-making shall be optimised regarding certain Key Performance Areas (KPAs). These are (EUROCONTROL, 2012):

- environment
- cost efficiency
- safety
- capacity.

The concept of Airport-Collaborative Decision-Making (A-CDM) plans to let different interest groups use the same base of data. Every involved stakeholder can access certain information, which is can be transparently used by everybody (EUROCONTROL, 2012), though the actual process of decision-making is not yet wholly constructed. The development of this concept is called Total Airport Management (TAM), which has been developed by the German Aerospace Centre (DLR) in cooperation with EUROCONTROL (2006). In the future stakeholder-groups' agents should be represented in a control room. It is called Airport Operation Control Center (APOC). In Figure 1 an example for an APOC is depicted. This is the Airport Control Center Simulator (ACCES) at the DLR in Braunschweig.

The aim is to analyse the current situation at the airport regarding certain criteria, so as to work out solutions in cooperative and collaborative processes and to derive plans. This control room can be virtual or real. Decisions should be made on the basis of a collective quantity of information and with the help of an Airport Operation Plan (AOP). For that purpose the AOP has to be iteratively designed for the best. This is promoted by communication and teamwork. According to Günther et al. (2006) this cooperative process is influenced by the finite resources of an airport, but also by the diverging goals of different stakeholders. Therefore the collective process of decision-making is made more difficult (Günther et al., 2006; Meinecke, 2011), furthermore to this day every interest group at an airport works in its own operational centres and because of the areal separation of every interest group that is involved in decision-making processes the collaboration is hindered.

Figure 1. Airport Control Center Simulator (DLR intern, 2015).

During negotiations and discussions in an APOC the decision-making process is relevant. In the following part some main ideas of decision making in general are presented.

Decision-Making Process

In general the process of decision-finding is divided into three approximate phases each of which consists of decision-making processes (Tramm & Rebmann, 1997):

- definition of the problem
- development of possible alternative options
- draw a conclusion, make a decision.

First of all the actual problem will be defined and the concrete problem described. In this phase the target system should also be refined, meaning that the desired output should be described more particularly. Afterwards suitable alternatives for actions are elaborated, which are rated according to certain criteria. In the end a decision will be made, meaning the best alternative is put in focus (Busse, 2012). The factors for measuring the quality of decision-making also focus on the quality of suggested solutions. Firstly, a baseline has to be declared, with the help of which you can evaluate the quality objectively. Moreover the operators' satisfaction has to be mentioned, the time until a decision is made may also be taken into account.

There are many factors, which act upon such a process of a group's interaction. The effect model for the exploration for the interactional process in groups (see Figure 2) illustrates this complex. This model differentiates between personal and situational factors. Personal factors for example refer to the motivation, the knowledge and the human processing capacity. Situational factors are group size and group structure amongst others. Both arrangements in groups act upon both, the interactional process and the actual result. Parameters, which are important in this contribution mostly refer to the sector of personal factors (Wolf, 2013).

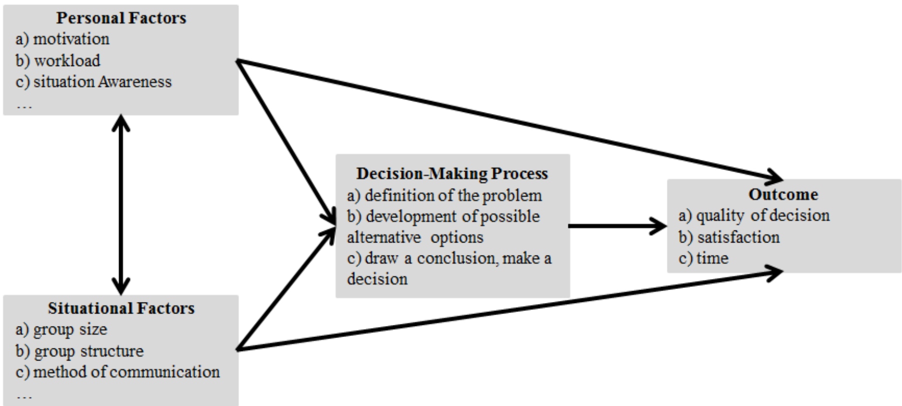


Figure 2. Decision-making process and influence factors (adapted from Wolf, 2013; Freese & Jipp, 2015).

Possible factors, which influence the process of decision-making are:

- motivation (e.g. Jungermann, Pfister & Fischer, 1998; Grafe, 2007),
- previous knowledge from experience (e.g. Gonzalez, Lerch & Lebiere, 2003),
- situational awareness (e.g. Endsley, 1988) and
- stress (e.g. Urban, Weaver, Bowers & Rhodenizer, 1996).

However, as already mentioned above, emotions are a crucial factor that should not be disregarded (also see Vollmeyer & Funke, 1999; Buchner, 1999). It is especially important to regard emotions since the connection between emotions and decision-making has not yet been terminally clarified (Jungermann, Pfister & Fischer, 1998; Spering, 2001; Töpfer, 2004).

Emotions

Even though the terms emotion, feeling, mood and affect are often used synonymously, they are used and defined differently in emotion psychological literature. It is important to define these terms for this work in order to comprehend the questionnaire's design and its results. Nevertheless, a selective definition is difficult since the boundaries of these terms are often flexible.

In general emotions describe a reaction to a specific and nameable incident. For instance someone could feel fear (reaction) on account of an encounter with a spider (incident). Emotions are felt intensively and are short-term in their duration (Kiefer, 2002). A general scientific definition has not (yet) been found. For that reason a working definition is commonly used such as the comparison that Kleinginna and Kleinginna (1981) developed. This comparison divides emotions into a four-level process.

> *"[…] (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as emotionally relevant perceptual effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goaldirected, and adaptive" (Kleinginna & Kleinginna, 1981, p. 355).*

The feeling is an aspect of an emotion which is limited in the subjective perception of an emotion. Above described four-level process is mostly disregarded. So you can say that feelings are directed inside, whereas emotions are outside directed cues to inform other people about those feelings (Angehrn, 2004).

Moods are experienced less intensively than emotions and they are long-term in their duration. No incident is necessary for a certain mood, so they do not need a concrete and nameable activator (Merten, 2003). Nervousness is an example of a negative mood; satisfaction is an example of a positive mood (Kiefer, 2002). In the literature there are different timespans to dissociate moods from emotions more accurately, although the duration of moods can vary according to different authors.

Affects differ from emotions since they have a definite direction. This term is used to describe fierce and uncontrolled emotional processes (Merten, 2003). Affects are reactions to strongly felt emotions. The perception is constricted and the attention is blurred.

For this work emotion was used since it is most specifically defined. Furthermore, researchers of the emotional psychology mostly agree on the concept of different levels of emotional processes, as defined by Kleinginna and Kleinginna (1981). Even if the term is not used as accurately in the everyday use, it is known and understandable to the community. Besides, it is used in many questionnaires. It is important to deal with emotions in the working context since they have a positive as well as a negative effect on job performance. However, research should concentrate not only on the emotions of individuals, since emotions can contribute to the development of a unique group dynamic effecting the job performance of everyone.

Decision-Making and Emotions

Regarding the different areas of research, it can be seen that it was already dealt with the influence of emotions on decision-making processes (Menges et al., 2008). Hereby two fundamental views can be differentiated. On the one hand there are views that have shown via studies that positive emotions (e.g. joy, satisfaction) lead to positive results (Isen, 1993a, 1993b). Furthermore, it was in studies empirically proved, that negative emotions (e.g. frustration, envy) lead to negative effects (Bedeian, 1995). On the other hand there is evidence in the literature, that assume that positive emotions lead to negative and that negative emotions lead to positive results (e.g. Menges et al., 2008). Concluding this means that emotions influence decision-making positively as well as negatively.

In summary, the research field of the ATM was presented with respect to actual decision-making processes of stakeholder groups. There are challenges regarding the optimisation of the decision-making process including the collaboration of different stakeholder groups. Decision-making processes are complex and different factors are supposed to have an influence on performance criteria. In this case not only workload and situation awareness are interesting. Especially, the field of emotions has hardly been considered in the field of ATM. In this paper it will address the research question of whether and which emotions have an influence on decision-making processes.

Method

To ask for the most relevant emotions which are occurring during decision-making processes in ATM, a questionnaire consisting of four parts has been developed. This questionnaire is mostly quantitative but also includes few qualitative questions for further information. The target group are stakeholders (e.g. air traffic control, airlines, airport operator and ground handlers) at an airport. For a bigger sample the questionnaire was written in German and in English and it has been sent to different airports in Europe. The questionnaire was optimised after an internal pre-test and feedback has been gathered at the DLR.

Online Questionnaire - Theoretical base

Since the survey's purpose was not to measure emotions in general but to measure emotions in the context of work, it was used the Job-related Affective Well-being Scale (JAWS). This survey consists of 30 items and was translated into different languages. The scale was developed by van Katwyk, Fox, Spector and Kelloway (2000). Its measured emotions are oriented on the two dimensions pleasure and arousal. These two dimensions are part of James A. Russell's Circumplex Model of Affect, which he defined in 1980. A simplified version of this model is shown in Figure 3.

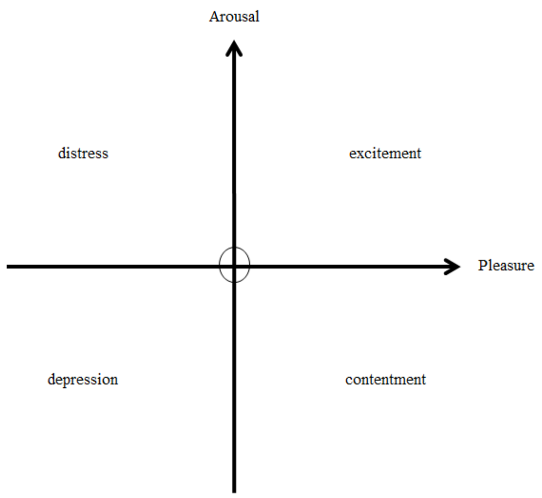


Figure 3. Simplified version of the Circumplex Model of Affect.

Online Questionnaire - content

The questionnaire consists of four parts:

In the questionnaire's first part demographic details were asked. By asking those questions possible differences concerning gender or age could be found out. Besides, the first part consists of questions about one's position and work experience. Moreover, the participants were asked how often they were in situations, in which they had to make decision in a group of different stakeholders. People were further asked what typical challenges in such situations were.

The longer version of the JAWS (30 items) was used for the survey's second part. It was enlarged by 18 items in both used languages. Those additional 18 emotional items were taken from the Circumplex Model of Affect. To avoid primacy and recency-effects the items were shown randomised. A five-stepped answering scale was used – never, rarely, sometimes, often, very often, but it was also possible to give no answer. To ensure an analysis of the items, a pre-test was conducted. In the pre-test (N=36) it was found out that the internal consistency (Cronbach's α) has a good value of 0.82. 14 items (of the 18 new items) were in the acceptance region (0.2-0.8). Just sad (0.15), delighted (0.22), gloomy (0.19) and appalled (0.08) were too difficult. Since delighted and appalled also had a bad selectivity, they were deleted from the real survey. All in all this part consists of 46 Items.

Directly asking the experienced emotions at the workplace was the main part of the survey's third part. The questions differ concerning duration (generally – in the past 30 days) and perspective (your colleagues – yourself). The participants had to assign three emotions of a list of different emotions into a ranking. The shown emotions had not to be used but could be enlarged by up to three new ones.

The short fourth part is mainly qualitative. Here it was asked if the participants had experienced situations in the past, where they were influenced positively or negatively by emotions at their workplace. A further description of this/these situation(s) could be given.

Participants

The questionnaire was sent to 23 European Airports via e-mail. Sixteen ATM-stakeholders filled out the online-questionnaire completely, 26 (six female) experts partly. This number of experts was enough to analyse their answers. Actually in the area of ATM experts a number about ten is good. The range of the age was between 32 and 63 years (mean: 46 yrs., standard deviation [SD]: 9 yrs.). In Table 1 the number of the certain stakeholder group is visible. The work experience in this employment ranged from 3 and 36 years (mean: 18 yrs., standard deviation [SD]: 10 yrs.).

Table 1. Distribution of ATM-stakeholders

	German		*English*	
	N	*in %*	*N*	*in %*
Air Traffic Control	/	/	1	10
Airline	12	75	1	10
Airport	4	25	8	80
∑	**16**	**100**	**10**	**100**

Note. The number represents an agent working as an Air Traffic Controller or Airline Agent or Airport operator, who answered the questionnaire.

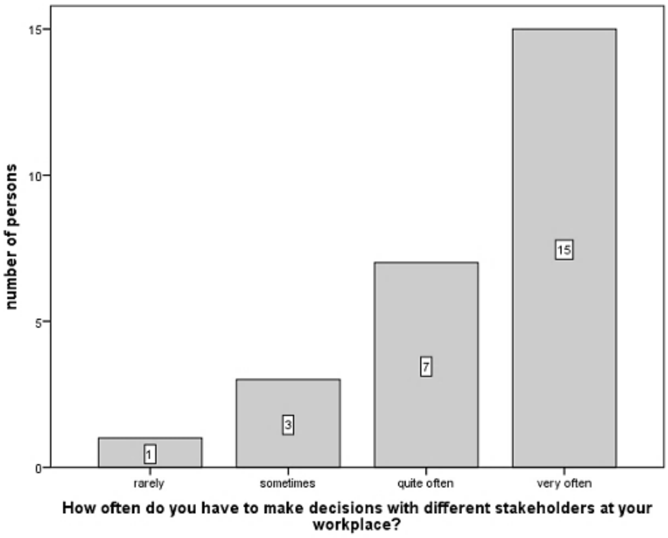


Figure 4. Frequencies of decision-making processes of ATM-stakeholders.

In Figure 4 the frequencies of decision-making processes are depicted. 58% are very often in situations, in which they have to decide in a team with different agents.

Challenges in such situations are especially “different goals and interests”, “time pressure” and “other factors like work overload” (Freese & Jipp, 2015). These results show the necessary of this topic and the further research of team work and possible influence factors in the area of ATM.

Results

As mentioned above, the questionnaire consists of four parts. After presenting the demographical aspects in the following part the frequencies of certain emotions are depicted. It is visible that positive emotions in particular were rated *quite often* and *extremely often* (see Table 2).

Table 2. Frequencies of emotional states (based on the Job-Affective Well-Being Scale))

In the past 30 days my job made me feel …	*never*	*rarely*	*sometimes*	*quite often*	*extremely often*	*no statement*
at ease	-	2	**13**	8	3	-
aggravating	7	**9**	7	2	-	1
aggrieved	7	**12**	6	-	-	1
agitated	-	**10**	**10**	3	2	1
amazed	3	6	**13**	3	-	1
angry	5	**10**	9	2	-	-
annoyed	4	5	**14**	3	-	-
anxious	4	**9**	**9**	3	-	1
aroused	**9**	7	7	1	-	2
bored	**14**	3	6	1	2	-
bright	-	6	5	**12**	1	2
cheerful	-	1	9	**11**	3	2
calm	-	-	9	**14**	3	-
confused	**14**	4	6	1	-	1
content		3	**12**	9	2	-
depressed	**17**	4	3	-	-	2
disgusted	**20**	5	-	-	-	1
discouraged	**10**	5	9	2	-	-
distressed	**12**	8	3	1	-	2
dozy	**9**	5	6	3	-	3
elated	2	4	**10**	5	3	2
energetic	-	4	**10**	9	2	1
excited	1	5	**8**	7	5	-
ecstatic	**14**	7	3	-	-	2
enthusiastic	1	6	**10**	7	1	1
exhausted	1	6	7	**8**	4	-
fatigued	**7**	5	6	5	2	1
frightened	**19**	5	1	-	-	1
frustrated	5	6	**8**	5	2	-

In the past 30 days my job made me feel …	*never*	*rarely*	*sometimes*	*quite often*	*extremely often*	*no statement*
furious	**12**	6	4	3	-	1
gloomy	8	**9**	6	3	-	1
happy	-	4	**11**	8	1	2
inspired	3	6	3	**13**	-	1
intimidated	**19**	5	-	-	-	2
miserable	**17**	5	2	-	-	2
morbid	**18**	5	1	-	-	2
pleased	3	6	**9**	6	-	2
proud	2	2	8	**12**	1	1
quiet	2	1	**12**	6	3	2
relaxed	-	5	**11**	8	1	1
sad	**14**	9	2	-	-	1
satisfied	-	5	5	**13**	2	1
strained	2	9	**11**	2	1	1
tense	1	6	**9**	8	2	-
tired	1	7	**8**	**8**	1	1
upset	6	**12**	6	1	-	1

Note. A number in bold mean that most of the experts rated the certain emotion with this frequency.

In the third part the participants had to assign a number of emotions into a ranking. In Table 3 as well as in Table 4 the number of rankings for the questions about noticed emotions from colleagues and from yourself at the workplace (without the time) can be seen. Because sixteen ATM-stakeholders filled out the online-questionnaire completely, now there are sixteen persons in total. It is noticeable that exhaustion and strain are ranked highly in both tables, which contrasts with the results of the JAWS. Frustration and irritation are other negative emotions, which were often chosen. Positive emotions were ranked less often generally.

Table 3. Rankings of emotional states (noticed from colleagues in general)

Emotion	*Rank1*	*Rank2*	*Rank3*	*Count*	*Missing*	*Total*
strain	1	1	1	3	13	16
angriness	0	0	0	0	16	16
agitation	0	1	0	1	15	16
aggravation	1	0	0	1	15	16
exhaustion	3	3	0	6	10	16
defeat	0	0	0	0	16	16
excitement	0	0	0	0	16	16
pleasure	0	0	0	0	16	16
tension	0	1	1	2	14	16
murk	0	0	0	0	16	16
intimidation	0	0	0	0	16	16
disgust	0	0	0	0	16	16
ecstasy	0	0	0	0	16	16
misery	0	0	0	0	16	16
energy	0	1	0	1	15	16
enthusiasm	1	1	0	2	14	16
discouragement	0	0	2	2	14	16

Emotion	*Rank1*	*Rank2*	*Rank3*	*Count*	*Missing*	*Total*
relaxation	0	0	0	0	16	16
delight	0	0	0	0	16	16
encouragement	0	0	0	0	16	16
arousal	0	0	0	0	16	16
amazement	1	0	0	1	15	16
joy	0	1	0	1	15	16
frustration	1	1	0	2	14	16
cheerfulness	0	0	1	1	15	16
serenity	2	2	0	4	12	16
irritation	1	1	5	7	9	16
satisfaction	0	0	0	0	16	16
brightness	0	0	0	0	16	16
inspiration	1	0	1	2	14	16
distress	0	0	0	0	16	16
boredom	0	0	1	1	15	16
tiredness	1	0	0	1	15	16
despondence	0	1	0	1	15	16
quiet	1	0	0	1	15	16
doziness	0	0	0	0	16	16
fear	0	0	0	0	16	16
anxiety	1	0	1	2	14	16
pride	0	0	0	0	16	16
sadness	0	0	0	0	16	16
fright	0	0	0	0	16	16
confusion	0	1	1	2	14	16
happiness	0	0	0	0	16	16
fury	0	0	0	0	16	16
contentment	0	0	1	1	15	16
upset	0	0	0	0	16	16
anxiety	0	0	0	0	16	16
demotivation	0	0	1	1	15	16
resignation	0	1	0	1	15	16
apathy	1	0	0	1	15	16

Table 4. Rankings of emotional states (noticed from yourself in general)

Emotion	*Rank1*	*Rank2*	*Rank3*	*Count*	*Missing*	*Total*
strain	0	2	0	2	14	16
angriness	0	0	0	0	16	16
agitation	1	0	0	1	15	16
aggravation	0	0	1	1	15	16
exhaustion	5	0	1	6	10	16
defeat	0	0	1	1	15	16
excitement	0	0	0	0	16	16
pleasure	0	0	1	1	15	16
tension	0	0	0	0	16	16
murk	0	0	0	0	16	16
intimidation	0	0	0	0	16	16
disgust	0	0	0	0	16	16
ecstasy	0	0	0	0	16	16

Emotion	*Rank1*	*Rank2*	*Rank3*	*Count*	*Missing*	*Total*
misery	0	0	0	0	16	16
energy	0	1	0	1	15	16
enthusiasm	2	0	1	3	13	16
discouragement	1	0	0	1	15	16
relaxation	0	0	0	0	16	16
delight	0	0	0	0	16	16
encouragement	0	0	0	0	16	16
arousal	0	0	0	0	16	16
amazement	0	1	0	1	15	16
joy	0	2	1	3	13	16
frustration	0	3	2	5	11	16
cheerfulness	1	1	0	2	14	16
serenity	2	0	1	3	13	16
irritation	1	0	2	3	13	16
satisfaction	0	0	0	0	16	16
brightness	0	0	0	0	16	16
inspiration	1	1	0	2	14	16
distress	0	0	0	0	16	16
boredom	1	1	0	2	14	16
tiredness	0	2	1	3	13	16
despondence	0	0	0	0	16	16
quiet	0	1	0	1	15	16
doziness	0	0	0	0	16	16
fear	0	0	0	0	16	16
anxiety	0	0	2	2	14	16
pride	0	0	0	0	16	16
sadness	0	0	0	0	16	16
fright	0	0	0	0	16	16
confusion	0	0	0	0	16	16
happiness	0	0	0	0	16	16
fury	0	0	0	0	16	16
contentment	0	1	1	2	14	16
upset	0	0	0	0	16	16
anxiety	0	0	0	0	16	16
isolation	0	0	1	1	15	16
stress	1	0	0	1	15	16

During the fourth part all participants were asked, whether they have ever reacted unusually due to the influence of positive or negative emotions at their workplace. They had the possibility to give an example of the noticed emotion and to describe the certain situation. In contrast to the answers of the JAWS it was visible, that frustration is one further emotion during these open questions, which is supposed to have an influence on decision-making. All in all, experts stated more negative emotions in that feedback, so that it seemed to be necessary to do further research in the analysis of both, positive and negative emotions.

Discussion & Outlook

The 26 asked experts reported which emotions they had experienced during decision-making situations. Especially positive emotions like joy, cheerfulness and enthusiasm were rated with often and extremely often. In contrast to this, negative emotions like frustration, exhaustion and irritation were rated often, especially during the open answers. These answers show the relevance of emotions during decision-making processes with different interests groups. Despite the fact that emotions have hardly been considered in the context of ATM, the results of the online-questionnaire show the importance of this topic. Especially ATM-experts, who have to deal with decision-making daily, can be influenced positively or negatively by their emotions. To measure the impact of emotions on decision-making, an experiment is already planned.

Even if only a small number of the targeted group was reached, it can be recognised, that especially in the open questions negative emotions have an impact. Therefore it has to be analysed, whether the answers of the JAWS could be based on social desirability effects. A general problem of this work's target group is that the number of experts is limited. The low return rate (N=26) might also be due to the fact that the contacted experts did not find the time to answer the survey or that they did see no necessity to do so. This low return rate means that the found results are not yet statistically verified.

The next step is to empirically confirm that joy and frustration have an influence on decision-making. To test this supposed connection, experiments are planned, in which positive and negative emotions will be induced during decision-making processes. By performing this experiment it will be possible to answer the question, if emotions have an influence on decision-making. If they influence decision-making, it can also be found out what kind of influence it is.

For the future, potential application of this research includes the development of an assistance system, which reacts adaptively on the users' emotional requirements. Adaptive Systems can serve as a co-worker and support the human, especially during decision-making processes in teams to guarantee the most efficient benefit.

References

Angehrn, A.B. (2004). *Emotionen im Team [Emotions in teams]*. Bern: Peter Lang AG.

Bedeian, A.G. (1995). Workplace envy. *Organizational Dynamics, 23*(4), 49-56.

Buchner, A. (1999). Komplexes Problemlösen vor dem Hintergrund der Theorie finiter Automaten [Complex problem solving in the context of the „Theorie finiter Automaten“]. *Psychologische Rundschau, 50*(4), 206-212.

Busse, D. (2012). *Nachhaltigkeitsaspekte in Theorie und Praxis der Entscheidungsfindung [Sustainability issues in theory and practice of decision-making]*. Wiesbaden: Gabler Verlag.

Endsley, M.R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society 32nd Annual Meeting* (pp. 97-101). Santa Monica, CA: Human Factors Society.

EUROCONTROL (2006). *Airport CDM Operational Concept Document.* URL http://www.euro-cdm.org/library/cdm_ocd.pdf [27.03.2015].

EUROCONTROL (2012). *European ATM Master Plan* (Edition 2). Available: http://ec.europa.eu/transport/modes/air/sesar/doc/2012_10_23_atm_master_plan_ed2oct2012.pdf [08.09.2014].

Freese, M. & Jipp, M. (2015). Zwischen Rational und Emotional – Entscheidungen und deren Einflussgrößen in kooperierenden Teams [Between rational and emotional - decisions and their influencing factors in co-operating teams]. In *Proceedings 4. Interdisziplinärer Workshop Kognitive Systeme, 23.-25. März 2015, Bielefeld, Deutschland*.

Gonzalez, C., Lerch, J.F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27*, 591-635.

Grafe, S. (2007). *Förderung von Problemlösefähigkeit beim Lernen mit Computersimulationen [Promoting the ability to solve problems while learning with computer simulations]*. Kempten: Klinkhardt.

Günther, Y., Inard, A., Werther, B., Bonnier, M., Spies, G., Marsden, A., … Niederstrasser, H. (2006). *Total Airport Management.* Available: http://www.bs.dlr.de/tam/Dokuments/TAM-OCD-public.pdf [08.09.2014].

Harris, D., & Stanton, N.A. (2010). Aviation as a system of systems: Preface to the special issue of human factors in aviation. *Ergonomics, 53*, 145-148.

International Air Transport Association (IATA) (2014). New IATA Passenger Forecast Reveals Fast- Growing Markets of the Future. Available: http://www.iata.org/pressroom/pr/Pages/2014- 10-16-01.aspx [13.10.2015].

Isen, A.M. (1993a). The influence of positive affect on cognitive organization: Some implications for consumer decision-making in response to advertising. In A. Mitchell (Ed.), *Advertising exposure, memory, and choice* (pp. 239-258). Hillsdale, NJ: Lawrence Erlbaum Associates.

Isen, A.M. (1993b). Positive Affect and decision making. In M. Lewis & J. Haviland (Eds.), *Handbook of emotion* (pp. 261-277). NewYork: Guilford.

Jungermann, H., Pfister, H.-R., & Fischer, K. (1998). *Die Psychologie der Entscheidung: Eine Einführung [Psychology of decision-making: introduction]*. Heidelberg: Spektrum Akademischer Verlag.

Kiefer, T. (2002). Die Macht positiver und negativer Gefühle in der Arbeitswelt [The influence of positive and negative feelings in the world of work]. *Personalführung, 12*, 49-55.

Kleinginna, P.R., & Kleinginna, A.M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5*, 345-355.

Meinecke, M. (2011). *Bewertung von Total Airport Management Konzepten mit Hilfe eines Spiel-basierten Ansatzes [Evaluation of Total Airport Management concepts using a game-based approach]*. Unveröffentlichte Bachelorarbeit, DLR, Braunschweig.

Menges, J., Ebersbach, L., & Welling, C. (2008). *Erfolgsfaktor Emotionales Kapital [Success Factor Emotional capital]*. Bern: Haupt Verlag.

Merten, B. (2003). *Einführung in die Emotionspsychologie* (1. Auflage) *[introduction in the emotion psychology]*. Kohlhammer. Stuttgart.

Qudrat-Ullah, H., Davidsen, P., & Spector, J.M. (2007). *Complex Decision Making: Theory and Practice.* Dordrecht: Springer.

Russell, J.A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology, 39*, 1161-1178.

Shetty, S. (2008). System of Systems Design for Worldwide Commercial Aircraft Networks. *26th International Congress of the Aeronautical Sciences (ICAS), Anchorage, Alaska (USA), September.*

Spering, M. (2001). *Emotionen und Kontrollüberzeugungen beim komplexen Problemlösen. Eine experimentelle Untersuchung anhand des computersimulierten Problemlöseszenarios FSYS 2.0 [emotions and control beliefs in complex problem solving. An experimental investigation on the basis of the computer-simulated problem-solving scenarios FSYS 2.0]*. Unveröffentlichte Diplomarbeit, Ruprecht-Karls Universität Heidelberg, Deutschland.

Töpfer, S. (2004). *Instanzenbasiertes Lernen: Der Einfluss von Zeitdruck und Persönlichkeitsmerkmalen beim Komplexen Problemlösen mit dem Water-Purification-Plant-Szenario [Instance based learning: The influence of time pressure and personality traits in complexes problem solving with the Water-Purification-plant scenario]*. Unveröffentlichte Diplomarbeit, Ruprecht-Karls-Universität Heidelberg, Deutschland.

Tramm, T., & Rebmann, K. (1997). Handlungsorientiertes Lernen in und an komplexen, dynamischen Modellen - Die Modellierungsperspektive als notwendige Ergänzung des handlungsorientierten Ansatzes in der Wirtschaftsdidaktik [Active learning in and of complex, dynamic models – perspective of modeling as a necessary complement of action-oriented approach in Wirtschaftsdidaktik]. In G. Lübke & B. Riesebieter (Hrsg.), Zur Theorie und Praxis des SIMBA-Einsatzes in der kaufmännischen Aus- und Weiterbildung (S. 1-39). Markhausen: Lübke.

Urban, J.M., Weaver, J.L., Bowers, C.A., & Rhodenizer, L. (1996). Effects of Workload and Structure on Team Processes and Performance: Implications for Complex Team Decision Making. *Human Factors, 38*(2), 300-310.

Van Katwyk, P.T., Fox, S., Spector, P.E., & Kelloway, E.K. (2000). Using the Job-Related Affective Well-Being Scale (JAWS) to investigate affective responses to work stressors. *Journal of Occupational Health Psychology, 5*(2), 219-230.

Vollmeyer, R., & Funke, J. (1999). Personen- und Aufgabenmerkmale beim komplexen Problemlösen [Personal and job characteristics in complex problem solving]. *Psychologische Rundschau*, *50*(4), 213-219.

Wolf, R. (2013). Interaktionsprozesse zur Entscheidungsfindung in virtuellen und face-to-face Gruppen [Interaction processes for decision making in virtual and face-to-face groups]. Available: http://nbn-resolving.de/urn:nbn:de:0111-opus-80746 [08.07.2014].

Work domain analysis of an intensive care unit: An Abstraction Hierarchy based on a bed-side approach

Ida Bodin, Camilla Fröjd, & Anders Jansson
Uppsala University, Sweden

Abstract

Work in intensive care units requires interaction with several medical devices and interpretation of dynamic information from several sources. The aim of the current study was to gain understanding of the work domain to support the development of a holistic information environment and further analyses of risky situations. A total of 18 hours of bed-side observations at an intensive care unit and interviews with three experienced intensive care unit nurses were conducted in order to receive input data for the modelling of the work domain. The domain was modelled in an abstraction hierarchy, as according to the first phase of the cognitive work analysis framework. The results show that the ultimate purpose of the work carried out in an intensive care unit is keeping patients alive while gaining time for treatment, but also to perform treatment and relieve symptoms. The purpose is represented at the top level of the model, and lower levels include functions as supporting the patients' vital functions and avoiding secondary complications. With this work domain analysis as a basis, three different design challenges identified can be dealt with systematically.

Introduction

The Swedish Accident Investigation Authority (2013) recently conducted an investigation on an accident resulting in the death of a patient in an intensive care unit at the Karolinska University Hospital, Sweden. A patient who had an external pacemaker implanted after surgery was after four days transferred from one division to another within the hospital. At the time of arrival, the patient was connected to the telemetric surveillance equipment. When the patient was registered in the system, the staff failed to indicate the presence of an external pacemaker, and there was no barrier in the interaction dialogue in the system checking for such an omission. Since the pacemaker was not marked explicitly, the presence of its pulses was difficult to discriminate from the patient's own heart beats, i.e., the pulses from the pacemaker were interpreted as the patient's own heart beats by the surveillance equipment. This in turn led to situations of cardiac arrest not being appropriately discovered by the surveillance system. During visual inspection of the patient by the night staff, the patient was found dead. The Authority concluded that factors at two different levels of the organization contributed to the evolvement of the accident. Deficits deducible to management level, that is to say latent errors at the blunt-end, resulted in unclear instructions regarding the telemetric surveillance equipment, in

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

turn creating uncertain conditions during transfer between divisions as well as when shifting staff. Despite the lack of clear instructions staff was forced to decide by themselves, since time did not allow consultation with colleagues or management. The other factor pointed to by the Authority was deficits deducible to the interaction between staff and the telemetric surveillance equipment, that is to say event-related errors at the sharp-end. The staffs' use of the telemetric surveillance equipment was not in accordance with the instructions since the alarm system resulted in a number of documented false alarms. This in turn resulted in reduced attention among staff, with the consequence that a correct alarm was rejected. The hypothesizes in the current project are that the following three challenges have to be met to ensure a better work environment where mistakes and errors like the ones identified by the Authority are reduced: (1) The design of single technical apparatus has to be adequate when it comes to usability and interaction; (2) a more the holistic design approach is needed with respect to the work environment where the users perform their work tasks; (3) a proactive safety culture that encourages learning from mishaps and lapses needs to be developed.

In order to investigate how the three challenges can be met, a long-term goal of the intensive care unit project discussed here is to employ a bed-side approach and produce data in order to document as much in detail as possible of actions, events and behaviours allocated to the interaction between the nurses and patients, as well as between nurses and the technical equipment. The assumption is that a bed-side approach for human factors research will support the development of a better understanding of the work conditions in intensive care units. Without a proper understanding of the dynamic decision making processes going on in an intensive care unit, it is probable that the important actions and series of events are overlooked.

The work domain analysis described in the current paper is a part of a first exploratory phase which will define the future direction of the intensive care unit-project. The aim of the work domain analysis is to gain knowledge about the intensive care nurses work domain to be able to address the different challenges mentioned above in a better way.

Even if the design of each and one of the technical apparatus is according to usability standards, the holistic work environment might suffer from usability flaws if the apparatus and softwares not are integrated, ad hoc implementation is therefore not appropriate from a sociotechnical systems design perspective. It is therefore important to adopt a systems perspective when working with improvements of the intensive care unit and from this reasoning a need for understanding the intensive care work is of great importance.

Work domain analysis in intensive care units

The work domain analysis is the first phase of the cognitive work analysis framework (Rasmussen & Pejtersen 1994, Vicente 1999), which is a framework for analysing complex systems. The work domain analysis phase identifies and describes the possibilities and constraints on work conducted within the work system, independent of situations and can therefore be used as a way to deal with novelty according to Vicente (1999). A work domain analysis is usually modelled in an abstraction-decomposition space which divides the system into five levels of system decomposition, represented in horizontal direction, and five levels of abstraction in vertical direction (Vicente, 1999). The five abstraction levels are from the overall purpose of the system at the top level, to the physical components of the system at the bottom, and the functions of the physical components of different abstractions in the levels in between.

Methods from the cognitive work analysis framework have successfully been used in many application domains and are often used to support interface design, according to a review by Read, Salmon, and Lenné (2012). Health care and even intensive care unit’s is an area where the cognitive work analysis framework has been used. For example Effken et al. (2011) conducted a cognitive work analysis by interviewing health care managers, to better understand environmental constraints and the implications for a decision support tool. Miller (2004) describes a work domain analysis modelling the intensive care unit patient, which further is used as a basis for information design for different media, by Miller et al. (2009). Gorges et al. (2013) modelled a work domain analysis for paediatric intensive care unit, which was used for rapid prototyping of a mobile patient monitoring application.

The current paper reports a work domain analysis with a bed-side perspective and a focus on the work conducted by intensive care nurses at an intensive care unit, because a good understanding of the work domain are useful in the development of a representative information environment. Therefore the focus is on understanding and modelling the sociotechnical system and work domain of the intensive care unit nurses, as opposed to the patient (as by Miller, 2004) or with a wider focus on organizational factors (as by Gorges, 2013).

Method

The work domain analysis of an intensive care unit, from a bed-side perspective, has been modelled based on information achieved through an observation study (figure 1), and through interviews and participatory analysis made by three subject matter experts. The subject matter experts’ have 12 to 16 years of experience working as specialized nurses at the intensive care unit. One of them (Camilla Fröjd) is co-author of this paper.

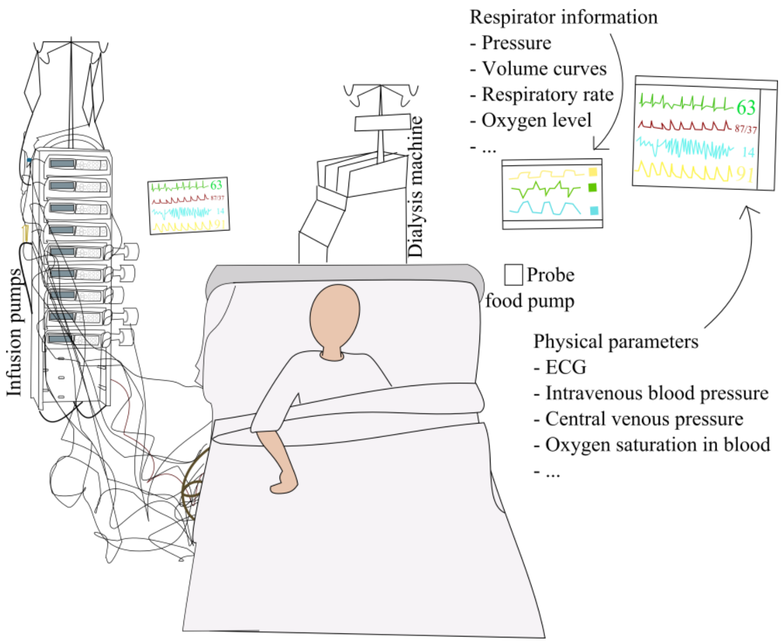


Figure 1. Illustration of the workplace in the intensive care unit.

The modelling of the work domain was a continuous process moving from a focus on collecting data towards reviewing the modelling of the work domain, see figure 2. The work domain has been modelled iteratively during and between the data collection sessions and the expert reviews in an abstraction hierarchy. The abstraction hierarchy includes five levels and system decomposition has also been included where it provides clarification. The names of the abstraction levels are developed by Rasmussen as reported by Reising (2000), and also described in Jenkins et al (2009) as the following: The top level contains the purpose of why the system exist, and is called the functional purpose. The second highest level is called values and priority measures, and can be seen as how the functional purposes can be measured. The middle level, the purpose-related functions, is the functions necessary to achieve the functional purposes described on an abstract level. The second lowest level is called object-related processes and is the functions of the physical components of the system, or why they exist. The lowest level of the abstraction hierarchy is called physical objects, and contains the physical components of the work system.

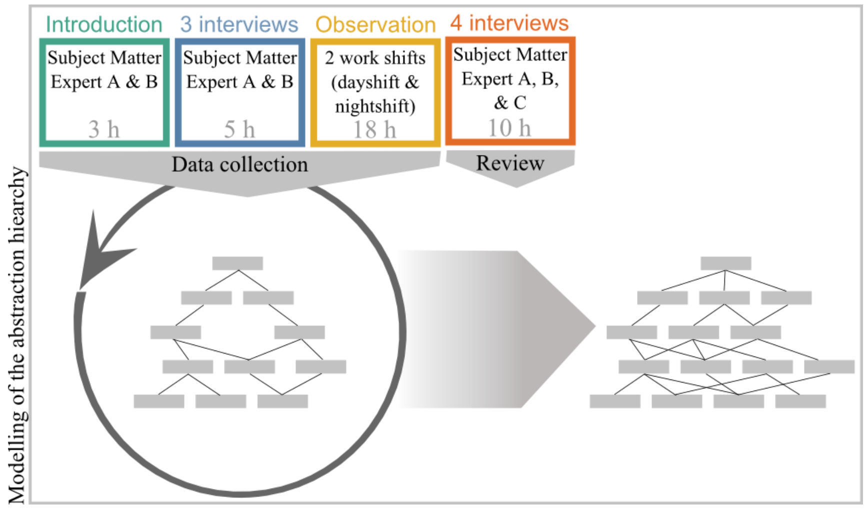


Figure 2. Illustartation of the process of data collection and modeling of the abstraction hierarchy describing the work domain in an intensive care unit.

The data collection started with a three hour introductory visit at the intensive care unit, guided by two of the subject matter experts. Here the basic layout of the intensive care unit, equipment, work routines, and tasks were described.

The data collection continued by interviewing the two subject matter experts separately during totally five hours divided into three sessions. The interviews were semi structured, and the interview questions were based on the questions suggested by Naikar, Hopcroft, and Moylan (2005). The abstraction hierarchy was modelled during and between the interviews, and was used as a base for the later interviews. In that way the subject matter experts could be asked to provide more information about *how* to fulfil the different functions or purposes described in the abstraction hierarchy, or *why* the physical components exist or what they can be used for.

More data were collected during an observation study that took place at the central intensive care unit at the University Hospital in Sweden. The observations took in total 18 hours distributed over two work shifts, first a day shift and then a night shift. Two to three different work teams were observed during each work shift, and questions about why tasks were performed were asked when the work situation allowed it. The notes taken during the observations were later translated into nodes and connection between nodes in the abstraction hierarchy.

In the review session, three subject matter expert were interviewed. They separately reviewed the nodes in the abstraction hierarchy and their connections, and the abstraction hierarchy was updated according to their feedback. All review sessions took in total ten hours divided into four work shifts.

Results

Functional purpose

The functional purpose of the work domain is in the abstraction hierarchy described as; “Treat, relieve symptoms, and gain time for treatment and the effect of treatment” (see figure 3). The functional purpose of the intensive care unit, the overall purpose of why the system exists, from the perspective of the worker in the system is to keep the patient alive while the patient receives the lifesaving treatment. Other units at the hospital have the specialization of treating different injuries or diseases, and therefore a physician from another unit is assigned the patient depending on the patients’ needs. However, the border for the actual treatment of the patients´ injury or disease and the intensive care is not a sharp edge, and some of the treatment is also carried out by the intensive care unit. What is characteristic for the intensive care unit is that patients need more support of the vital functions to stay alive than can be given at other hospital units. An additional purpose of the intensive care unit is to relieve the patient’s symptoms.

Values and priority measures

The three values and priority measures and their connections are visualized in Figure 3. The value and priority measure “Person centred care” is perhaps not an obvious part of gaining time for treatment or treat a patient’s disease or condition. It is anyways modelled as connected to the functional purpose (instead of described as a functional purpose itself) because the patient’s privacy and wellbeing is an important part of how the intensive care is conducted and defines how well it is performed.

The two other values and priority measures are directly connected to keeping the patient alive and enable treatment of the patient’s disease or condition, or to not cause other damage because of the intensive care itself. These values and priority measures are “Prevent secondary complications” and “Support the patient's basic & vital functions”. Both values and priority measures are further decomposed into categories of secondary complications or the patient’s body functions that might need support, respectively.

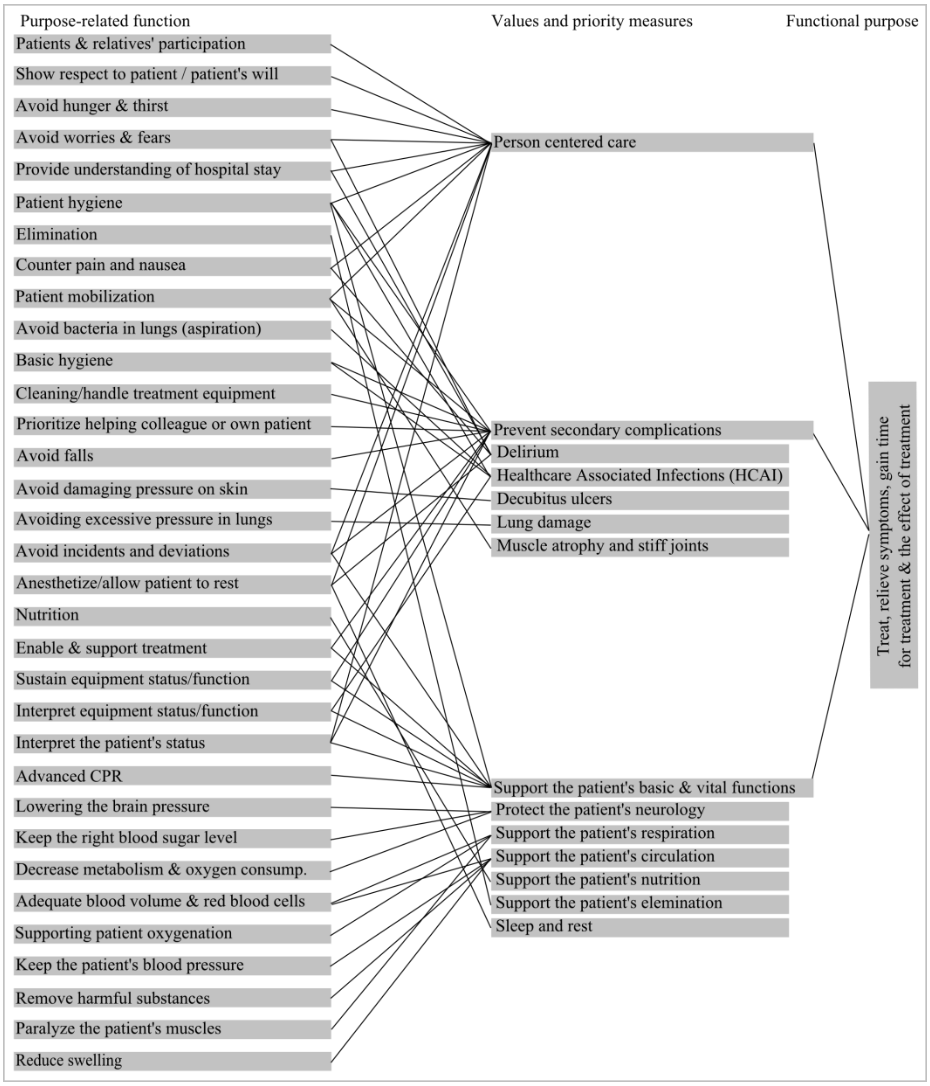


Figure 3. A simplified version of the top three levels of the abstraction hierarchy modelled for the work domain of an intensive care unit.

Purpose-Related Functions, Object-Related Processes and Physical Objects

The purpose-related functions, and their connections to the values and priority measures, can also be seen in Figure 3. Those are the functions of the work system, described on an abstract level, which in this system mean that they are functions described from a care or medical perspective. The object-related processes' are instead described from a technical perspective, which is more related to the medical technology or physical objects of the work system. The object-related processes' and the physical objects are not included in Figure 3 due to space limitations.

Examples of functions in the intensive care work domain are to interpret the patient's clinical status, interpret equipment status, and perform intensive care treatment. Those functions require interaction with several medical devices, and interpretation of information from several sources. Examples from the abstraction hierarchy in full is that the purpose-related functions “Interpretation of patient’s status” are supported by the object-related processes “Monitoring of vital signs (physiological data) & information about patient status”, which is among other physical objects implemented through blood gas measurements and other analysis results, x-ray, and patient data presented on a monitoring display. To interpret the patient’s status also examination of the patient and the information provided from the nurse that worked the shift before is very valuable. Interpretation of equipment status regards for example that the ventilator settings fit the patient, and that all infusion pumps distribute the drugs as appropriate. The purpose-related functions related to the implementation of IC prescription and treatment, as for example “Keept the patient´s blood pressure” and “Remove harmful substances”, are implemented by many technical functions where one of them regards administration of drugs to the patient (object-related process) implemented with for example infusion pumps (physical object).

Discussion

The future work concerns meeting the three challenges mentioned in the introductory section: (1) Design of single technical apparatus has to be adequate when it comes to usability and interaction; (2) a more holistic design of the work environment that supports the users in performing their work tasks; (3) a proactive safety culture that encourage learning from mishaps and lapses.

Constraints imposed on the work in the sociotechnical system are identified and modelled in the abstraction hierarchy, from the physical objects, as medical technology, to the higher purpose of gaining time to treat the patient’s injuries by for example supporting the patients’ vital functions. Higher level goals for intensive care unit could instead be expressed with a focus on the whole organization, but for the intensive care nurses’ work domain the higher level goals on a patient level are central, and therefore is the bed-side perspective important for the understanding of the nurses work domain and to be able to support an holistic design, which is mentioned earlier as challenge one and two. The model of the work domain contributes to an understanding of the work conducted within the work system, as well as information requirements and functions needed, which can be used to support development of an interface design integrating the different parts of the sociotechnical system. Therefore an understanding of the work domain is useful to support interface design (challenge 1) and especially the focus on the holistic work design (challenge 2) because the visualization of the relations between important domain functions. But further analyses are needed to be able to support the development of the interface design. Analysis that investigate the control tasks need to be conducted and possible strategies to complete the work tasks, which are a continuation in accordance with the cognitive work analysis framework.

The focus on the intensive care nurses work domain was chosen because of the need to gain a larger understanding of the nurses work domain, and bring the knowledge through to design and in the continuation of the project. Further the work domain analysis can be extended to also include data from interviews with physicians working within the intensive care unit, which would be a useful complement because the functions in the work domain connected to physicians are different from the functions connected to the nurses. Therefore, the physicians input would be from another perspective, and could give other requirements for the system design.

Regarding the third challenge about supporting development of an proactive safety culture, the domain analysis did not give direct input. In the future work, that will also include safety culture aspects, the understanding of the work achived through the domain analysis will be used when discussing risks for incidents with nurses using the Collegial Verbalization method (Erlandsson & Jansson, 2013; Jansson et al., 2013; Jansson, Erlandsson & Axelsson, 2015; Jansson, Olsson & Erlandsson, 2006) and when analyzing those from a system perspective. The future work is also to include a workshop study with the nurses at the intensive care unit, to increase the learning from incidents and abnormalities.

References

Effken, J.A., Brewer, B.B., Logue, M.D., Gephart, S.M. & Verran, J.A. (2011). Using cognitive work analysis to fit decision support tools to nurse managers' work flow. *International Journal of Medical Informatics*, *80*, 698-707.

Erlandsson, M. & Jansson, A. (2013). Verbal reports and domain specific knowledge: a comparison between collegial and retrospective verbalisation. *Cognition, Technoly & Work, 15, 239-254.*

Gorges, M., Morita, P.P., Burns, C.M., & Ansermino, J.M. (2013). Mobile patient monitoring for the pediatric intensive care unit work domain analysis and rapid prototyping results. In *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on* (pp. 3765-3770). Piscataway, USA: IEEE.

Jansson, A., Olsson, E., & Erlandsson, M. (2006). Bridging the Gap between Analysis and Design: Improving Existing Driver Interfaces with Tools from the Framework of Cognitive Work Analysis. *Cognition, Technology & Work, 8, 41-49.*

Jansson, A., Erlandsson, M. & Axelsson, A. (2015). Collegial verbalisation – the value of an independent observer: an ecological approach. *Theoretical Issues in Ergonomics Science, 16, 474-494.*

Jansson, A., Erlandsson, M., Fröjd, C., & Arvidsson, M. (2013). Collegial collaboration for safety: Assessing situation awareness by exploring cognitive strategies. In *Proceedings from the 14th Interact Conference, Workshop on Human-Work Interaction* (pp. 35-40). Berlin, Germany: Springer

Jenkins, D., Stanton, N., Salmon, P. & Walker, G. (2009). *Cognitive work analysis: coping with complexity*: Ashgate Publishing Company.

Miller, A. (2004). A work domain analysis framework for modelling intensive care unit patients. *Cognition, Technology & Work, 6*, 207-222.

Miller, A., Scheinkestel, C., & Steele, C. (2009). The effects of clinical information presentation on physicians' and nurses' decision-making in ICUs. *Applied Ergonomics, 40*, 753-761.

Naikar, N., Hopcroft, R., & Moylan, A. (2005). Work domain analysis: Theoretical concepts and methodology: DTIC Document.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive systems engineering*. Wiley.

Read, G.J.M., Salmon, P.M., & Lenné, M.G. (2012). From work analysis to work design: A review of cognitive work analysis design applications. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 368-372) Thousand Oaks, USA: SAGE Publications

Reising, D.V.C. (2000). The abstraction hierarchy and its extension beyond process control. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 194-197) Thousand Oaks, USA: SAGE Publications

Swedish Accident Investigation Authority. (2013). Dödsfall på hjärtklinikens hjärtintensivavdelning N25 vid Karolinska Universitetssjukhuset i Solna, Stockholms län, den 17 oktober 2010 *(RO 2013:02).* Retrieved from http://www.havkom.se/virtupload/reports/RO%202013_02.pdf.

Vicente, K.J. (1999). *Cognitive work analysis: toward safe, productive, and healthy computer-based work.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Characterising influences on safety culture in military aviation: a methodologically grounded approach

Anthea Bennett[1,3], Elizabeth Hellier[2], & Andrew Weyman[3]
[1]Institute of Naval Medicine, [2]University of Plymouth,
[3]University of Bath
UK

Abstract

Historically, much effort has been expended in safety culture / climate research toward identifying a generic core set of components, predominately using the self-administered questionnaire approach. However, no stable unified model has emerged, and much of this research has taken a methodologically top-down approach to depicting organisational safety culture. In light of this, the benefits of qualitative exploration as a precursor to and foundation for the development of quantitative climate measures are increasingly recognised. When grounded in the viewpoint of employees, qualitative data driven techniques can provide an insight into how those within an organisation make sense of their work environment and how this impacts their understanding of safety. The current research aimed to address issues of ecological validity by using a qualitative approach to exploring and characterising military aviation employee perspectives on safety culture and risk taking prior to development of a quantitative measurement tool. A thematic analysis of twelve focus groups (N=89), conducted with military employees in a semi-structured manner, was undertaken. This insight into how these personnel interpret their working world was characterised by six nameable constructs: 1. Policy and procedures, 2. Pressure, 3. Management ownership of safety, 4. Individual responsibility and risk perception, 5. Communication and 6. Organisational commitment. Interpretation of these constructs and implications for the future development of a quantitative measurement tool are discussed.

Introduction

Despite intense research interest over the past three decades, the theoretical basis for the concept of safety culture remains indistinct (Kim & Wang, 2009). Efforts toward development of a generic set of components to characterise safety culture have proved generally unsuccessful (Guldenmund, 2007); findings and exploitation of safety culture research have tended to be inconsistent (Cox & Flin, 1998; Pidgeon, 1998) and increasingly fragmented. This lack of consistency in research findings supports the supposition that understanding of safety has a high contextual specificity, where generic measure approaches will have little utility (Jeffcott et al.,

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

2006). The lack of a universally accepted definition of safety culture and its boundaries are likely to exacerbate these inconsistencies.

Safety culture-common constructs

Safety culture research is dominated by psychometrically derived tools designed to measure a multi-dimensional construct. However, studies vary considerably in the number and content of these definitions (Guldenmund, 2007). In an effort to coalesce this research Flin et al. (2000) consider common headline variables through interpretation of (often idiosyncratically labelled) constructs. Common variables identified include management commitment, workforce involvement, training / communication, employee risk perception, nature of the work environment and policy / procedures. Many safety climate questionnaire tools have taken a methodologically top down approach to development, often based on already existing measures or constructs (Cox and Flin, 1998), which often date back to seminal work in this area (Zohar, 1980). Top-down measurement development has been criticised for running the risk of imposing theoretical concepts and underplaying the potential for social difference (Seo et al., 2004), with negative implications for instrument validity which has proved elusive for many instruments across sectors or professions (Guldenmund, 2000).

The benefits of data driven approaches, grounded in employee perspectives, are claimed to have the potential to cast light on how factors impact culture, and in what ways identified variables influence employee decision making and behaviour in relation to risk (Cox & Flin, 1998; Weyman et al., 2006). There are, as yet, only a small number of published studies using qualitative methods (Jeffcott et al., 2006; Blazsin & Guldenmund, 2015; Nordlof et al., 2015), however the rich detail in these provides contextual insight into understanding of safety culture within their respective organisations.

Safety culture in aviation

Aviation has received a great deal of interest in this area (McDonald et al., 2000; Falconer, 2006; Petterson & Aase, 2008; Goodheart & Smith, 2014), however even within this industry there is a lack of convergence on common constructs (O'Conner et al., 2011). Understanding military attitudes to safety and risk are important given the requirement for these personnel to have a positive attitude toward safety yet willing to take risks (Borjessin et al., 2011). This is dissimilar to civilian life, in which much of the safety culture research is grounded, although many of the regulatory and legislative requirements may be common to both. The interplay of military requirements and regulatory policy may result in different interpretations of safety than have previously been identified in civilian studies (Turner & Tennant, 2009). The Defence Aviation Environment (DAE) has undergone considerable changes in recent years as a result of recommendations made in the Nimrod Review (Haddon Cave, 2009), a detailed investigation into the loss of the Nimrod MR2 XV230 aircraft in 2006. Implementation of these recommendations has resulted in changes to risk ownership structures, processes associated with safety management, regulatory authorities and error management (Ministry of Defence, 2012) while safety culture has become a key focus as a result.

Materials and methods

The aim of the current research was to use qualitative methods to enhance insight into defining influences on safety culture and risk taking within a military aviation context and to use this as the basis for the development of a more ecologically grounded and valid quantitative safety climate measurement tool. The study reported here relates to findings from initial foundation qualitative research.

Study design

An exploratory, qualitative approach was adopted for the current study. As the focus was on shared understanding of factors related to safety, focus groups were selected as the method of enquiry. This method has several advantages; it provides insight into shared sense making, norms, values and attitudes, facilitates group dynamics to allow synergistic amalgamation of ideas (Kitzinger & Barbour, 1999) and allows the complexity of employee experience to be captured (Jeffcott et al., 2006). A semi structured interview schedule was developed through discussion with subject matter experts (SMEs) within the organisation, as well as the findings from previous studies and the wider safety culture / climate literature.

Participants and recruitment strategy

The study was undertaken with ethical approval from both the Ministry of Defence Research Ethics Committee (MODREC) and the University of Bath. Participants were an opportunity sample of military personnel attending safety training courses. Twelve focus group sessions were conducted, lasting between forty five and sixty minutes. The sample (N=89) was predominately aircrew (such as pilots) and aircraft maintainers (known broadly as engineers), with a small number of Air Traffic Control (ATC) personnel. Details of the samples can be seen in Table 1.

Table 1. Trade and number of participants in focus group sessions

Focus group	Trade	Number of participants
1	Engineer	9
2	Aircrew	8
3	Engineer	6
4	Aircrew	9
5	Engineer/Air Traffic Control	5
6	Aircrew	9
7	Aircrew	7
8	Engineer	6
9	Engineer	6
10	Engineer	5
11	Engineer	11
12	Engineer	8

As personnel in these three departments were unlikely to interact during their normal work, it was decided to conduct the departmental groups separately. As there were only a small number of ATC personnel, they were accommodated within a focus group with Engineering personnel. In accordance with the ethical protocol, participants were volunteers, fully briefed prior to consent, able to withdraw at any time and assured confidentiality. The main author facilitated the focus groups, and an assistant took notes.

Data analysis

The sessions were audio recorded and transcribed verbatim. Thematic analysis was selected to explore the data as it is well suited to identifying, analysing and reporting patterns in the data (Braun and Clarke, 2006) while accommodating interpretation of the research topic (Boyatzis, 1998). The analysis was both empirically and theoretically driven, with all themes being grounded in data provided by participants (Biggs et al., 2013).

The analysis process was guided by methods described by Braun and Clarke (2006). Steps were reviewed iteratively and the method of constant comparative analysis (Glaser and Strauss, 1967) was used to encourage conceptualisation of the relationships between the data (Thorne, 2000). The transcripts were imported into the Nvivo 9 software where initial analysis and coding of the data on a small number of transcripts (N=2) were undertaken independently by the first author and an assistant, after which differences in classification and interpretation were discussed to challenge assumptions. Iterative cycles of grouping initial codes into clusters and subsequently themes resulted in a thematic framework. On completion of this thematic framework, an inter-rater reliability assessment was undertaken on sample transcripts (N=3). Cohen's kappa statistic was used to assess the degree of concordance which resulted in a coefficient of 0.62. The definition and boundaries of the themes was revisited, with a subsequent assessment returning a coefficient of 0.72.

Results & Discussion

Thematic analysis of the focus group data resulted in identification of six nameable themes that are considered to characterise how these military aviation employees articulate headline influences on workplace safety culture and risk taking (Table 2). Each theme embodied a number of related facets which will be discussed in more detail. Illustrative quotes from the focus group sessions are identified in *"italics*", with clarification notes in (parentheses) and the author's occupational group in [square brackets]. Given the difference in nature of both the work environment and the work tasks carried out by Aircrew and Engineers, it was considered important to note any differences between these groups during the thematic analysis. Within this section, where themes were applied, or interpreted differently between the two groups, this is noted. The small number of ATC participants meant that a similar comparison could not be made with this group when considering the themes identified.

Table 2. Themes and sub themes

Theme	Sub-theme
Policy & procedures	Legitimacy Barriers
Pressure	Organisational pressure Interdependence of functions
Management ownership of safety	Supervisory/line management Senior management
Individual responsibility & risk perception	Camaraderie Perceived consequences
Communication	Reporting Just culture
Organisational commitment	Pride

Policy and procedures

Respondents portrayed close adherence to policy and procedures as an accepted and strongly socially legitimised facet of the workplace context, playing a key role in safety perception and implementation throughout the organisation. This theme was apparent in all twelve focus groups and is considered to be characterised by two sub-themes, legitimacy and barriers. Respondents appeared to see policy and procedures as necessary and important, with adherence and compliance being imperative, this was interpreted as these having legitimacy. Sentiments expressed indicated that their legitimacy was enhanced by the belief that *"many of these rules….have been brought about because of accidents happening in the past, i.e. lessons learned"* [engineer]. This acceptance of the procedures was interpreted as positive, given the highly procedural nature of much of the work *"nearly everything that we do in aviation is very procedural"* [aircrew] which requires rule compliance to support the safety strategies (Hopkins, 2010). This is plausibly reinforced by the high consequences of failure and the cognitive availably of repercussions.

Despite the high legitimacy afforded, respondents juxtaposed this with descriptions of barriers to compliance. The most common of these was the claimed difficulty in keeping up to date with policy and procedures *"I don't think everybody necessarily knows exactly what procedures or what policies or regulations are necessarily applicable.."* [engineer]. This was attributed to the high volume and perceived constant alteration, with information spread across several sources *"Each week more policy comes out and everyone's expected to read it and take it in"* [engineer]. As a result, respondents portrayed themselves as vulnerable, in the sense of blame and accountability arising from inadvertent non-compliance. A lack of knowledge may lead to unintentional violations that are unrelated to inappropriate action or lack of motivation (Laurence, 2005), yet this has received little attention in the literature (Dahl, 2013).

Difficulties in compliance were also attributed to procedures that were lagging behind technical advances, changes that *"hadn't been acted on"* [engineer], and contradictory procedures which necessitate *"applying common sense"* [engineer] were identified as further barriers. Involvement of frontline personnel in rule creation/alteration encourages compliance (Blazsin & Guldenmund, 2015). Changes should be acted on promptly as where inadequate, incompatible procedures are not changed in a flexible manner, locally accepted ways of working may result. This was evident in this organisation *"A lot of it has become the norm so you don't think about it…"* [engineer] which is a concern as this embodies an "*unknown risk"* [engineer], leading to situations where people *"may not really consider the implications"* [engineer] of what they are doing. Situations such as these may increase the potential for error and unintended circumstances (Reason, 1997) yet little is understood about how these 'norms' develop or can be managed.

Pressure

A general tension or trade-off between productivity and safety objectives has been identified in diverse industries (Weyman et al., 2003; O'Dea et al., 2010) and this has been a common thread in safety climate assessments (Flin et al., 2000). However, the definitions and sources of pressure vary between industries, and how these are communicated and interpreted may be context specific. In the current study non-compliance was generally seen as meeting with strong cultural and institutional disapproval *"I don't think many people, if any, would knowingly do that* (not adhere to standard operating procedures) *in this day and age"* [Aircrew], yet still two main sources of pressure were cast as being at odds with this. The first of these was allied to achieving organisational demands (called capability which can include defence, search and rescue, training and humanitarian efforts). The high perceived importance of capability was considered to be a shared group experience between respondents given the perceived common awareness that *"we are in the military-we have to be operationally effective, we have to have that capability"* [aircrew]. This pressure was considered to be implicit as personnel cast their experience as being highly conscious of their military purpose *"the reason we're here is defence of the country"*. Military work can be considered an outlier in comparison to civilian organisations, and the acceptance of risk may differ in training and routine tasks to combat activities (Turner & Tennant, 2009). This is arguably likely to affect views on safety that may not be reflected in the, predominately civilian, published research literature.

During routine tasks, safety was presented as often creating additional time pressure; *"a lot of this policy adds time on to jobs"* [engineer] both at team and individual levels, which was reinforced through management. For engineers particularly, failure to achieve these goals was described as risking reputational damage for individuals *"people believe they have to do a job in a certain time otherwise they will be looked at in a negative light"* [engineer] but also at a team and organisational level. This was evidenced in descriptions of the military 'can-do' attitude which has been observed elsewhere (Bosak et al., 2013). This attitude was presented as pervasive despite management directive to the contrary *"now the* (senior management) *has turned around and said I don't want you to have a can-do attitude"*. This may represent a difference between senior management command and local management implementation, or could highlight self-fulfilling attitudes, both of which would benefit from management consistency and open communication. This type of pressure may lead to reduced attention to rules that are seen to impede progress, potentially circumventing a safety barrier (Bosak et al., 2013). Furthermore, in cases where a negative outcome does not occur, this may be incorrectly perceived as 'safe' and adopted again in future.

Management ownership of safety

A key construct in many safety culture assessments, good management involvement in safety has been shown to have a positive relationship with safety behaviours (Cheyne et al., 1998; Flin et al., 2000). The current study inductively identified two levels of management within respondent accounts. These two levels were the supervisory / line management level, and the senior management level, characterised by contrasting expectations and respondent experiences. Supervisors were seen as most influential in day to day business, being role models for appropriate behaviour *"The different techniques that people have to supervise and the way you've been taught and the procedures that they follow are passed on* (to subordinates)*"* [engineer] and passing on normative behaviour. These individuals are important in determining potential safety performance in personnel, making individuals *"more comfortable* (about) *being able to speak up"* [engineer]. Positive management behaviours have been seen to be important in determining compliance (Mearns et al., 1997; Simard & Marchand, 1997).

Contrastingly, senior management ownership was only discussed through impressions of primacy given to human factors and safety training and safety promotional material – this was cast positively as senior management support. Findings from other sectors suggest that senior management are under pressure to show that they are dedicated to safety (Biggs et al., 2013) and 'walk the walk', not just 'talk the talk'.

Individual responsibility and risk perception

Analysis of respondent accounts of their experiences showed a high, often intrinsically motivated, personal accountability for safety of oneself and one's colleagues *"we're our brother's keeper and look out for each other"* [aircrew] as well as the general public. Of note were references to small team size and general familiarity with colleagues as drivers of this safety awareness. This was interpreted as feelings of camaraderie – a sense of looking out for each other. Although this intrinsic motivation presented as strong, extrinsic motivation of legal responsibility for actions was also cast as important; personnel perceived personal liability as more pertinent due to changes borne out from the Nimrod report (Haddon Cave, 2009).

Respondent experiences suggested an inherent acceptance that aviation is a high risk activity and has the potential for severe consequences. Similar themes were observed in aviation personnel by Patankar (2003), suggesting an overarching industry wide risk perception may exist. Interestingly, while respondents accepted their involvement in high risk activities, this was cast as being managed through *"due risk assessment"* [engineer]; military activities are weighed against potential gains to operational capability (Turner & Tennant, 2009). This concept of organisationally determined acceptable risks has not yet been discussed in the literature-further investigation may provide insights into understanding of safety culture.

Communication

Communication about safety issues was presented by respondents as required to encourage organisational learning, such as using reported mistakes as *"really good learning tools"* [engineer]. However, despite acknowledged management encouragement, some respondents claimed a *"stigma of being labelled"* if they were to self-report mistakes. The embarrassment of admitting a mistake may affect behaviour, even in employees who acknowledge the importance of reporting (Nordlof et al., 2015). In Defence Aviation, the concept of a 'just culture' (treating individuals in a fair and consistent manner and applying appropriate sanctions) is promoted in order to encourage reporting - this concept has been widely encouraged and promulgated in communications in the study organisation.

It was this issue of 'just culture' in which it was identified that perceptions may diverge along functional boundaries. Differences between experiences of Aircrew and Engineers were noted. Aircrew generally agreed that *"we know we must admit a mistake, that we must learn from it and that we will only get punished if we have absolutely contravened regulation"* [aircrew] while many aircrew suggested that this may be different for the engineering function. This was partially supported, with a suggestion that the just culture was *"more alive and well at the top of the chain, but less so at the bottom*" [engineer], but this discussion did not extend to the junior engineers. Given the institutional requirement to report safety related incidents, it is unlikely that people would admit to under-reporting, yet there are indications that perceptions of the just culture concept may not be homogenous. This concept of sub cultures is increasingly being acknowledged in safety culture research (Antonsen, 2009), with cultural boundaries being identified between functional departments or specialisations (Gherardi et al., 1998; Mearns et al., 1998; Richter & Koch, 2004).

Organisational commitment

The final identified theme was named organisational commitment and was defined as the shared sense of safety purpose between employees, a sense of belonging to a group. This manifested through discussions about pride in the safety reputation of the organisation *"I think we all see it as our heritage…and we would never try and put a negative view on it"* [aircrew] where safety was cast as an integral part of the organisation which was instilled early on in personnel training *"it's been so engrained from the time that you join that it's something that becomes second nature to you"* [engineer]. This was summarised by one respondent suggesting that *"Safety reputation…is all important- we think we're reasonably safe, but not complacent"*. Similar findings have not been explored in many safety culture measurement tools, however a similar construct named 'Pride in company' was found by Patankar (2003) using statements such as 'I am proud to work for this company' and 'My company is the best in the business'.

Conclusion

The grounded approach to characterising safety culture in a military aviation population has provided insights that support the requirement for contextualisation of safety culture. These insights will be used to support the development of an ecologically valid quantitative safety climate measurement tool. In contrast to other sectors, policy and procedures were shown to have high legitimacy, despite some organisational barriers to compliance being identified. Exploration of employee interpretation of the priority of organisational goals and potential sub-cultural differences will be facilitated through a more quantitative tool to determine the generalizability of these findings to a wider sample. The study may be limited by the separation of Aircrew and Engineering participants into separate groups. Participants were therefore able to discuss shared experiences, but potential differences between these groups may be better identified through combining these functions into a single focus group. The identified themes would benefit from future application of psychometric techniques to test the suitability of novel themes (such as the high level of commitment to organisational safety reputation and integration of safety awareness throughout the military aviation career) and previously established themes (such as management commitment and policy & procedures) as constructs underpinning safety culture.

References

Blazsin, H., & Guldenmund, F. (2015). The social construction of safety: comparing three realities. *Safety Scien*ce, *71* (Part A) 16-27.

Börjesson, M., Österberg, J. & Enander, A. (2011). Risk and safety attitudes among conscripts during compulsory military training. *Military Psychology*, *23*, 659-684.

Bosak, J., Coetsee, W.J. & Cullinane, S.J. (2013). Safety climate dimensions as predictors for risk behaviour. *Accident analysis and prevention*, *55*, 256-264.

Boyatzis, R. (1998). *Transforming qualitative information: Thematic analysis and code development*. Thousand Oaks, CA: Sage

Braun, V. & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77-101.

Cox, S., & Flin, R. (1998). Safety Culture: Philosopher's stone or man of straw? *Work and Stress, 12*, 189-201.

Flin, R., Mearns, K., O'Connor, P. & Bryden, R. (2000). Measuring safety climate: identifying common features. *Safety Science*, *34*, 177-192.

Gherardi, S., Nicolini, D. & Odella, F. (1998). What Do You Mean By Safety? Conflicting Perspectives on Accident Causation and Safety Management in a Construction Firm. *Journal of Contingencies and Crisis Management*, *6*, 202-213.

Glaser, B.G. & Strauss, A.L. (1967). *The discovery of grounded theory.* New York: Hawthorne.

Guldenmund, F.W. (2000). The nature of safety culture: a review of theory and research. *Safety Science*, *34*, 215-247.

Guldenmund, F.W. (2007). The use of questionnaires in safety culture research-an evaluation. *Safety Science*, *45*, 723-743.

Guldenmund, F.W. (2010a). (Mis)Understanding and exploring safety culture. *Risk Analysis*, *30*, 1466-1480.

Haddon Cave, C. (2009). *The Nimrod Review*. London: The Stationery Office.

Hopkins, A. (2010). Risk Management and Rule Compliance Decision Making in Hazardous Industries, *Working Paper 72*. National Centre for OHS Regulation, Canberra.

Jeffcott, S., Pidgeon, N., Weyman, A., & Walls, J. (2006). Risk, trust and safety culture in UK train operating companies. *Risk Analysis, 26*, 1105-1121.

Kim, S., & Wang, J. (2009). Three competing paradigms: vertical and horizontal integration of safety culture research. *International Review of Public Administration*, *14*, 63-86.

Kitzinger, J. & Barbour, R.S. (1999). Introduction: the challenge and promise of focus groups. *Developing focus group research: Politics, theory and practice*, 1-20.

Ministry of Defence (2012). Military Aviation Authority (MAA) external audit panel (MEAP) report. The Stationery Office, London. https://www.gov.uk/government/publications/military-aviation-authority-maa-external-audit-panel-meap-report. Accessed Aug. 21, 2015.

Ministry of Defence (2014). Air Safety Annual Report, Aug 13-Aug 14. The Stationery Office, London. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/394167/DG_MAA_air_safety_annual_report_2013_2014.pdf. Accessed Aug 21, 2015.

Nordlof, H., Wiitavaaea, B., Winblad, U., Wijk, K. & Westerling, R. (2015). Safety culture and reasons for risk-taking at a large steel-manufacturing company: Investigating the worker perspective. *Safety Science*, *73*, 126-135.

O'Connor, P., O'Dea, A., Kennedy, Q. & Buttrey, S. (2011). Measuring safety climate in aviation: A review and recommendations for the future. *Safety Science, 49*, 128-138.

Pidgeon, N. (1998). Safety Culture: Key theoretical issues. *Work and Stress*, *12*, 202-216.

Richter, A. & Koch, C. (2004). Integration, differentiation and ambiguity in safety cultures. *Safety Science, 42*, 703-722.

Seo, D.C., Torabi, M.R., Blair, E.H., & Ellis, N.T. (2004). A cross-validation of safety climate scale using confirmatory factor analytic approach. *Journal of Safety Research*, *35*, 427-445.

Simard, M., & Marchand, A. (1997). Workgroups' propensity to comply with safety rules: The influence of micro-macro organisational factors. *Ergonomics*, *40*, 172-188.

Using Cognitive Work Analysis to design smart grid interfaces

Christine Chauvin, Philippe Rauffet, Mathilde Tréhin,
Pascal Berruet, & Julie Lassalle
Lab-STICC UMR CNRS 6285, University of South Brittany
France

Summary

Smart grids are electricity networks that can intelligently integrate the behaviour and actions of all users connected to them in order to deliver sustainable, economic, and secure electricity supplies efficiently. They provide a tool for consumers to control their consumption better and, in the end, to save energy. The issue is that electricity-consuming activities are habitual and routinized ones, and modifying these habits is extremely difficult. This paper indicates how the Cognitive Work Analysis framework could be used to design an interface facilitating users' comprehension of energy consumption and subsequent adoption of new behaviours.

Introduction

The promotion of sustainable consumption is an important aspect of sustainable development. However, sustainable electricity consumption appears to be a particularly difficult challenge, and households seem to constitute a particularly difficult target group (Fischer, 2008). Smart grids could be an opportunity to address energy challenges. The concept relates to "an electricity network that can intelligently integrate the behaviour and actions of all users connected to it, in order to efficiently deliver sustainable, economic and secure electricity supplies" (SmartGrids European Technology Platform, 2013). Smart grids may provide tools for consumers to control their consumption better and, in the end, to save energy. A smart grid system may transform passive consumers into decision-makers who will play a positive role in environmental issues. They might thus become "prosumers" (Mah et al., 2012) or "consum'actors", that is to say, responsible consumers. In France, Electricité Réseau Distribution France (Electricity Distribution Network France) is modernizing the electrical grid and substituting smart meters for standard meters. Pilot projects, like the smart grid SOLENN project carried out in the area of Lorient (South Brittany), aim to develop and test information and support tools for consumers equipped with smart meters. These tools should help households become more aware of their electricity consumption.

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Several studies (Mah et al., 2012; Toft et al., 2014; Perlaviciute & Steg, 2014) have identified obstacles as well as enabling factors that influence the acceptability and acceptance of smart grids. Obstacles are related to the users' fears. Some people fear invasion of their privacy due to data breaches, a degradation of the quality of service due to the possibility of power modulation, and the complexity of the system. The main drivers are related to financial incentives on the one hand and to social or environmental motivations on the other. Several authors (Kobus et al., 2013; Goulden et al., 2014) have emphasized the role that may be played by information systems in the process of acceptance of smart grids. This role is negative when they deliver data that are not easy to understand and when they seem to be "opaque". In contrast, they may facilitate behaviour changes when they are intuitive, flexible, and when they provide frequent feedback. Kobus et al. (2013) stressed that electricity is used within the context of routinized actions (turning the light on, for example), which rely on automatic processes. The major issue is therefore to design interfaces that could spark and support the development of new consumption habits.

In this paper, the Cognitive Work Analysis (CWA) methodology is used to define the main principles of an interface that could facilitate habit-changing processes. The CWA methodology was proposed by Rasmussen (1986), Rasmussen et al. (1994), and further developed and codified by Vicente (1999). This framework is used to design "ecological interfaces" designed to help knowledge workers adapt to change and novelty (Vicente, 2002). It has already been used in a large number of systems. To our knowledge, however, it has never been employed to model a smart grid system. It is a formative constraint-based approach, consisting of five successive stages: Work Domain Analysis, Control Task Analysis, Strategies Analysis, Social Organization and Cooperation Analysis, and Worker Competencies Analysis. Three of these stages are presented in this paper: Work Domain Analysis (WDA), Control Task Analysis (ConTA) and Worker Competencies Analysis (WCA). The work presented here is carried out within the context of the SOLENN project.

Method

As recommended by Stanton and Bessell (2014), interviews were used as primary source of information for construction of the products in CWA. Since the system doesn't exist yet, a semi structured interview format was used that is similar to the format described by Bisantz et al. (2003). In such an interview, questions put to experts are motivated by the concepts of the Work Domain Analysis. Experts questioned were the project manager and three information tools designers of the SOLENN project. Information collected during collective meetings, as well as documents analysis, served also to consolidate the analyst's understanding. After phase one (WDA) was completed, the functions identified were used in the subsequent phases (ConTA and WCA), to offer different perspectives on the system.

Work Domain Analysis

The WDA is the most important stage of the CWA methodology. WDA deals with the constraints that are placed on actors by the functional structure of the field or the environment in which the work occurs. This phase is associated with a modelling tool, the Abstraction Hierarchy, which can be used to break down any work domain in terms of:

- *ends* (purposes, goals) and *means* (to reach the goals) according to an implementation hierarchy;
- *whole* and *parts* according to a decomposition hierarchy.

The implementation hierarchy enables the description of a work domain in terms of five levels of abstraction: functional purpose (the purpose of the work domain, its “raison d’être”), priority measures/ abstract functions, general functions, physical processes and activities, and physical resources and their configurations. Each level is connected by a structural means-end framework linked to the next upper or lower level. It is a causal structure in physically coupled systems, obeying the laws of nature. Hence, the future system states may be predicted. The hierarchy is an intentional structure in human-activity systems such as the smart grid one. In these cases, “causality is observed through the interaction of social rules between groups of participants, and future states of the system cannot be similarly predicted” (Wong et al., 1998, p. 147).

The decomposition hierarchy is destined to break a domain down into sub-systems, then each sub-system into functional units, each unit into sub-sets, and finally each sub-set into components. Both hierarchies are used to define the informational content and structure of an interface (Vicente & Rasmussen, 1992).

Control Task Analysis

ConTA is related to the activity required for meeting the purpose of a system. Naikar et al. (2006) proposed to characterize this activity as a set of recurring work situations, work functions, or control tasks. Work functions are related to functions to be performed in a work system. They are defined at the purpose-related function level or at the object-related process level in the abstraction hierarchy (Jenkins et al., 2008). They may be performed in different work situations.

Worker Competencies Analysis

Worker competencies are related to the modes of cognitive control that may be required to realize a control task. WCA relies on the Skill-Rule-Knowledge taxonomy proposed by Rasmussen (1986) to distinguish three kinds of cognitive control modes:

- the skill-based level involving the use of automated behaviours with no conscious control (such as mental math calculations) and patterns of automated and highly integrated actions;
- the rule-based level involving the correspondence of an “if-then” type between signs and an appropriate action (if such a sign is present, then such an action is executed);

- the knowledge-based level involving declarative knowledge. This level corresponds to sequential and analytical reasoning that is based on an explicit representation of goals and a mental model of the functional properties of the environment. Using it is costly because it requires focused symbolic attention.

Results

Work Domain Analysis and Abstraction Hierarchy

Table 1 shows the Abstraction Hierarchy of a smart grid system. The system was refined into three levels: the whole system (smart grid at territory level), sub-systems (each household fitted with a smart meter), and the function units in which electricity is used.

The main purpose of a smart grid system (i.e. functional purpose) is to deliver sustainable, economic, and secure electricity supplies (Toft et al., 2014). The SOLENN project has two main purposes: *i)* securing the electricity supplies in order to decrease the risk of load shedding, and *ii)* optimizing the energy consumption; the latter concerns the system at both territory and household level.

Values and priority measures represent the criteria that must be respected for a system to meet its functional purposes. Criteria are fundamental laws, principles, or values that can serve as a basis for the evaluation. In a smart grid system, the main criteria concern the measurement of the energy demand: *i*) at territory level, energy demand should be less than the maximum electricity production capacity; *ii*) at the household level, energy demand should be less than the available kVA power, consumption should be as limited as possible, given the main features of the household (number of persons), the dwelling (surface, year of construction), and the environment (location, season, outside temperature). Consumption must decrease towards an optimum bounded by an acceptable level of comfort. Several reference values (in kWh per year) could be considered: a theoretical optimum, the mean consumption of similar profiles, past consumption of a specific household in a similar context. Consumption decreasing can also be translated into expense decreasing (in euros) and carbon footprint decreasing (CO_2). It is important to represent these functional relations explicitly on the interface, so that operators can determine when the process constraints are broken (Vicente & Rasmussen, 1992).

The third level (Purpose-related functions) represents the functions that a system must be capable of supporting, so that it can fulfil its functional purposes. A smart grid system can modulate the available power remotely (in case of network congestion or incident), deliver information to the energy producer, the supplier, and the consumer, and provide information for the management of energy consumption. At the household level, the main function consists in managing the electricity consumption. This function can be considered at the level of function units: managing the electricity consumption related to heating, producing warm water, etc.

The fourth level (Object-related processes) represents the functional processes or the functional capabilities or limitations of the physical objects in a system. Among the objects listed at the fifth level, it is important to notice that the number of the current clamps is limited (to three or four). It is therefore not possible to know the consumption of each device. Furthermore, information is not transmitted continuously but according to discrete time steps. The power of each electric device is also a feature that must be taken into account.

Table 1. Abstraction Hierarchy of the smart grid system in the SOLENN project.

	Smart grid at territory level	Household level	Function units
Functional purposes	Securing the electricity supply, managing the electricity demand	Optimizing electricity consumption (i.e., obtaining reasonable consumption with a good level of comfort)	
Values & priority measures	Avoiding load shedding; energy demand< maximum energy production; minimizing consumption (GWh); reducing peak load, straightening load curb	Energy demand < kVA power; reducing electricity consumption (kWh), electricity expenses (€) and carbon footprint (CO2)	
Purpose-related functions	Modulation of the available power; consumer information and coaching.	Managing the electricity consumption	Heating, producing warm water, cooking, lighting, cooling, washing dishes, washing clothes, drying, cleaning, using electronic devices
Object-related processes		Level of information breakdown, time step of information delivery	Power of each device, power demand and duration of use
Physical objects	Linky information system	Smart meters (Linky) Electric switch Sub-metering system Website, applications	Electric devices Computers, tablets

The last level may represent physical or artificial objects (such as artefacts and infrastructure). In the SOLENN project, the main physical and artificial objects that may be installed in the household or made available to the consumers are:

- a smart meter (enabling a two-way communication between the meter and the central system and supplying information concerning the daily consumption);

- current clamps and a sub-metering device (enabling the measure of the individual circuit of energy demand and providing information regarding the consumption of specific devices or groups of devices);
- electric devices;
- websites and individual applications (providing information regarding consumption at defined time steps, showing the load profile);
- individual and collective coaching (offering consumers advice).

The Abstraction Hierarchy (AH) provides an informational basis, since the model may be converted into a list of variables. Its main benefit is to provide information that would be useful to cope with unanticipated events. As explained by Bisantz and Vicente (1994, p. 84), AH is intended to represent the set of goal-relevant constraints governing the operation of the controlled system. This type of representation can be described as event-independent, since it provides information about system structure that is independent of any specific event or consequence of events. This is in contrast to representations that are event-dependent, consisting of the symptoms or corrective procedures associated with a set of events, or classes of events, which must be identified beforehand. This last type of work domain representation cannot, by definition, help operators consistently cope with unanticipated events.

In the case of smart grid systems, specific recommendations have already been given concerning the information that should be delivered to the consumer (Lewis et al., 2012; Bouchet & Chauvin, 2015). The Abstraction Hierarchy model adds new recommendations concerning the display of the process constraints and the display of possibilities for actions within these constraints. They are, at the household level, *i*) the display of the electricity demand compared to the kVA available power, and *ii*) the display of electricity consumption results compared to a reference value.

Control Task Analysis

As said before, activity required for meeting the purpose of a system may be characterized as a set of work situations, work functions, or control tasks.

In the case of a smart grid system, two main situations may be distinguished: normal and incidental situations. The incidental situations are related to constraints affecting the electric grid such as peak loads. In such cases, three work functions are expected: power modulation, consumer information, and consumer response /action by reducing consumption levels.

The decision ladder is used to decompose activity into a set of control tasks for each work situation and/or work functions. It uses the formalism defined by Rasmussen to model a diagnosis and decision task (Rasmussen, 1986). In this formalism, rectangular boxes represent information-processing activities and circles represent states of knowledge resulting from these activities.

Figure 1 shows the different stages of the task realised in order to manage the electricity consumption in incidental situations. A similar figure could be drawn for normal situations.

The ascending left side of the ladder brings together all the steps of situation analysis (from detecting abnormal conditions to evaluating consequences on the system status). The descending right side of the ladder relates to the various steps of action planning (task specification, then procedure) and ends with the performance of the action itself.

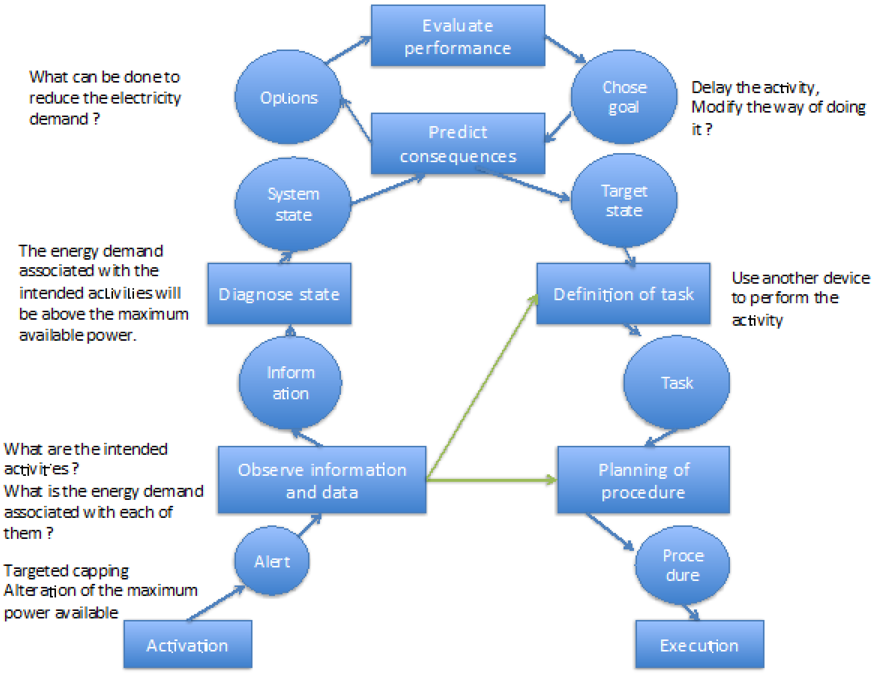


Figure 1. Model of the "management of electricity consumption" control task in incidental situations.

In incidental situations, the control task begins when consumers are informed about a targeted capping or about a change in the maximum power they can draw. The stages following the activation consist in: *i)* estimating future consumption by considering intended uses or activities, *ii)* comparing the desired consumption with the new limit and determining whether it is under or above it, *iii)* predicting consequences (are the foreseen activities possible or not?), *iv)* examining what could be done to reduce the electricity demand, *v)* choosing a goal, which could be either to delay the activity or to modify the manner of doing it, and *vi)* in the latter case, deciding to use another device to perform the activity.

The decision ladder also shows alternative routes (i.e., shortcuts) connecting the two sides, thus signalling expert operators' heuristic decision making. Heuristic decision making corresponds to operators' know-how and rests upon inductive reasoning that associates states of the environment to actions that have been shown as successful in similar situations. It thus depends on empirical correlations between evidence and actions observed in familiar scenarios.

From a practical point of view, such a model brings useful elements to the identification and display of important information (the new maximum power, the energy demand associated with a given activity, etc.); in that sense, it complements the Work Domain Analysis. It also leads to considering the possible ways to support expert behaviours such as helping users understand the relation between particular activities, devices, and energy demand (see Figure 2). In that sense, task analysis complements the Worker Competencies Analysis.

Worker Competencies Analysis

The Skill-Rule-Knowledge taxonomy is highly relevant for the design of smart grid interfaces. One of the main issues is to break up routinized behaviours that are not reflected upon and that may be seen as "environmentally detrimental habits" (Matthies, 2005; Fischer, 2008) and to induce a conscious decision so that new norms and considerations should be taken into account. This approach induces, first, extra effort but the creation of new routines is expected at a medium term.

Fischer (2008) indicated that several kinds of feedback may be used to support such a decision process, assuming that "feedback is most effective if it: *i*) successfully captures the consumer's attention, *ii*) draws a close link between specific actions and their effects, *iii*) activates various motives that may appeal to different consumer groups, such as cost savings, resource conservation, emissions reduction, competition, and others" (p. 83). Concerning the second point, Fischer explained that successful feedback involves appliance-specific breakdown. Costanza et al. (2012) showed that consumers go beyond the disaggregation of appliance loads, and deal with higher levels of abstraction such as "oven roast dinner". Such reflections about consumption patterns facilitate the creation of rules (i.e., rule-based level) associating specific consumption events (described in terms of start and end timestamps and the amount of energy consumed) and specific activities involving the use of one or more electrical appliances.

Towards an ecological interface

The twofold objective of an ecological interface is to encourage the use of skill-and rule-based behaviour while providing support for otherwise more effortful behaviour to cope with unfamiliar and unanticipated situations (Vicente, 2002). In the case of a smart grid system, one of the main goals is to provide support for effortful behaviour, so that the user can elaborate rules - or action schemes - facilitating understanding and decision-making in normal or incidental situations. To this end, several information elements must be presented together. The Abstraction Hierarchy and the Control Task Analysis facilitate their identification. These information elements are the criteria to be respected and their value, all the active work functions at the function unit level (for example, washing clothes, heating, producing warm water, washing dishes), the devices used (washing machine, heaters, boilers, dishwasher) and their characteristic features (typical duration of use and consumption).

As pointed out earlier, the number of the current clamps is limited, and it is therefore not possible to measure the consumption of each device directly. However, users could manually note their energy consumption log on a load curve, as proposed by Costanza et al. (2012). The system could then analyse the consumption associated with a specific event and display this information. Figure 2 shows a summary representation of the information collected. It compares devices used for the same activity with numerical and graphic representations and highlights the amount of energy used for each activity. This representation would help users determine the most energy-intensive devices for a given period. It intentionally uses the two notions of power (W) and energy (Wh) to facilitate the learning of those concepts.

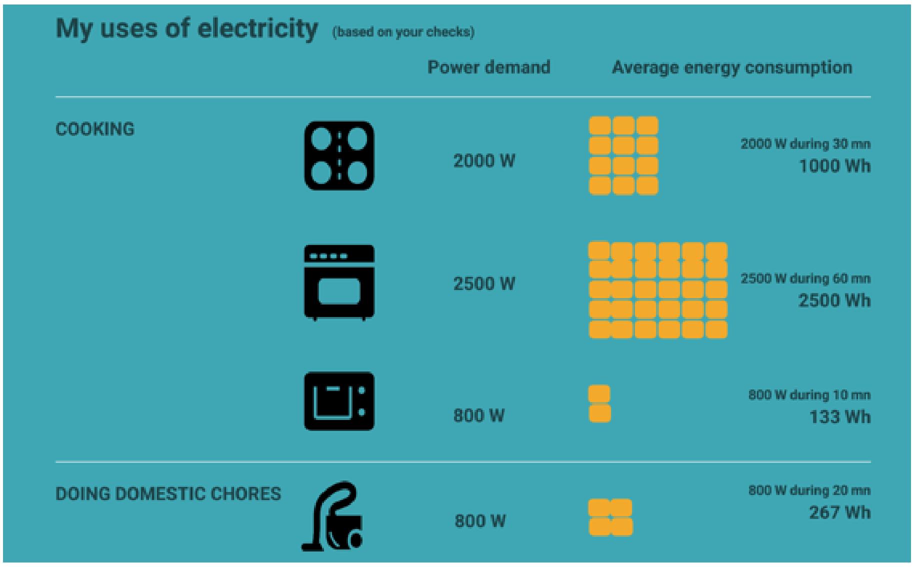


Figure 2. Displaying the average energy consumption per device and activity. Icons were created in the context of the Noun Project (https://thenounproject.com/).

This information would help consumers evaluate their possibilities of use in an incidental situation. Figure 3 shows that they could manipulate boxes representing specific uses, in order to check what is possible, given the reduction of available power. This function would help users plan their domestic tasks in a constrained situation, owing to the simultaneous representation of *i)* the energy demand associated with each activity and *ii)* the maximum available energy. It is not simply a static representation, as in the proposal by Costanza et al. (2012), since users could play an active role. By manipulating blocks representing specific uses, consumers could check the configurations of devices that would be allowed, given that the available power will be reduced. In this way, users could anticipate, develop skills, and adopt new reflex actions in restrictive situations. Once again, what is expected thanks to repeated use of such an interface is the creation of new consumption habits.

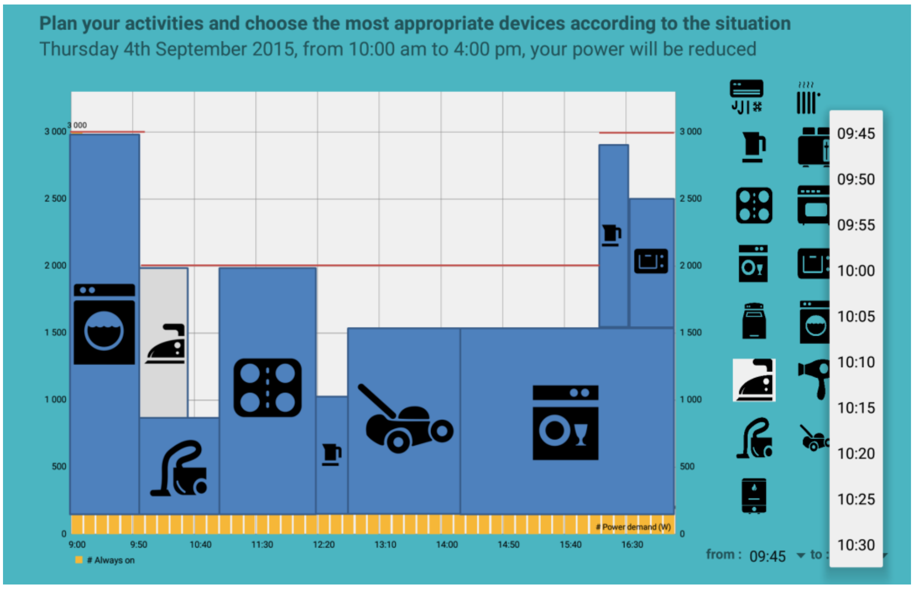


Figure 3. Displaying the range of possibilities given the available power

Conclusion

Bennett and Flach (2011) summarized the goal of an ecological interface when viewed through the lens of Work Domain Analysis on the one hand and of the decision ladder on the other. At the WDA level, it is to make the constraints at all levels of the abstraction hierarchy visible; ideally, "the operator should be able to see the state of the work domain in relation to the goals, the costs, and the fields of possibilities associated with physical and regulatory laws and organizational layout" (Bennett & Flach, 2011, p.103). When considering the decision ladder, they recommended that the representation provide signals and signs that map directly onto states/ constraints of the work processes to support productive thinking (e.g., chunking, automatic processing, and recognition-primed decisions). It seems possible and relevant to apply these principles to the design of a smart grid interface. Showing the consumption associated with specific use as well as the maximum available power should help consumers understand their consumption and adopt new uses. This proposal will be tested with the trial participants taking part in the SOLENN project.

Cognitive Work Analysis is useful for designing a new, first-of-a-kind system. It is based, in this case, on Engineering-Expert-Knowledge. In the framework of the SOLENN project, this analysis will be completed by interviews with users equipped with a Linky smart meter and by observations focusing on their use of the information and support tools. At this stage of the study, we will investigate the utility and usability of the interface, its capacity to improve the users' understanding and management of their consumption, as well as the system's acceptance.

Besides Cognitive Work Analysis, other approaches could be used to design information tools for smart-grid systems. An alternative could be to induce the desired behaviour thanks to pervasive technologies (Fogg, 2009) or with positive reinforcements and indirect suggestions as advocated by the "nudge" approach (Thaler & Sunstein 2008).

Acknowledgments

This work was carried out within the context of the SOLENN project supported by the Environment and Energy Mastery Agency (Agence de l'Environnement et de la Maîtrise de l'Energie - ADEME) as part of the French government-led Investments for the Future Programme.

References

Bennett, K.B., & Flach, J.M. (2011). *Display and interface design: Subtle science, exact art*. Boca Raton: CRC Press.
Bisantz, A.M., & Vicente, K.J. (1994). Making the abstraction hierarchy concrete. *International Journal of Human-Computer Studies*, *40*, 83-117.
Bisantz, A.M., Roth, E., Brickman, B., Gosbee, L.L., Hettinger, L., & McKinney, J. (2003). Integrating cognitive analyses in a large-scale system design process. *International Journal of Human-Computer Studies*, *58*, 177-206.
Bouchet, M., & Chauvin, C. (2015). IHM persuasive et pilotage des actions de Maîtrise de l'Energie. In N. Bonnardel, L. Pellegrin, and H. Chaudet (Eds.), *Actes du Huitième Colloque de Psychologie Ergonomique (EPIQUE 2015)*. Paris, France: Arpege Science Publishing.
Costanza, E., Ramchurn, S.D., & Jennings, N.R. (2012). Understanding domestic energy consumption through interactive visualisation: a field study. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 216-225). ACM.
Fischer, C. (2008). Feedback on household electricity consumption: a tool for saving energy? *Energy Efficiency*, *1*, 79-104.
Fogg, B.J. (2003). Persuasive Technology. *Using computers to change what we think and do.* Amsterdam: Morgan Kaufmann.
Goulden, M., Bedwell, B., Rennick-Egglestone, S., Rodden, T., & Spence, A. (2014). Smart grids, smart users? The role of the user in demand side management. *Energy Research & Social Science*, *2*, 21-29.
Jenkins, D.P., Stanton, N.A., Salmon, P.M., Walker, G.H., & Young, M.S. (2008). Using cognitive work analysis to explore activity allocation within military domains. *Ergonomics*, *51*, 798-815.
Kobus, C.B., Mugge, R., & Schoormans, J.P. (2013). Washing when the sun is shining! How users interact with a household energy management system. *Ergonomics*, *56*, 451-462.

Lewis, P.E., Dromacque, C., Brennan, S., Stromback, J., & Kennedy, D. (2012). *Empower demand 2: Energy efficiency through information and communication technology – best practice examples and guidance*. Technical Report. VaasaETT, Global Energy Think Tank. [Online] Available from: www.esmig.eu/sites/default/files/final_empower_2_demand_report_final_distr 2.pdf.
Mah, D.N.Y., van der Vleuten, J.M., Hills, P., & Tao, J. (2012). Consumer perceptions of smart grid development: results of a Hong Kong survey and policy implications. *Energy Policy*, *49*, 204-216.
Matthies, E. (2005). Wie können Psychologlnnen ihr Wissen besser an die PraktikerIn bringen? Vorschlag eines neuen, integrativen Einflussschemas umweltgerechten Alltag shandelns. [How can psychologists improve their outreach towards practitioners? A suggestion for a new, interactive model of environmentally sound everyday practices]. *Umweltpsychologie, 9,* 62 – 81.
Naikar, N., Moylan, A., & Pearce, B. (2006). Analysing activity in complex systems with cognitive work analysis: concepts, guidelines and case study for control task analysis. *Theoretical Issues in Ergonomics Science*, 7, 371-394.
Perlaviciute, G., & Steg, L. (2014). Contextual and psychological factors shaping evaluations and acceptability of energy alternatives: integrated review and research agenda. *Renewable and Sustainable Energy Reviews*, *35*, 361-381.
Rasmussen, J. (1986). *Information processing and human–machine interaction*. Amsterdam: North Holland.
Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive systems engineering*. New York: Wiley.
Smart Grids - European technology platform. (2013). European technology platform for the electricity networks of the future - FAQs about ETP Smart Grids.
Stanton, N. A., & Bessell, K. (2014). How a submarine returns to periscope depth: Analysing complex socio-technical systems using Cognitive Work Analysis. *Applied Ergonomics, 45*, 110-125.
Thaler, R.H, Sunstein, C.R. (2008). Nudge: Improving decisions about health, wealth, and happiness. *Constitutional Political Economy*, *19*, 356-360.
Toft, M.B., Schuitema, G., & Thøgersen, J. (2014). Responsible technology acceptance: model development and application to consumer acceptance of Smart Grid technology. *Applied Energy*, *134*, 392-400.
Vicente, K.J. (1999). *Cognitive work analysis: toward safe, productive, and healthy computer-based work*. Mahwah, NJ: Erlbaum.
Vicente, K.J. (2002). Ecological interface design: progress and challenges. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *44,* 62-78.
Vicente, K.J., & Rasmussen, J. (1992). Ecological interface design: theoretical foundations. *IEEE Transactions on Systems, Man and Cybernetics, 22*, 589-606.
Wong, W.B., Sallis, P.J., & Hare, D.O. (1998). The ecological approach to interface design: applying the abstraction hierarchy to intentional domains. In P.B Calderon and B. Thomas (Eds.), *Computer Human Interaction Conference, 1998. Proceedings. 1998 Australasian* (pp. 144-151). Los Alamitos, CA: IEEE Computer Society Press.

Driving with a multi stage warning system in the head-up display – How do drivers react upon it?

Susann Winkler, Juela Kazazi, & Mark Vollrath
Technische Universität Braunschweig, Germany

Abstract

Driver warnings are a good way to reduce accidents in urban areas as they support drivers in complex and critical situations. This was shown in a previous driving simulator study of the research project UR:BAN. Some critical situations might demand emergency braking, others might require mere attention re-allocation or moderate braking. Thus, a multi stage warning system (*warning* and *acute warning*) for the head-up display was developed inducing different driver reactions depending on the situation’s criticality and intervention requirements. A driving simulator study (N = 24 drivers; including eight urban scenarios) was conducted, in order to examine whether drivers understood this multi stage warning system and to what extent learning was required. The test included a first drive without any warning support, a learning phase and an end phase with previously unknown scenarios. The data show that the warning system is intuitively understood by the drivers, without learning being noticeably required and drivers differentiating well between the two warning stages.

Introduction

Assisting drivers in urban areas and their safety-critical situations (e.g., with collision warning systems) is a good way to promote safe driving. This is one of the aims of the research project UR:BAN (Manstetten et al., 2013; www.urban-online.org), from which the present study arose. Most accidents still occur in urban areas as compared to other road types. For example, they amounted up to 73.3% of all accidents on German roads in 2013 (Statistisches Bundesamt, 2014). The safety-critical situations, which one experiences in urban areas, are not only challenging for drivers, but also for assistance system developers. They show very diverse characteristics, such as concerning their location (e.g., at an intersection, on a straight or curved road), the type of critical objects involved in them (e.g., pedestrians, bicyclists, obstacles) and their behaviour (e.g., static or dynamic, accelerating or decelerating) as well as their criticality (low to high). Assistance systems have to gather and evaluate the respective information correctly as fast as possible in order to supply the drivers with the needed support (Gläser et al., 2014; Heimes & Nagel, 2002; Herrmann & Schroven, 2012). Depending on the safety-critical situations, the reactions required by the drivers to avoid a possible collision

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

also differ (Schmidt et al., 2009; Khanafer et al., 2009), especially as the reaction intensity needed increases with the decrease of time-to-collision (TTC) (Dingus et al., 1998; Najm et al., 1994).

Assistance systems for collision avoidance, as focussed here, can be staged on a continuum with different urgency levels (Dingus et al., 1998) allowing drivers to optimally adapt their behaviour according to the respective requirements of critical urban situations. Such collision warning systems might start at a very early purely informative level, but become gradually intrusive with drivers further approaching the safety-critical situation, in order to activate them more and more strongly (Campbell et al., 2007; Naujoks & Neukum, 2014; Zarife, 2014). Finally, right before the imminent collision an acute warning or eventually an automatic vehicle intervention like an automatic emergency brake might be triggered (General Motors Corporation & Delphi-Delco Electronic Systems, 2002). On this described continuum many further warning stages are possible, from advising drivers to raise their attention and allocate it properly (*information* and *prewarning*) to demanding them to decelerate, start braking moderately (*warning*) and eventually to eliciting an emergency brake (*acute warning*) (see also Diederichs et al., 2010; Petermann-Stock & Rhede, 2013; Werneke et al., 2014; Winkler et al., 2016). However, in a given situation and depending on the reactions of the drivers, not all warning stages have to become active. Thus, the output of the collision warning system would be adapted accordingly, in order to support drivers optimally.

With such a warning system and especially by help of the earlier warning stages in the described continuum (e.g., *prewarning* or *warning*), drivers might already be able to prepare for upcoming safety-critical situations before they become highly critical, if they can be detected far in advance (for example by means of car2X communication, see also Engel et al., 2013; Nöcker et al., 2000; Röglinger & Facchi, 2009; Weiß, 2011). As soon as drivers adapt their driving behaviour appropriately, for example by releasing the accelerator pedal or even braking moderately, later warning stages like an *acute warning* can be avoided, which usually succeed earlier ones. However, a situation might also be very time critical and hard to detect for the system, so that a late warning (*acute warning*) might have to be given directly, in order to enable drivers to react very fast and strong, for example with an emergency brake. Therefore, a multi stage warning system, which supports drivers integratively and adaptively to the situation’s criticality and intervention need, seems optimal for drivers (Jones & Hansman, 2007; Naujoks & Neukum, 2014; Zarife, 2014), especially in urban areas.

According to these ideas and based upon the results of previous studies (Kazazi et al., 2015; Winkler et al., 2015), a multi stage warning system (for urban areas) was examined in this study with the focus on the two warning stages *warning* and *acute warning*. In less critical situations, like a lead vehicle braking in front of the ego vehicle keeping a safe distance, it triggers a first warning stage (*warning*), rather appealing to the drivers’ attention and readiness to brake if need be. Whereas, in highly critical situations or such with increasing criticality, a more forceful driver reaction is demanded, activating drivers strongly and quickly (second warning stage: *acute warning*). A head-up display (HUD) projects both warning stages into the windshield as displaying sequential warning stages integratively in one display improves drivers' response time onto warning presentations (Singer et al., 2014). The HUD as a location for warning presentation additionally allows drivers to perceive them right in their field of view without taking their eyes off the road (Ablaßmeier et al., 2007) and thus saves time for drivers to react upon them (Watanabe et al., 1999). Aim of the multi stage warning system is to trigger strong and fast brake reactions in drivers in order to avoid collisions or reduce collision severity. Therefore, a stop sign is presented in the HUD accompanied by an additional acoustic warning signal as an *acute warning* (second warning stage), when an upcoming situation demands it due to its increased criticality. However, preferably the system warns drivers sufficiently in advance so that the situation can be resolved by a moderate driver reaction, as elicited by a caution sign displayed in the HUD as a *warning* (first warning stage). Whether drivers intuitively understand such a multi stage warning system (even when experiencing it for the first time) or to what extent it has to be learned by drivers in order to trigger an appropriate reaction, still has to be investigated yet. Furthermore, it is still unclear how (well) drivers actually react upon both warning stages and can differentiate between them. Thus, these questions are examined in the presented study.

Summing up, this driving simulator study evaluates a multi stage warning system for urban areas based upon drivers' behaviour in eight different safety-critical situations. Half of these safety-critical situations are analysed in detail for this paper, while the other half will be reported in a further paper. Firstly, possible learning effects through driving with the multi stage warning system in four repeated safety critical situations with different criticalities are investigated. The question is whether the frequent experience of the multi stage warning system in various critical situations might influence the driver reactions over time (e.g. faster brake reactions when the same situation is experienced again). Secondly, the differentiation between the two warning stages (first stage: *warning*, second stage: *acute warning*) is. As the first warning stage is presented quite early in a less critical situation, a more relaxed driver reaction is expected compared to the second warning stage, which would be presented in a more critical situation or if the former less critical situation becomes more critical. Thus, the driving behaviour (e.g., brake reaction time, time to maximum braking and maximum braking value) should differ between both warning stages.

Method

Multi stage warning system

Drivers were supported by a multi stage warning system, consisting of two warning stages (*warning* and *acute warning*). For its visualization a multicolour HUD was used. In form of traffic signs measuring maximum 15x15 cm (4° visual angle), both warning stages were projected driver-centered into the windshield right above the engine hood overlapping with the simulated driving scenery. For the first warning stage ("Warning the driver"), a caution sign was displayed in the HUD, when the time-to-collision (TTC, time left until the ego vehicle would collide with the safety-critical object, if the speed difference stays the same; Hayward, 1972) between the ego vehicle and the safety-critical object went below 8 s, but stayed above 2 s (see

Table 1). If the drivers did not react sufficiently or the situation was already that critical so that the TTC went below 2 s, a second warning stage (“Eliciting a last driver reaction”) was triggered. As can be seen in Table 1, in the second warning stage a stop sign was displayed in the HUD accompanied by an additional acoustic warning signal. The *acute warning* lasted until the ego vehicle crossed the virtual crossing point of the safety-critical object or collided with it.

Table 1. Implemented multi stage warning system with the two warning stages: warning (W_1; solely visual) and acute warning (W_2; visual and acoustic) including their timing based on the time-to-collision (TTC, Hayward, 1972) of the ego vehilce and the safety-critical object.

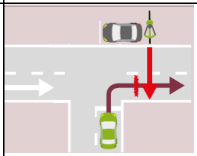
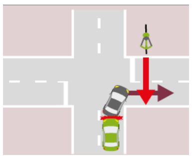
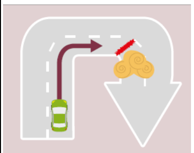
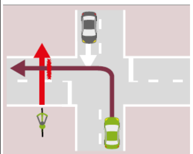
Warning stage	*Visual*	*Acoustic*	*Timing*
Warning (W_1)		-	2 s < 8 s
Acute warning (W_2)		1 kHz (“Beep”)	< 2 s

Driving simulator and implemented scenarios

For the study the fixed base medium fidelity driving simulator of the Technische Universität Braunschweig was used. Its seat box is equipped with typical car interiors like a steering wheel with force feedback, accelerator and brake pedals as well as three LCD screens with 1400x1050 px resolution serving as rear-view mirrors (left, middle and right). Moreover, four cameras are installed at the seat box, covering the drivers’ face, feet (for pedal operation), the scenery and a time stamp running with the scenery to synchronise the video with the driving data. The virtual urban scenery is projected onto three silver screens (left, ahead, right), providing the drivers with a 180° field of view at about 2.1 m distance from the driver’s seat. The driving simulation is created by the SILAB 4.0 software (from WIVW, Krüger et al., 2005; see www.wivw.de) and accompanied by an acoustic simulation of traffic sounds like wind and engine noises for a rather realistic impression of sitting in a real car.

An urban area wih a speed limit of 50 km/h served as the test track. Drivers went straight unless a voice output and an arrow near the speedometer told them to turn. The four scenarios analysed are a subset of in sum eight examined scenarios with various criticality, which will be reported in a further paper. As described in Table 2, this paper comprises three less critical scenarios (S_3W_1, S_4W_1 and S_5W_1) with a bicyclist, a lead vehicle and an obstacle as safety critical objects and one quite critical scenario with a crossing bicyclist (S_6W_2). As the critical scenarios demanded drivers to react immediately with an emergency brake, a direct *acute warning* (W_2) was presented (see Table 1). In the less critical scenarios, drivers were solely required to decelerate moderately. Therefore the *warning* (W_1) was used. If the drivers did not react accordingly, eventually the *acute warning* followed.

Table 2. Depiction of the four examined scenarios and (first triggered) warning stages: S_3W_1) Bicyclist, S_4W_1) Lead vehicle and S_5W_1) Obstacle (requiring a first stage warning – warning) as well as S_6W_2) Bicyclist (demanding a direct second stage warning – acute warning).

Scenario and warning stage	*Picture*	*Description*
S_3W_1) Bicyclist 		When turning right at a t-junction with oncoming traffic a bicyclist hidden by parking vehicles crosses the ego vehicle's path from left to right
S_4W_1) Lead vehicle 		When following a lead vehicle decelerating onto 8.5 m/s and indicating a right turn, it stops suddenly at the intersection for a bicyclist crossing from left to right
S_5W_1) Obstacle 		When driving straight ahead, a hay bale blocking the ego vehicle's path suddenly becomes visible after being hidden from drivers' sight by a curve
S_6W_2) Bicyclist 		When turning left at an intersection with oncoming traffic, a bicyclist (activated by the ego vehicle crossing the centre line) crosses the ego vehicle's path from left to right (contrary to in Germany allowed direction)

Experimental design and dependent variables

Drivers went through four learning phases (L, see Table 3), in order to be familiarised with the warning system. As can be seen in Table 3, in the first three learning phases (L1-L3) they were confronted with a repetition of scenarios in randomised order (within-subjects design), whereas the last learning phase (L4) comprised two new scenarios (one quite critical and one less critical scenario), onto which the learned knowledge was to be applied. Furthermore, L1 was directly attached to the training drive, comprising a quite critical and a less critical scenario, and in contrast to all following learning phases, drivers were not supported by the warning system in L1. In sum, every driver encountered sixteen safety-critical situations (two in L1, six each in L2 and L3, two new scenarios in L4). The scenarios experienced in L1 and L4 were interchanged between the drivers, leading to two groups of drivers (see Table 3), while the order of the scenarios within each learning phase was randomised.

Table 3. Organisation of the eight scenarios ($S_{1...8}$) over the four learning phases (L1-L4), being randomised within each learning phase, with the three specific warning system support forms (without – W_0; warning – W_1; acute warning – W_2) and the two groups of drivers A and B, with the four scenarios analysed for this paper in bold ($S_{3...6}$).

Group	*L1*	*L2*	*L3*	*L4*
A	S_1W_0, S_2W_0	S_1W_1, S_2W_2, **S_3W_1, S_4W_1, S_5W_1, S_6W_2**	S_1W_1, S_2W_2, **S_3W_1, S_4W_1, S_5W_1, S_6W_2**	S_7W_1, S_8W_2
B	S_7W_0, S_8W_0	S_7W_1, S_8W_2, **S_3W_1, S_4W_1, S_5W_1, S_6W_2**	S_7W_1, S_8W_2, **S_3W_1, S_4W_1, S_5W_1, S_6W_2**	S_1W_1, S_2W_2

In order to see how drivers react upon the multi stage warning system, the driving behaviour in each scenario was recorded. Table 4 describes the three driving parameters analysed in detail.

Table 4. Analysed driving behaviour variables.

Variable	*Unit*	*Description of variable*
Brake reaction time	s	Time from onset of the *warning/acute warning* until the pressing of the brake pedal
Time to maximum braking	s	Time from onset of the *warning/acute warning* until the maximum braking value is reached
Maximum braking value	%	Maximum pressing of the brake pedal after onset of the *warning/acute warning* until its offset in percent of the sample maximum

Participants

A total of twenty-four drivers were tested in a mixed design (13 female, 11 male; M = 26.8 years, SD = 8.2 years). The average driving experience was 9.2 years (SD = 8.5 years) and the annual mileage was mainly less than 3000 km. All participants had normal or corrected-to-normal visual acuity. Simulator training was required for participation in order to avoid simulation sickness. Drivers were compensated with 10 € for a successful participation or received course credits (if students at the Technische Universität Braunschweig).

Procedure

After being welcomed, the drivers received a written instruction about the objectives and the procedure of the experiment, signed a consent form and filled out a demographic questionnaire. Then the drivers received a 15 min training to get accustomed with the driving simulator. When the training was successfully completed, drivers were instructed about the multi stage warning system and evaluated it (system acceptance) a priori based on their expectations. The following test drive took about 15 min. Afterwards drivers filled out two post-hoc questionnaires (system acceptance and subjective ratings). Subsequently, an interview about the warning system finished the experiment. Finally, drivers were thanked for their participation and compensated.

Data analysis

For data analysis IBM SPSS Statistics 21 was used. This paper analysed the four scenarios, which all 24 drivers experienced twice in learning phase L2 and L3. The results of the driving behaviour with the two warning stages are presented. The driver reactions upon the warning system in L2 and L3 were compared by paired t-tests (or Wilcoxon signed-rank tests if normality was not given) in order to look for learning effects over time. Furthermore, the driver behaviour was examined, in order to analyse how well drivers differentiate between the two warning stages. Concerning the warning stage differentiation, paired t-tests (or Wilcoxon signed-rank tests if normality was not given) compared the scenarios S_3W_1 and S_5W_1 with a *warning* to scenario S_6W_2 with a direct *acute warning* for L3. Learning phase L3 was considered here, as drivers by then experienced the warning system multiple times, so it supposedly had been learned sufficiently if needed.

The number of participants in the groups was reduced (see Table 5 in the results section), as drivers had to be excluded from statistical analyses and categorized as nonreacters, if they initiated a brake reaction before the onset of the warning system (including brake reaction times under 0.2 s), did not brake at all or were extreme outliers. For significant results, *r* is given as an estimate of the effect size. A significance level of $p \leq .05$, corrected by Bonferroni due to multiple tests, was adopted for all statistical tests. As only a few accidents occurred in the four scenarios considered, the number of collisions is not further regarded here (S_3W_1-L2: N = 1; S_4W_1-L3: N = 1; S_6W_2-L2: N = 6, L3: N = 2). The other scenarios, learning phases, and the subjective data will be reported in another paper.

Results

Learning to drive with the multi stage warning system

Table 5 shows the number of drivers who reacted to the *warning*, the *acute warning* or to neither of the warning stages (e.g., they started braking before the warning system onset) in the learning phases L2 and L3 for all four examined scenarios. In scenario S_3W_1, almost half of all drivers reacted consistently onto the *warning* in both learning phases. Yet, from the five drivers reacting to the *acute warning* in L2 none reacted upon it in L3 anymore, but rather shifted to reacting upon the *warning* (N = 3) or even beforehand (N = 1; see Table 5). This might indicate some learning effects. Similarly, in S_4W_1 half of the ten drivers, who had reacted to the *acute warning* in L2, already reacted to the *warning* in L3 and one even earlier. However, in general this scenario is rather diverse concerning the according driver reactions. As Table 5 further shows, for scenario S_5W_1 consistently most drivers reacted to the *warning* in both learning phases (N = 21), without anyone receiving an *acute warning*. In scenario S_6W_2 sixteen out of almost all drivers initiating a brake reaction upon the *acute warning* in L2, also did so in L3 (see Table 5). However, another seven drivers avoided an *acute warning* in L3, which again might indicate some learning effects. Overall, the difficulty of the three scenarios with a *warning* seems to be slightly different. Especially in scenario S_4W_1, it is difficult to compare the driving behaviour from L2 to L3, as for example only three drivers showed a break reaction upon the *warning* in both learning phases, which is why this scenario S_4W_1 is not further regarded for the following statistical within comparisons.

Table 5. Number of drivers showing brake reactions onto the two warning stages (warning and acute warning) and excluded drivers (/) over the two learning phases (L2 and L3).

Learning phase	*L3*					
	S_3W_1			S_4W_1		
L2	Warning	Acute warning	/*	Warning	Acute warning	/*
Warning	11	0	4	3	1	4
Acute warning	3	0	2	5	4	1
/	3	0	1	2	1	3
	S_5W_1			S_6W_2		
L2	Warning	Acute warning	/*	Warning	Acute warning	/*
Warning	21	0	2	-	-	-
Acute warning	0	0	0	-	16	7
/	0	0	1	-	1	0

* Drivers excluded due to inappropriate reactions (see above).

Regarding the driver behaviour over the learning phases L2 and L3 with the *warning* in scenario S_3W_1 there were no significant learning effects (see Fig. 1; brake reaction time: $t(10) = 1.31$, $p = .220$; time to maximum braking: $t(10) = 1.04$, $p = .321$; maximum braking value: $t(10) = -0.32$, $p = .758$). The same holds true for scenario S_5W_1 (brake reaction time: $t(20) = -0.40$, $p = .694$; time to maximum braking: $z = -1.23$, $p = .217$; maximum braking value: $t(20) = 0.03$, $p = .978$). Similarly for the *acute warning*, no significant learning effects could be found in scenario S_6W_2 (brake reaction time: $t(15) = 1.89$, $p = .078$; time to maximum braking: $t(15) = 0.33$, $p = .743$; maximum braking value: $t(15) = 2.01$, $p = .063$; see Fig. 1).

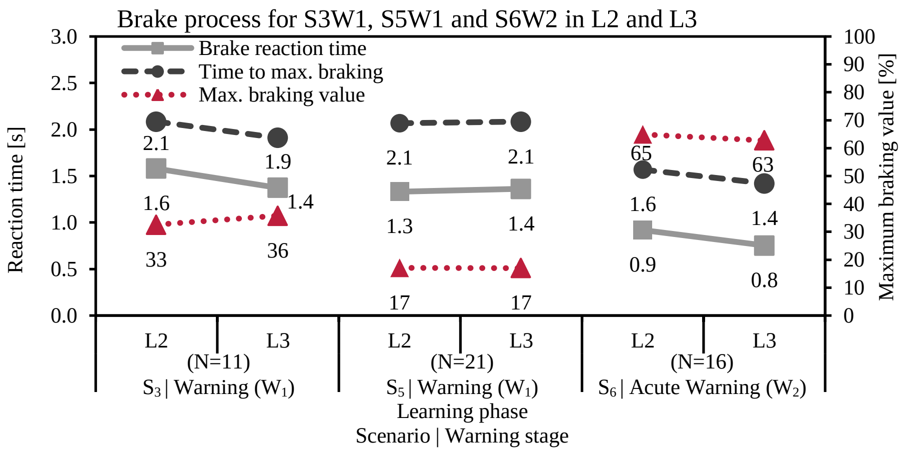


Figure 4. Mean brake reaction time, mean time to maximum braking and mean maximum braking value within scenario S_3W_1, S_5W_1 and S_6W_2 over the learning phases L2 and L3, displaying the number of participants (N) included at the bottom.

Differentiation between warning and acute warning

When comparing the two scenarios with a *warning*, S_3W_1 and S_5W_1, to scenario S_6W_2 with an *acute warning* in learning phase L3, all three driving parameters differed significantly between the two warning stages in both within-subjects comparisons, S_3W_1 vs. S_6W_2 (brake reaction time: $t(13) = 5.90$, $p < .001$, $r = .85$; time to maximum braking: $t(13) = 3.54$, $p = .004$, $r = .70$; maximum braking value: $t(13) = -4.68$, $p < .001$, $r = .79$) and S_5W_1 vs. S_6W_2 (brake reaction time: $z = -3.52$, $p < .001$, $r = -.88$; time to maximum braking: $t(15) = -2.75$, $p = .015$, $r = .58$; maximum braking value: $t(15) = -5.16$, $p < .001$, $r = .80$, see Fig. 2). As can be seen in Figure 2, in general the brake reactions were substantially faster and stronger with the *acute warning* and the time to maximum braking was also shorter for the *acute warning* compared to the *warning*.

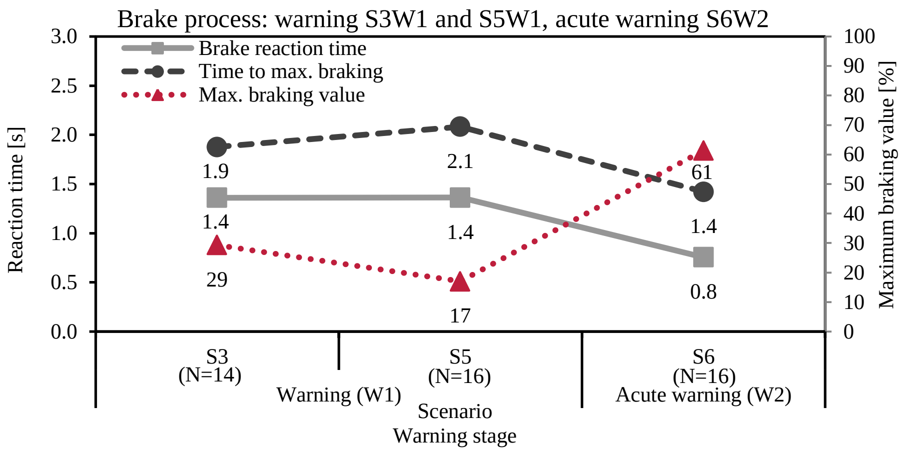


Figure 5. Mean brake reaction time, mean time to maximum braking and mean maximum braking value with the warning in scenario S_3W_1 and S_5W_1 and the acute warning in scenario S_6W_2 for learning phase L3, displaying the number of participants (N) included at the bottom.

Discussion

This paper presents a driving simulator study on how drivers react upon a multi stage warning system (comprising two warning stages: *warning* and *acute warning* presented in the HUD) for collision avoidance in urban areas. The main question was whether drivers understood the difference between the two warning stages and if they profited from increasing experience with the warning system in different critical situations.

Almost all drivers reacted adequately to both the *warning* and the *acute warning*, when experiencing different safety-critical situations for the first time, which shows they are intuitively understandable. Yet, there were still some indications of a learning effect. For example, in their first experience of two out of three scenarios with a *warning* some drivers only showed a brake reaction when the *warning* had escalated to the *acute warning*. In their second experience, most drivers reacted already to the *warning*. Thus, it seems that some situations are harder to comprehend with drivers either learning what the different warning stages mean or adapting their driving behaviour so that the situations escalate slower, which allows them to react already to the *warning*, the first warning stage of the multi stage warning system. Similarly, somewhat de-escalating driver behaviour is found in the quite critical scenarios with a direct *acute warning*, although a significant learning effect from learning phase L2 to L3 cannot be supported statistically.

Moreover, drivers can clearly distinguish between the two warning stages, which becomes obvious in their driving behaviour. The maximum braking value in the *acute warning* is significantly greater than in the *warning*. Similarly, the brake reaction time and the time to reach the maximum braking value in the *acute warning* are significantly shorter as well. While this corresponds to the intended difference between the two warning stages, it may also be due to the situations being of clearly different criticalities. Thus, drivers might also react differently as they are aware of this difference in the situations. Probably, both factors contribute to the difference in the drivers' reactions. However, as was shown in previous studies with these quite critical situations there is also a benefit of the *acute warning* as compared to a control group without an *acute warning* (Kazazi et al., 2015; Winkler et al., 2015). Thus, it seems the concept and aim of the *acute warning* is well understood and supports the drivers to react adequately. Moreover, the drivers also seem to understand and react appropriately to the *warning*, as they thereupon brake less but in most cases sufficiently, in order to de-escalate the situation and prevent an *acute warning*. Consequently, the proposed multi stage warning system can be recommended as drivers show the intended driving behaviour.

To further undermine the benefit of the proposed multi stage warning system, more analyses based on the data recorded in the present study will follow in another paper. For example, the other scenarios and learning phases as well as subjective ratings of the warning system should also be considered. In order to see how well drivers apply their warning experience onto new situations, comparisons of the driving behaviour of unsupported (L1) and supported driving, at the beginning of the learning phase (L2) or at the end of all learning phases (L4), would be of interest. Likewise, the multi stage warning system should be investigated with drivers experiencing more stages succeedingly or when the system is actually integrated with other assistcance systems having different aims like driver comfort (e.g., lane change or constriction assistance) or ecological driving (e.g., traffic light assistance). Consequently, further research on how to integrate and prioritise different assistance systems is needed.

References

Ablaßmeier, M. Poitschke, T. Wallhoff, F. Bengler, K., & Rigoll, G. (2007). Eye Gaze Studies Comparing Head-up and Head-down Displays in Vehicles. In *IEEE Multimedia and Expo* (pp. 2250-2252).

Campbell, J.L., Richard, C.M., Brown, J.L., & McCallum, M. (2007). *Crash Warning System Interfaces: Human Factors Insights and Lessons Learned (DOT HS 810 697)*. Washington D.C., USA: National Highway Transportation Safety Administration.

Diederichs, F., Marberger, C., & Hinder, V. (2010). Iterative design and assessment of an audio visual warning concept for Car2x communication systems. In *Proceedings of the ITS World Congress on Intelligent Transportation Systems*. Busan, Korea.

Dingus, T.A., Jahns, S.K., Horowitz, A.D., & Knipling, R. (1998). Human factors design issues for crash avoidance systems. In N.V. Barfield, & A. Dingus (Eds.), *Human factors in intelligent transportation systems* (pp. 55-94). Mahwah, NJ, USA: Erlbaum.

Engel, S., Kratzsch, C., & David, K. (2013). Car2Pedestrian-Communication: Protection of Vulnerable Road Users Using Smartphones. In J. Fischer-Wolfarth, and G. Meyer (Eds.), *Advanced Microsystems for Automotive Applications 2013 – Smart Systems for Safe and Green Vehicles* (pp. 31-41). Cham, Germany: Springer.

General Motors Corporation & Delphi-Delco Electronic Systems (2002). *Automotive Collision Avoidance Operational Field Test Warning Cue Implementations Summary Report* (DOT HS 809 462). Washington D.C., USA: National Highway Transportation Safety Administration.

Gläser, C., Michalke, T., Bürkle, L., & Niewels, F. (2014). Environment Perception for Inner-City Driver Assistance and Highly-Automated Driving. In *Proceedings of the IEEE Intelligent Vehicles Symposium* (pp. 1270-1275). Dearborn, Michigan, USA.

Heimes, F., & Nagel, H.-H. (2002). Towards Active Machine-Vision-Based Driver Assistance for Urban Areas. *International Journal of Computer Vision, 50*(1), 5–34.

Herrmann, S. & Schroven, F. (2012). Situation Analysis for Driver Assistance Systems at Urban Intersections. In *Proceedings IEEE International Conference on Vehicular Electronics and Safety* (pp. 151-156). Istanbul, Turkey.

Heller, O. (1982). Theorie und Praxis des Verfahrens der Kategorienunterteilung (KU). In O. Heller (Ed.), *Forschungsbericht 1981* (pp. 1-15). Würzburg, Germany: Psychologisches Institut der Universität Würzburg.

Jones, E.M., & Hansman, R.J. (2007). Approaches to Enhance Driver Situational Assessment Aids (ICAT-2007-3). Available from http://hdl.handle.net/1721.1/38869 (18.08.2015).

Kazazi, J. Winkler, S., & Vollrath, M. (2015). Accident prevention through visual warnings. How to design warnings in head-up display for older and younger drivers. In *Proceedings of the IEEE International Conference on Intelligent Transportation Systems* (pp. 1028-1034).

Khanafer, A., Balzer, D., & Isermann, R. (2009). A rule-based Collision Avoidance System - A Scene Interpretation, Strategy Selection, Path Planning and System Intervention. *SAE International Journal of Passenger Cars – Mechanical Sys*tems, *2*(1), 389-397.

Krüger, H.-P., Grein, M., Kaussner, A., & Mark, C. (2005). SILAB – A task-oriented driving simulation. In *Proceedings of the Driving Simulator Conference* (pp. 232-331). Iowa City, IA, USA.

Lee, J.D., McGehee, D.V., Brown, T.L., & Reyes, M.L. (2002). Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors*, *44*, 314-334.

Lerner, N., Robinson, E., Singer, J., Jenness, J., Huey, R., Baldwin, C., & Fitch, G. (2014). Human factors for connected vehicles: Effective warning interface research findings. (Report No. DOT HS 812 068). Washington, DC: National Highway Traffic Safety Administration.

Manstetten, D., Bengler, K., Busch, F., Färber, B., Lehsing, C., Neukum, A., Petermann- Stock, I., & Schendzielorz, T. (2013). "UR:BAN MV Human Factors in Traffic" – a German Research Project to Increase Safety in Urban Traffic. In *Proceedings of the ITS World Congress on Intelligent Transport Systems* (pp. 596-605). Tokyo, Japan.

Najm, W.G., Koziol, J.S., Tijerina, L., & Pierowicz, J.A. (1994). Comparative Assessment of Crash Causal Factors and IVHS Countermeasures. In *Proceedings of the IVHS America Annual Meeting, Volume 1* (pp. 412-421). Washington D.C., USA: IVHS America.

Naujoks, F., & Neukum, A. (2014). Timing of in-vehicle advisory warnings based on cooperative perception. In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting* (pp. 193-206). Available from http://hfes-europe.org (20.08.2015).

Nöcker, G., Hermann, D. & Hiller, A. (2000). Telematics Based Traffic Organization. In *Proceedings of the ITS World Congress on Intelligent Transportation Systems*. Turin, Italien.

Petermann-Stock, I. & Rhede, J (2013). Intelligente Strategien für nutzerzentrierte MMI Konzepte im urbanen Raum. In VDI Gesellschaft Fahrzeug- und Verkehrstechnik (Eds.), *Der Fahrer im 21. Jahrhundert* (VDI-Berichte, No. 2205, pp. 263-286). Düsseldorf, Germany: VDI-Verlag.

Röglinger, S., & Facchi, C. (2009). How can car2x-communication improve road safety. University of Applied Sciences Ingolstadt, *Arbeitsberichte – Working Papers, vol. 15*.

Schmidt, G.J., Khanafer, A., and Balzer, D. (2009). Successive Categorization of Perceived Urgency in Dynamic Driving Situations (No. 2009-01-0780). In *Proceedings of the SAE 2009 World Congress*, Detroit, Michigan.

Singer, J., Lerner, N., & Robinson, E. (2014). Effects of Multiple-Device Integration Strategies on Driver Comprehension of Crash Warnings. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, no. 1* (pp. 1923-1927).

Statistisches Bundesamt (2014). Traffic. Traffic accidents 2013 [Verkehr. Verkehrsunfälle 2013]. Fachserie 8 Reihe 7. Wiesbaden, Germany: Statistisches Bundesamt.

Watanabe, H., Yoo, H., Tsimhoni, O., & Green, P. (1999). The Effect of HUD Warning Location on Driver Responses. In Proceedings of the *ITS World Congress on Intelligent Transport Systems*.

Weiß, C. (2011). V2X communication in Europe – From research projects towards standardization and field testing of vehicle communication technology. *Computer Networks, 55*, 3103 –3119.

Werneke, J., Kleen, A., & Vollrath, M. (2014). Perfect timing – Urgency, not driving situations influence the best timing to activate warnings. *Human Factors, 56*, 249-259.

Winkler, S., Kazazi, J., & Vollrath, M. (2015). Distractive or supportive – how warnings in the head-up display affect drivers' gaze and driving behaviour. In *Proceedings of the IEEE International Conference on Intelligent Transportation Systems* (pp. 1035-1040).

Winkler, S., Werneke, J., & Vollrath, M. (2016). Timing of early warning stages in a multi stage collision warning system: Drivers' evaluation depending on situational influences. *Transportation Research Part F: Traffic Psychology and Behaviour, 36*, 57-68.

Zarife, R. (2014). *Integrative Warning Concept for Multiple Driver Assistance Systems.* Ph.D. Thesis, Julius-Maximilians-Universität Würzburg, Würzburg, Germany.

Investigating the impact of attentional declines on road-crossing strategies of older pedestrians

Rebecca Wiczorek, Jan Siegmann, & Florian Breitinger
Technische Universität Berlin,
Germany

The main reason for older pedestrians for being involved in accidents is related to attentional deficits. The aim of the current research is to investigate the link between attention and road crossing strategies of older persons. Two group discussions were carried out to investigate what types of crossing locations are perceived as dangerous. Additionally, an observation interview was conducted including 102 pedestrians of three age groups at six different crossings. Based on these investigations a questionnaire was developed examining attention, behaviour and risk perception in road crossing. The questionnaire consists of eight different crossing situations. Participants have to estimate the risk and indicate their crossing strategies by choosing possible actions from a predefined set. Additionally, they answer twelve items with regard to their experience of attention-critical traffic situations. Seventy-eight participants filled in the questionnaire, half of them were over the age of 65 years, the other half younger than 35 years. It turned out that road crossing behaviour varies with age as well as with the type and the size of the road. In contrast, estimation of risk does not depend on age. Surprisingly, reported attention deficits turned out to be more present in younger than in older participants. Potential reasons for the findings and the relevance of different behavioural strategies for traffic safety are discussed.

Introduction

Everyday pedestrian mobility is a crucial need for older persons' ability to live independently. With an increase in age, more and more quotidian activities are carried out by foot (e.g. Limbourg & Matern, 2009). In contrast, older pedestrians are at high risk in traffic. In Germany in 2012, more than half of the pedestrians dying after a crash were older than 65 years (Statistisches Bundesamt, 2013). While they are not involved more often in crashes than younger adults, their risk of being injured in relation to their average walking distance is higher (Rytz, 2006). A lack of paying attention to traffic was identified as the main reason why older pedestrians are involved in crashes with motorists (Statistisches Bundesamt, 2013).

The aim of the research group FANS (Fußgänger-Assistenzsystem für ältere Nutzerinnen und Nutzer im Straßenverkehr - Pedestrian Assistance System for Older Road User) is to develop and evaluate an assistance system for older pedestrians supporting them in road crossing. The process of development is based on an

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

iterative user-centred approach. Therefore, it is necessary to investigate the behaviour of older pedestrians and identify crucial factors. Crossing strategies may change as a function of increasing age-related deficits. Thus, the system should not prevent people from using their own coping strategies, but provide assistance in critical situations.

The aim of the studies presented here was to develop a questionnaire that links road crossing behaviour to attentional deficits of older people. Past studies with regard to road crossing behaviour have investigated aspects such as gap acceptance, misperception of distances, ability of behavioural adaptation, etc. (cf. Papadimitriou, Yannis & Golias, 2009). Most of these studies focused on the miscalibration of behaviour and situation. Others investigated the impact of declines in motor and cognitive abilities (including attention) on road-crossing decisions in the ongoing traffic (e.g. Dommès & Cavallo, 2011). In contrast, the current approach aims to understand the reasons for overlooking potential dangerous road users. Research shows that several different attention abilities decline with an increase in age. Therefore, problems in crossing roads may be attributed to those age-related attentional declines (cf. Ball et al., 1990) as some of them can be linked to traffic requirements. Examples are difficulties in the estimation of the direction of a moving object, problems to distinguish between relevant and irrelevant stimuli, etc. Additionally, older persons’ performance in visual search tasks decreases more strongly compared to younger people when a parallel visual task is carried out (e.g. Ziefle et al., 2008; Musselwhite & Haddad, 2010). The same seems to be true for additional motor tasks even though they require different cognitive resources (e.g. Beurskens & Bock, 2012). The questionnaire should provide information with regard to attentional deficits in traffic and to the visual search strategies applied by younger and older pedestrians.

In order to develop and to use such a questionnaire, some open questions had to be answered in advance. The first question refers to the traffic environment from the perspective of older pedestrians. Two group discussions were carried out in order to learn more about everyday mobility of older pedestrians and to identify safety-critical aspects of traffic and environmental conditions. Second, as a precondition for administering a questionnaire study, it had to be investigated whether people are aware of their own road crossing strategies (i.e. able to report them). To answer this question, pedestrians of different age groups were observed while crossing a road and asked to describe their behaviour afterwards. This procedure should allow controlling for the accuracy of the described behaviour, to identify systematic age-related differences in behaviour or the description of it, and to control for reporting normative instead of realistic behaviour. Procedures and results of group discussions and observation-interview are used for the development and conduction of the questionnaire study.

Group discussions

The theoretical framework for the group discussions is based on the combination of Hägerstrand's (1970) time geography and Giddens' (1997) theory of structuration. Both theories are located within the social sciences and mainly used in the fields of human geography and sociology. From Hägerstrand's (1970) point of view, time and space restrict freedom of action. He defines three kinds of constraints limiting the spatial range of human actions: *capability constraints* describe biological needs such as eating and sleeping, *coupling constraints* refer to the environmental context of every action, and *authority constraints* subsume general rules or laws. However, constraints are exclusively restricting human actions. Thus, Giddens (1997) theory of structuration was integrated, because it includes also enabling factors that support peoples' actions. The combination of both theories offers a more holistic view on the everyday mobility of older pedestrians.

Method

Two group discussions were carried out. Focal points were older persons' motivation to participate in traffic by foot, factors that impair or support their behaviour, and the resulting consequences for everyday mobility. However, only results with regard to road crossing will be reported here.

The group discussions were conducted based on the methods suggested by Zwick & Schröter (2012). Criteria based selection was done according to previous findings (e.g. Limbourg & Matern 2009). They identified gender, age, walking ability, and place of residence as relevant criteria for everyday mobility of older pedestrians. Both groups differed with regard to number and expertise of participants. The aim of the first group was to benefit from the aggregated view of representatives of organisations dealing with the concerns of pedestrians or older persons. The focus of the second group was on the individual experiences of older pedestrians in the city of Berlin.

Participants

Four representatives (all of them older than 55, half of them female) took part in the discussion of the first group. The participants of the second group were eight senior citizens between the ages of 60 and 81(M=69; SD=7.21) with experience in walking through the city of Berlin on a regular basis. Half of them were female, two were using a walking aid and they lived in districts, which differed with regard to accident frequencies.

Discussion-guidelines

Two guides with questions were prepared as stimuli for the discussions. Specific questions for participants were different for the two groups, but the main subjects were the same: (1) importance of walking, (2) traffic safety, and (3) actual behaviour. Each subject contained several main questions with additional sub questions to ensure a continuous discussion without influencing its direction too much.

Analysis of the discussions

Discussions were analysed using the qualitative content analysis as stated by Gläser and Laudel (2010). A first category system was generated based on the theoretical considerations (Hägerstrand, 1970; Giddens, 1997). It was then iteratively complemented and adapted based on the stepwise analysis of the discussion transcripts. The complete category system includes motives for actions, actions themselves, and factors relevant for action (impeding and enabling). The latter category consists of person-related and environment-related factors. The environment-related factors are further subdivided into other individuals, social regulations, ambient conditions, and infrastructure. Relevant for the current considerations is only the latter category: infrastructure.

Results and discussion regarding the impact of infrastructure on road crossing

Statements, relevant for road crossing behaviour of older pedestrians, subsumed under the category infrastructure can be distinguished in (1) type of street, (2) pedestrian crossings, (3) condition of the road surface, and (4) obstructions:

(1) Type of street refers to the number of lanes (to be crossed) and traffic density. Uncertainty and even fear are growing with the number of lanes and the traffic load. In complex and confusing situations, older pedestrians do not know when to cross the street and often hesitate, which may result in risky behaviour due to late decisions.

(2) Official pedestrian crossings are sometimes inconveniently located forcing people to do detours. Most pedestrians are not willing and some are not even able to choose longer ways, which is especially true for those with walking disabilities. This may lead to taking unsafe shortcuts.

Pedestrian crossings with traffic lights are very suitable for crossing roads with two lanes or more. However, older pedestrians experience the green phases as being too short. Therefore, they hurry in order to reach the other side or centre island in time. This can compromise their attention to traffic. This problem has already been identifies in other countries such as Ireland and the UK (Romero-Ortuno et al., 2010; Asher et al., 2012).

On pedestrian crossings with a centre island, older pedestrians tend to look out for traffic in only one direction. Neglecting the other direction may lead to overlooking road users such as cyclists who ride on the wrong side of the road.

(3) The condition of the road surface can be problematic for older pedestrians, especially those with an impaired walking ability. Pavement damages, high kerbstones, and cobblestones can impede the progress of older persons. Problematic surface conditions can lead to uncertainty and may reduce resources need to attend to traffic.

(4) Obstructions mostly refer to plantings at the roadside and the centre island. The view into traffic can be greatly impaired forcing pedestrians to step on the road to gain better sight or even to cross the street without looking for approaching cars.

Observation-interview

Method

Participants
One hundred and two pedestrians of three different age groups (i.e. 34 in each group), participated in this study. Younger group: ≤35 (19 male, 15 female), middle-aged group: 35≥65 (17 male, 17 females), older group: ≥65 (13 male, 21 female).

Procedure
Two persons conducted the observation-interview: an observer and an interviewer. Participants' behaviour was logged while crossing one of six different roads of similar size with one lane for both directions. Afterwards, questions about their previously shown behaviour were asked.

Observation-questionnaire
The observer focused on the same behavioural components as the interviewer: Gaze frequencies (how often participants looked for traffic) and whether a person stopped before crossing. To facilitate participant's responses, gaze frequencies were categorized into four groups: did not look at all, looked only in one direction, looked in both directions, and looked several times in both directions.

Dependent measures
Observed behaviour in terms of gaze frequencies and number of stoppings were analysed. *Observed and self-reported behaviour* were compared. *Incorrect estimations* were further analysed with regard to their tendency to describe normative behaviour.

Results

A one-way ANOVA was performed to analyse age differences in observed gaze frequencies. In addition 3x2 χ^2-tests of independence were used to check for age differences in observed stopping behaviour and in the accuracy of the self-assessments of stopping and looking. A 2x2 χ^2-test of independence was used to compare observed and reported stopping behaviour.

Observed Behaviour
No difference in gaze frequencies and number of stops was found between age groups. On average, participants looked for traffic 2.84 times and 27.4% of them stopped before crossing the road.

Relation between observed and self-estimated behaviour
Accuracy of estimation gaze frequencies did not differ between age groups. The same was true for estimation of number of stops. Overall, only 45.8% of the participants were able to estimate correctly how often they looked for traffic. 73.5% estimated correctly whether they had stopped before crossing.

Incorrect estimations

Significant differences were found between age groups with regard to their tendency to describe normative behaviour, $\chi^2(2,N=53)=6.144$, $p=.046$. Taking only into account participants with incorrect estimation, older participants overestimated in 84.6% of the cases how often they looked for traffic, while the younger and the middle-aged group underestimated their gaze frequencies in 63.6% and 61.1% of the cases. No age differences were found with regard to incorrect estimation of stops. However, participants had a tendency to report normative behaviour, $\chi^2(1,N=100)=4.233$, $p=.04$. Of those who did not stop, 30.6% stated that they did, while only 10.7% of participants who stopped reported they did not stop before crossing the road.

Road-crossing Questionnaire

The questionnaire was developed based on the results of the previous literature review, the group discussions, and the observation interview. Questions regarding attention abilities and attention deficits were derived from previous findings respecting different types of attention deficits of older persons (cf. Ball et al., 1990). The presented behavioural alternatives included both aspects of visual search and aspects of motion, as being potential relevant aspects of road crossing of older pedestrians (e.g. Musselwhite & Haddad, 2010; Beurskens & Bock, 2012). The stimulus material for the questionnaire was created based on results of the group discussions. It includes different aspects of infrastructure relevant for road crossing, such as number of lanes, architecture and surface of pedestrian crossings vs. absence of pedestrian crossings, problematic surface conditions, and obstacles. Findings of the observation interview served for decisions on sample and questions. Given that ageing is a continuous process, the younger group instead of the middle-aged group was chosen as reference in order to assure the absence of age-related declines in the control group. Observed behaviour in terms of gaze frequencies and stopping were not related to age, but seemed to depend more on the actual traffic conditions. One possible reason for the absence of age effects might be the similarity of the road design used for the observation. Therefore, participants were asked to provide descriptions of behaviour in different situations. Comparison of observed and reported behaviour revealed no effects of age, which is a precondition for a systematic comparison of answers in the questionnaire. As percentages of correct estimations were rather low, questions were revised and questions were asked about behaviour that is more general excluding details such as gaze *frequency*. Further analysis of incorrect estimations showed that older persons had a stronger tendency to report normative behaviours, while younger persons tended to the opposite. In order to prevent these biases or at least reduce them, instructions given to the participant emphasized the importance of honest responses. Additionally, a scale of risk estimation was included in order to understand whether potential differences in reported behaviour resulted from different strategies or from diverging perceptions of the situations. It was hypothesised that differences in behaviour would occur based on the type and the size of the road. It was expected that certain conditions such as uneven surfaces, obstacles and traffic lights would lead to different behaviours in younger and older participants. Older participants were expected to report higher estimations of risk than younger participants. Furthermore, it was expected that older participants would experience more often situations were attentional deficits become manifest than younger participants.

Method

Participants
Seventy-eight participants of two age groups attended the study. The younger group (≤35) consisted of 39 students (20 male, 19 female); their age ranged from 18 to 34 years (*M*=26.3; *SD*=3.7). The older group (≥65) also consisted of 39 participants (16 male and 23 female), ranged in age from 64 to 83 years (*M*=72.6; *SD*=4.7), a participant younger than 65 was included because recruiting was based on the year of birth). Further characteristics of younger and older participants are, respectively: driver licence (27 vs. 28), regular drivers (10 vs. 19), regular cyclists (28 vs. 20), walked regularly (39 vs. 39), had experienced a fall on the pavement outside (11 vs. 15), and were involved in an accident as pedestrians (4 vs. 3).

Description of the road-crossing Questionnaire
Photographs of the same person in eight different road crossing situations were shown to participants (see Figure 1). Situations differed systematically with regard to crossing type (4 with and 4 without pedestrian crossings) and road size (from small to large in four steps, ranging from one lane for two directions to two lanes for each direction). Additional variations were: height of kerbstone (1 high, 7 low), surface damage (5 with damages, 3 without), line-of-sight obstruction (5 with free sight, 3 with obstacles), centre islands (3 of the pedestrian crossings), and traffic lights (2 of the pedestrian crossings). Photographs were presented in a random order. For each photograph, participants were asked to indicate their estimation of risk in the specific situation. Furthermore, for each crossing situation, they had to indicate, which of the ten predefined actions they would show and in which order. The actions were either related to visual search, relevant for save crossing (check for traffic coming from the left side or coming from the right side), related to additional visual search with regard to safe walking (control for asperity on the pavement or on the road, pay special attention to the kerbstone), or related to motions (accelerate; slow down; stop; continue; step onto the road).

The questionnaire included an additional part assessing potential attention deficits in traffic situations. It consisted of 12 items. Each item described a different attention-critical traffic situation related to one of the known attentional deficits occurring with increasing age. Participants had to indicate how often they experience such situations. For example: 'When a car is far away, I have difficulties deciding whether it moves towards me.' This statement referred to the finding that older people, on average, have greater difficulties detecting the direction of movements than younger people.

Figure 1. *Three of the eight photographs used in the questionnaire with and without pedestrian crossings.*

Procedure

Participants were tested in groups of up to six persons. After filling in the demographical part of the questionnaire, participants received detailed written instructions. They were instructed to choose any number of actions they would perform to cross the road from the predefined set, and to sort them into the order in which they would be executed. “Road crossing” was finished when reaching the other side of the road or, in case of a centre island, when reaching this point. After having described their behaviour, participants estimated the risk of the current situation. Finally, participants filled in the part of the questionnaire related to attention. After finishing the study, they were thanked for their participation and received financial compensation.

Design and Dependent Measures

The behavioural part consisted of a 2 (age) x 2 (road type) x 4 (road size) design with repeated measures on the second and third factor. For the actions related to visual search for safe walking, roads were grouped according to surface damage and high kerbstones (5 with and 3 without), resulting in a 2 (age) x 2 (surface damage) design. Attention values were compared only between age groups.

Frequencies of the single actions served as behavioural measure. The variables ‘accelerate’ and ‘slow down’ were adjusted before the analysis. Instances when participants chose to slow down before accelerating were not taken into account for ‘accelerate’. Instances when participants chose to stop after slowing down were not taken into account for ‘slow down’.

Risk estimation was assessed on a 5-point Likert scale ranging from 1 = very low to 5 = very high.

Attention was assessed on a 5-point Likert scale ranging from 1= very rarely to 5 = very often. Answers of all 12 items were summed up to an overall attention score.

Results

Behavioural data and risk estimation were analysed with 2x2x4 ANOVAs for repeated measures, actions related to surface control were analysed with a 2x2 ANOVA with repeated measures. *A priori* defined single t-test comparisons with regard to surface damage and line-of sight obstruction were calculated for risk estimation. Attention was analysed with a t-test for independent samples. Additional 2x2x4 and 2x2x2x4 ANOVAs with gender as a control variable were calculated.

Behavioural Data

A list of all main effects and all single interaction effects with regard to visual search for safe crossing and motions are displayed in Table 1. Three-way interactions were not part of the hypothesis and are therefore not reported.

Table 1. F-values of 2x2x4 ANOVAs of main effects (with *df*(1,76)) and interaction effects (with df (3,228)) of single actions with regard to visual search for safe crossing and to motions (* $p< 0.05$, ** $p<0.01$, *** $p<0.001$)

	Age Group (A)	*Road type (T)*	*Road size (S)*	*AxT*	*AxS*	*TxS*
Stop	13.84***	62.71***	17.09***	1.921	.49	18.63***
Slow down (corrected)	9.50**	<1	8.06***	<1	<1	11.18***
Accelerate (corrected)	<1	76.86***	23.58***	3.39	1.15	22.31***
Check right side	<1	23.85***	30.00***	<1	<1	37.31***
Check left side	<1	27.58***	6.07**	<1	<1	7.48***

As can be seen in Figure 2, no effect of age group was found with regard to visual search for safe crossing. Percentages of persons checking the left and the right side were the same in both groups. All participants tended to look to both sides in situations without pedestrian crossings. With pedestrian crossings, they chose more often to look only to the left side. That was especially true when roads had a centre island and thus, cars were coming only from the left. Few participants decided not to look for traffic at all, mainly on roads with traffic lights. No interaction effects of age and road type or road size were found.

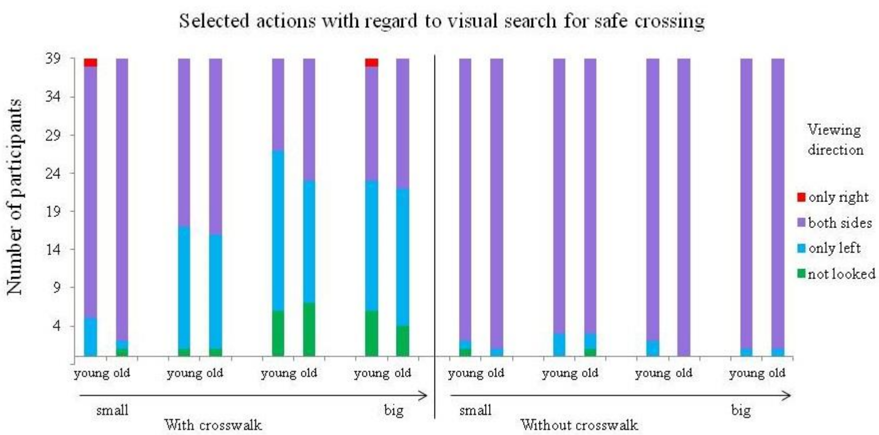


Figure 2. Selected actions with regard to visual search for safe crossing with and without pedestrian crossings. Actions and combination of actions are categorized into four groups. Categories are mutually exclusive.

As shown in Figure 3, more older persons said they stopped before crossing the road, while more younger persons chose to slow down instead. No effect of age group was found for accelerating. Both age groups accelerated most often on the two large roads with pedestrian crossings that had green traffic lights. Both groups stopped less often on traffic lights, thus they were found to stop more often on smaller roads with pedestrian crossings, while stopping more often on larger roads without pedestrian crossings. Similarly, both groups did not slow down often on traffic lights. Apart from that, they showed a tendency to slow down more often instead of stopping on smaller roads compared to larger roads. No interaction of type or size of road with age was found.

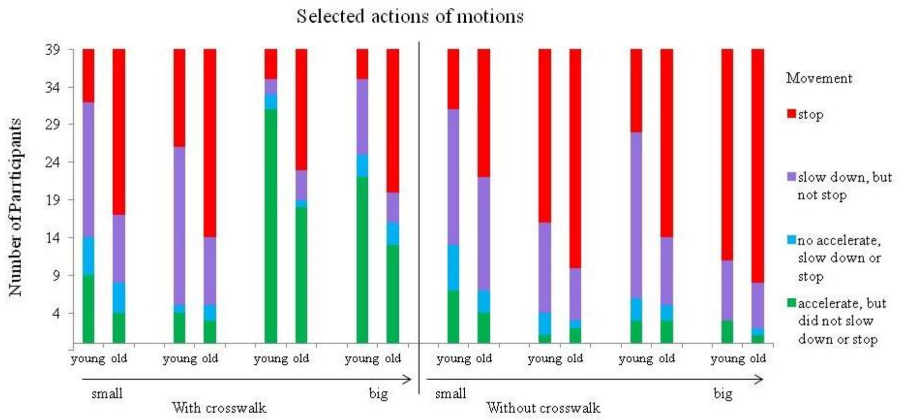


Figure 3. Selected actions of motions for single roads with and without crosswalk. Actions and combinations of actions are categorized into four groups. Categories are mutually exclusive.

Main effects and interaction effects with regard to visual search for safe walking are presented in Table 2.

Table 2. F-values of 2x2 ANOVAs of main and interaction effects (with *df*(1,76)) of single actions with regard to safe walking (* *p*< 0.05, ** *p*<0.01, ****p*<0.001)

	Age group	*Underground (U)*	*AxU*
Check pavement	13.37***	13.39***	6.71*
Check street	9.06**	15.67***	.05
Pay attention kerbstone	9.24**	3.60	.10

In Figure 4 means of different surface checking actions are displayed. Significant age differences were found for all three checking actions. More of the older participants checked the pavement, the road, and the kerbstone compared to the younger participants. Both groups checked the road more often when its surface was uneven. This difference was more pronounced for the older than the younger. Checking of kerbstones did not vary between streets with even and uneven surfaces. However, only one of the streets had comparatively high kerbstone.

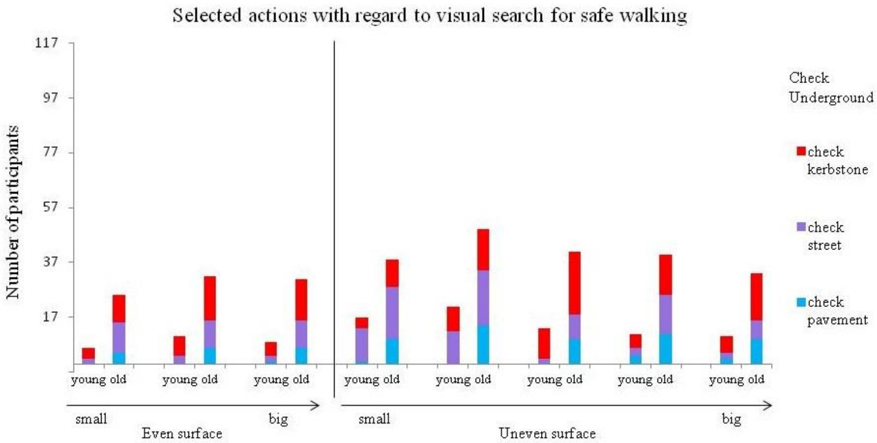


Figure 4. Selected actions with regard to visual search for safe walking for single roads with and without uneven surface. Participants were able to select any number of the three variables.

Risk Estimation

Younger and older adults did not differ with regard to their risk estimation. Roads without official crossing were considered riskier, $F(1,76)=30.29$, $p<.001$, $\eta^2=.285$, as well as larger roads, $F(1,76)=46.591$, $p<.001$, $\eta^2=.380$. Roads with uneven surface were considered riskier than roads with even surface, $t(77)=3.42$, $p<.001$. Roads without blocked sight were considered riskier than roads with blocked sight, $t(77)=-7.22$, $p<.001$.

Attention

Younger participants reported to experience more attention-critical situations ($M=24.87$) than older participants ($M=22.49$). This effect was marginally significant $t(76)=1.903$, $p=.061$.

Gender as control variable

In order to control for gender effects, ANOVAs including gender as additional between-subjects factor were calculated for all dependent variables. No effect of gender was found for risk estimation, attention and behaviour with one exception. Women were found to stop more often before crossing than men, $F(1,74)=6.120$, $p<.05$, $\eta^2=.076$.

Discussion

The questionnaire study was conducted to investigate problems related to attentional demands in road crossing and strategies of older pedestrians in comparison to younger pedestrians. Results revealed a methodological weakness of the attention part of the questionnaire. Analyses of the mean frequency of experiencing critical situations in traffic due to potential attentional declines indicated more attention-related problems for younger than for older participants. This finding is not in line with the broad knowledge regarding the decrease of attentional functions in the later lifespan. Possible reasons for this unexpected finding either may be attributed to the assessment method or may be due to older pedestrians' general coping strategies. It is possible that older persons learn to avoid attention-critical situation. Research in the driving context show that older persons avoid rush hours, bad weather conditions, complex intersections, unprotected left turns, etc. (McGwin & Brown, 1999). Same coping mechanism may be applied when walking in order to compensate for age-related difficulties. In future studies a different method will be used, which assesses attention abilities more directly in order to avoid ambiguous results.

However, results regarding behavioural strategies offer interesting new insights. The analyses of actions chosen by participants in order to cross roads of different type and size show differences between the age groups as well as for different traffic environment situations. Problematic strategies as well as coping strategies of older pedestrians were identified. It is important to note that differences in behaviour of the two age groups did not derive from differences in perception of the respective situation, because estimated risk did not differ between groups.

In line with previous findings of the observation interview, people of different age groups behaved similar with regard to safety-related visual search. Younger and older participants looked for traffic with the same frequencies. Not checking both directions when only crossing until a centre island was first identified during the group discussions. Results of the questionnaires confirmed this visual search behaviour. Both age groups showed this particular behaviour.

Regarding visual search for safe walking, a clear difference emerged between the age groups. Older participants checked the pavement more often than younger and both checked more often when the surface was damaged. This is a clear indication for problematic strategies. Participants use resources to check the pavement that are actually needed to observe the traffic situation.

Results for actions related to motions show that younger people tend to do both tasks: visual search and walking in parallel. More often than the older, they chose only to reduce their walking speed instead of stopping before checking traffic. Older pedestrians stop more often first and then look for traffic. This sequential task completion represents an adequate coping strategy for declines in the multitask capacities. Both groups stopped more often on larger than smaller roads, without pedestrian crossings. Pedestrian crossings with green traffic lights led both groups to accelerate their walking velocity. This finding confirms earlier results of the group discussions. It is especially problematic for older persons because of their reduced ability to check for traffic parallel to their walking direction. Albeit pedestrians 'right' to cross with green traffic lights, they are nonetheless in risk to be run over by a turning car or cyclist violating the traffic rules.

Results of this pedestrian study are in line with previous findings regarding divided attention and multitasking in older adults (Ball et al., 1990; Musselwhite & Haddad, 2010; Beurskens & Bock, 2012). They indicate that support of an assistance system should primarily focus on two things: the avoidance of multitask requirements while crossing a road and the older persons' probability of overlooking potential dangerous road users not behaving in the anticipated way. As a next step assistance strategies will be developed meeting these criteria and will then be evaluated in a laboratory and field setting.

Acknowledgement

We thank Mandy Dotzauer for proofreading. Furthermore, thanks are due to Stefanie Boxhorn for acting as pedestrian model.

References

Asher, L., Aresu, M., Falaschetti, E., & Mindell, J. (2012). Most older pedestrians are unable to cross the road in time: a cross-sectional study. *Age and Ageing, 41*, 690–694.

Ball, K.K., Roenker, D.L., & Bruni, J.R. (1990). Developmental changes in attention and visual search throughout adulthood. *Advances in psychology*, *69*, 489-508.

Beurskens, R., & Bock, O. (2012). Age-related deficits of dual-task walking: a review. *Neural plasticity*, 2012.

Dommès, A., & Cavallo, V. (2011). The role of perceptual, cognitive, and motor abilities in street-crossing decisions of young and older pedestrians. *Ophthalmic and physiological optics, 31*, 292-301.

Giddens, A. (1997). *Die Konstitution der Gesellschaft* [The constitution of society]. Frankfurt/New York: Campus Verlag.

Gläser, J. & Laudel, G. (2010). *Experteninterviews und qualitative Inhaltsanalyse* [Expert interviews and qualitative content analysis]. Wiesbaden: VS Verlag für Sozialwissenschaften.

Hägerstrand, T. (1993). What about people in regional science. *Papers of the Regional Science Association*, *24*, 7-21.

Limbourg, M. & Matern, S. (2009). *Erleben, Verhalten und Sicherheit älterer Menschen im Straßenverkehr* [Experience, behaviour and safety of older people in traffic]. Köln: Eugen-Otto-Butz-Stiftung..

McGwin Jr, G., & Brown, D. B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis & Prevention*, *31*, 181-198.

Musselwhite, C., Haddad, H. (2010). Exploring older drivers' perceptions of driving. *European Journal of Ageing*, *7*, 181-188.

Papadimitriou, E., Yannis, G., & Golias, J. (2009). A critical assessment of pedestrian behaviour models. *Transportation research part F: traffic psychology and behaviour, 12*, 242-255.

Romero-Ortuno, R, Cogan, L., Cunningham, C.U., & Kenny, R.A. (2010). Do older pedestrians have enough time to cross roads in Dublin? A critique of the Traffic Management Guidelines based on clinical research findings. *Age and Ageing, 39*, 80–86.

Rytz, M. (2006). *Senioren und Verkehrssicherheit* [Seniors and traffic safety]. Bern: VCS Verkehrs-Club der Schweiz.

Statistisches Bundesamt (2013). *Unfallentwicklung auf Deutschen Strassen 2012* [Accident development on German roads in 2012]. Begleitmaterial zur Pressekonferenz am 10.Juli 2012 in Berlin.https://www.destatis.de/DE/Publikationen/.

Ziefle M., Pappachan P., Jakobs E., Wallentowitz H. (2008). Visual and auditory interfaces of advanced driver assistant systems for older drivers'. *In Proceedings of Computers Helping People with Special Needs: 11th International Conference* (pp. 62-69). ICCHP, Linz, Austria.

Zwick, M.M. & Schröter, R. (2012). *Konzeption und Durchführung von Fokusgruppen am Beispiel des BMBF-Projekts "Übergewicht und Adipositas bei Kindern, Jugendlichen und jungen Erwachsenen als systemisches Risiko"* [Conception and conduction of focus groups described at the example of the BMBF-project „Overweight and obesity in children, adolescents and young adults as a systemic risk"]. Wiesbaden: VS Verlag für Sozialwissenschaften.

Ambient light based interaction concept for an integrative driver assistance system – a driving simulator study

Marc Dziennus, Johann Kelsch, & Anna Schieben,
German Aerospace Centre
Braunschweig
Germany

For today's vehicles several advanced driver assistance systems are on the market supporting the driver in critical driving situations or automating parts of the driving tasks. In the future there will be even more. Currently, those assistance systems do not use a common and consistent interaction strategy to communicate with the driver. The goal of the present study is to present and to evaluate a concept using ambient light for presenting information of different assistance systems in an integrated way. Research on visual peripheral warnings showed positive effects on driver reaction times in demanding situations. This paper presents results of a driving simulator experiment, in which an ambient light concept using peripheral visual perception were tested. A 360° LED stripe was installed around the driver in a fixed-based driving simulator providing interaction signals via peripheral vision. The developed ambient light display should support the driver in different driving situations by a consistent colour-coded interaction design. In a between-group design 41 participants (21 with and 20 without ambient light) drove eight different highway scenarios to test the display. Results of the ambient light interaction design regarding driver reactions and subjective evaluation regarding comprehensibility of the ambient light concept are reported and discussed.

Introduction

More and more of today's car manufactures develop assistance functions with one main goal: automated driving. Numerous driver assistance systems are already on the market. While some driver assistance systems support drivers in critical driving situations only (Scott & Gray, 2008, Lee et al., 2004), others are able to take over whole driving tasks, such as longitudinal or lateral control (Adaptive cruise control (ACC), lane keeping assist (LKAS)). In specific situations both lateral and longitudinal control can be delegated to the automation (traffic jam assistant). Ironically, the integration of the driver becomes a crucial point on the way to automated driving (Bainbridge, 1983) as the vehicle automation might not be able to handle all driving situations under all conditions when introduced into the market. Thus, different automation levels (SAE international, 2014) will be available in one and the same car requiring transitions of control between the driver and the

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

automation. Essential for achieving a good integration of the driver is a good driver-automation interaction design to allow the driver to build up a correct mental model of the overall system and to choose and perform the correct actions if necessary.

For a longer period of time some automated driving functions, e.g. for highway driving, might exist in parallel to various other driving assistance systems that support the driver only in specific driving tasks and situations until automation is available for most of the driving environments. Often, these assistance systems are not designed as an integrative driving assistance system which results in different and inconsistent ways to communicate with the driver. The usage of different modalities, symbols or positions for information could lead to a confusion of the driver (Tretten & Gärling, 2011) which results in difficulties to interpret the information and warnings correctly (Cummings et al., 2007).

The goal of this study is to present and to evaluate an interaction concept for an integrative driver assistance system, which combines information from different assistance systems and present them consistently and understandably to the driver (Utesch, 2014; Maier, 2014, Zarife, 2014). To ensure that presented information are perceived and understood by the driver the right communication modality is crucial. In the present study an interaction concept is used which is based on peripheral vision. While driving, the visual modality is one of the most important but also most demanded modality (Sivak, 1996). This may be true for foveal vision, but peripheral vision could offer a new way to communicate information without overloading drivers. Existing literature on peripheral vision for driver assistance systems show positive effects regarding reaction time in demanding situations (Henning et al., 2008; Utesch, 2014; Maier, 2014; Laquai et al. 2011; Kienast et al., 2008). That is why some researchers develop and evaluate applications of peripheral information systems using different kinds of light, colours, animations and locations in the vehicle (Löcken et al., 2015; Pfromm, et al., 2013, Kienast et al., 2008, Laquai et al, 2011, continental, 2013). If peripheral vision is used to communicate information, some physiological characteristics need to be considered. Humans are not able to see sharp in the periphery. This means that text messages or symbols would not be suitable for this for peripheral vision. On the other hand, humans are very sensitive for light, movements and stimuli changes in their peripheral field of view (Remington et al., 1992). Following this, the presented integrative interaction design uses colour-coded lights and animations to deliver peripheral information.

While existing concepts use the ambient light only to communicate specific information like warnings in critical situations (Utesch, 2014, Maier, 2014; Pfromm et al. 2013, Laquai et al 2011) or lane change information (Löcken, 2015) the approach used in the present study has the goal to present different kinds of information from different assistance systems in an integrated way. Following this approach our colour-coded interaction design is based on the theory of ecological information by Gibbson (1966). Regarding Gibbson's theory the perception of the environment and its affordances are a critical factor for the behaviour. Kelsch and Dziennus (2015) described how to create affordances by the colour-coded design of the ambient light. Following this approach the present study investigates whether peripheral ambient light signals could stimulate desired driver behaviour compared to a baseline without any assistance.

Method

Experimental setup and participants

The ambient light
An ambient light concept using a 360° LED stripe to communicate interaction signals via peripheral vision was installed and tested in a fixed-based driving simulator (see Figure 1a). The ambient light surrounded the driver. Following this, the ambient display was divided into four different segments: front, rear, left and right (see Figure 1b).

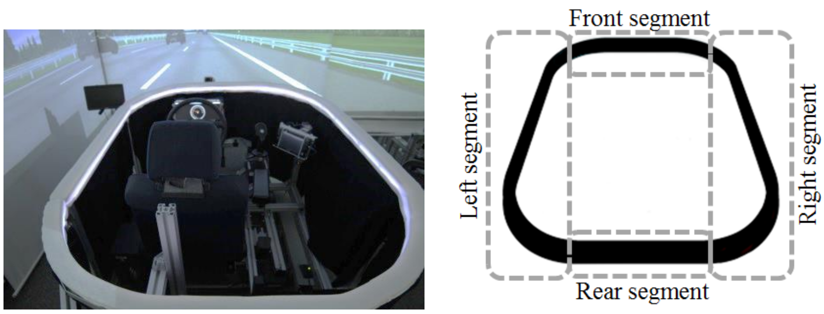


Figure 6. a) Ambient display surrounding the driver, b) Segments of the ambient display

Signals of the ambient light were displayed in the particular segment to create affordances for the drivers to initiate (recommendations) or to stop (warnings) driving manoeuvres. To achieve this goal, a colour-coded interaction design was used. According to that the colour white was used to illustrate that the ambient display was active (Figure 2a). To indicate that a driver should stop a driving manoeuver the particular segment of the ambient light turned red and gave a directional warning (e.g. in case of a longitudinal warning the front segment turned red, Figure 2b). With rising criticality of the situation the ambient light changed from only red (Time headway< 2s) into a blinking red (THW< 1s). Furthermore it was possible to communicate recommendations for driving manoeuvre to the driver. In this case the ambient light changed the colour of the segment into green (Figure 2c). Making recommendations more salient to the driver the ambient light started blinking (for 4s) if the driver did not react to a recommendation within four seconds. Affordances created by the ambient light should support drivers to perform the correct reaction (e.g. brake reaction or lane change). The ambient light design used for this study was a reduced design which included no head-down display or auditory feedback.

Figure 7. a) Ambient light active, b) Longitudinal warning, c) Lane change recommendation

Simulator
The experiment was conducted in a fixed-based simulator in the Interaction Design and Ergonomics lab (IDeELab) at the DLR Braunschweig. The simulator was a single fixed simulator equipped with an active steering wheel and pedals.

Scenarios
To test the impact of the ambient light on driver behaviour created eight different scenarios were created. Even though the warning or recommendation scenarios were different, the interaction design for warnings and recommendations was identical. The ambient display should provide generic information to guide the attention to a specific direction. The scenarios were divided into four warnings and four recommendation scenarios (see Table 1). Every participant drove on the same curved three-lane highway and experienced identical scenarios in the same order. The scenarios took between 30 and 80 seconds and the vehicle was stopped automatically to allow the investigator to switch between the scenarios via software control. While participants in the experimental group were supported by the ambient light, participants of the baseline experienced no support at all.

Table 6. Number of used scenarios

	Warning	*Recommendation*
Longitudinal	2	2
Lateral	2	2

Longitudinal warnings
In the first longitudinal recommendation scenario participants drove with 100 km/h on the right lane. Participants approached a traffic sign which restricted the velocity to 80km/h. To warn the driver that he was too fast the ambient display changed the colour of the front segment into red. In the second scenario participants drove with 100km/h on the right lane. Suddenly the front vehicle started braking at a time-headway of 1.5s. In the ambient light condition the driver was supported by a red pulsing (until THW>1s) or a red blinking (THW<1s) LED segment in the front.

Lateral warnings
In the first lateral warning scenario participants started with 100km/h on the right lane. After ten seconds they received an auditory command to change the lane. In parallel two vehicles on the middle lane crashed into each other and made a lane change to the middle lane very dangerous. Again the left segment of the ambient display turned red to signalize that a lane change was not safe anymore. If participants achieved a lateral deviation of more than 0.6m or used the indicator, the ambient light started blinking in red to underline that a lane change is not safe at all.

The second lateral warning scenario started on the right lane of motorway. Participants drove with 100km/h and experienced upcoming fog after ten seconds (see Figure 3). After 18s participants got an auditory command to change to the middle lane. At the same time the ambient light turned the left segment red to indicate that a lane change is not safe anymore. If participants achieved a lateral deviation of more than 0.6m or used the indicator, the ambient light started blinking in red to underline that a lane change is not safe. Six seconds after the auditory command or four seconds after the lateral deviation a truck appeared on the middle lane and took over with a velocity of more than 180km/h. If participants changed the lane, they would collide with the truck.

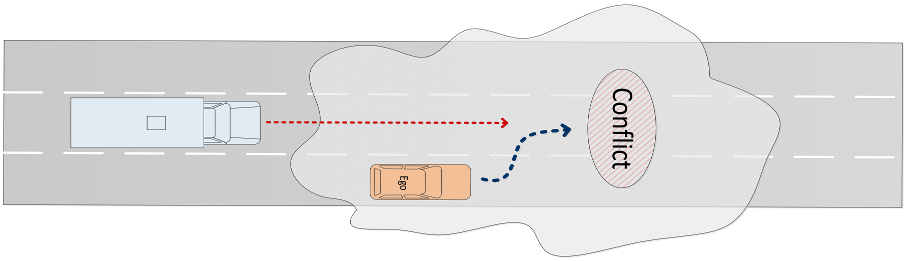


Figure 8. Lateral warning scenario

Longitudinal recommendations
In the first longitudinal recommendation scenario participants drove in a traffic jam with a speed of 20km/h. When participants reached the end of the traffic jam, surrounding traffic started to accelerate and the front segment of the ambient light changed into green to signalize that participant should speed up. In the second scenario participants started with 80km/h on the right lane. After ten seconds they saw a traffic sign which indicated a speed limit of 120km/h. Again the frontal element of the ambient light changed its colour into green.

Lateral recommendations
In the first lateral recommendation scenario participants drove with 100km/h on the right lane of the motorway and approached a slower vehicle on their lane. The left segment of the ambient display turned into green to recommend a lane change. In the second lateral recommendation scenario, participants drove with 100km/h on the middle lane of the three lane highway with surrounding traffic on the middle and left lane. After ten seconds the ambient light changed the colour of the right segment into green to give a lane change recommendation.

Participants
A total of 41 (20 male, 21 female) participants with an age between 19 and 64 (*M*= 36.8; *SD*=14.13) took part in the present study. Participants were randomly distributed to an experimental group with ten males and eleven females and a baseline group with ten males and ten females. Regarding driving experience 56% of the participants reported to drive more than 10000km per year. Most of the participants (88%) also indicated to have no or little experience with driver assistance systems such as Adaptive Cruise Control (ACC). All participants were

recruited from the participant pool from the German Aerospace Centre (DLR). For the participation in the experiment participants were paid 10€ per hour.

Experimental design

In the present study a one factor design, with the factor *system support,* was used. All participants drove eight driving scenarios in manual mode. The presence of the assistance system "ambient light" defined the factor *system support*. This between factor had two conditions: "ambient light" and "baseline" (see Table 2).

Table 7. Experimental design of the study

System support	Ambient light warning *Lateral & Longitudinal*	Ambient light Recommendation *Lateral & Longitudinal*
Baseline	4	4
Ambient Light	4	4

Procedure

Each participant completed the experiment in approximately 1.5h on a single day. At first, all participants filled out a consent form and a demographic survey. Participants of the ambient light group were instructed that the purpose of the study was to test a new assistance system for motorways. After that, all participants experienced a test drive of 15min to familiarize themselves with the simulator. Afterwards participants of the ambient light condition were instructed that they will be supported by a new, light based assistance system while driving. The exact functioning of the system was not instructed. Participants of the baseline were not supported by any assistance system. After each scenario participants of the ambient light condition had to fill out a questionnaire regarding their understanding of the signals presented by the ambient light and its usability. At the end of the presented study a questionnaire regarding the general attitude towards the ambient light was handed out. For participants of the baseline the experiment ended at this point. Participants of the ambient light condition were now instructed about the method of operation of the ambient light and drove all scenarios again in the same order to experience the ambient light one more time (Only results of the first drive are reported). After this participants had to fill out a sort questionnaire regarding acceptance of the system. After a short debriefing participants received their payment.

Dependent Variables

In the present study the number of correct reactions was used as dependent variable. A correct reaction was defined as a specific reaction that should be triggered by the ambient light. In longitudinal warning scenarios the correct reaction was to decelerate or brake. In lateral warning scenarios participants performed a correct reaction when they didn't execute a lane change. In longitudinal recommendation scenarios participants who accelerated performed a correct reaction. Regarding lateral recommendations participants should change the lane.

Additionally, the participants rated the comprehensibility, meaning, and usability of the ambient display after each scenario on a seven-point scale (-3 to 3) with ten semantic differentials (comfortable vs. uncomfortable, safe vs. unsafe, distracting vs. not distracting, useful vs. not useful, nice vs. disruptive, permissive vs. prohibitive, forcing vs. relenting, inhibiting vs. supporting, avoiding vs. affording, limiting vs. expanding).

Results

The number of correct reactions and subjective data on the two lateral and longitudinal warning scenarios and the two lateral and longitudinal recommendation scenarios are clustered and reported in the following section.

Lateral warnings

For the lateral warning scenarios seven participants of the baseline and three of the ambient light condition were not considered in the data analyse due to technical problems. Figure 4 show that that the ambient display led to significantly less critical lane changes (six lane changes) (x^2=11.86; p=.001) compared to the baseline condition (18 lane changes). Participants evaluated lateral warnings as not limiting or expanding (*M*=0, SD=1.96) and not avoiding or affording (*M*=0, SD=2.15). Further they rated the lateral warnings as supporting (*M*=1.2, SD=1.8), forcing (*M*=1.3, SD=1.25), prohibitive (*M*=-2.2, SD=0.9), nice (*M*=0.85, SD=1.52), useful (*M*=1.38, SD=1.89), not distracting (*M*=0.74, SD=1.9), safe (*M*=1.76, SD=1.48) and comfortable (*M*=0.88, SD=1.41).

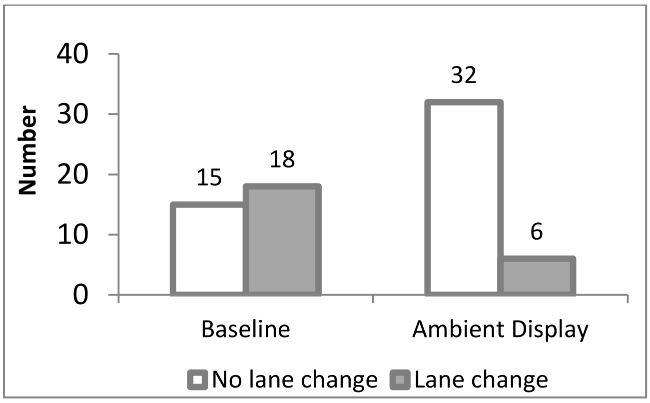


Figure 9. Number of lane changes in lateral warning scenarios

Longitudinal warnings

Regarding the question if the ambient light could trigger a braking manoeuvre the brake reaction of the drivers were compared between the conditions (Figure 5). Results showed no significant differences between the groups regarding the number of brake reactions (baseline=33, ambient display=38) (x^2=1.122; p=.289). Nevertheless, participants rated longitudinal warnings as limiting (*M*=-0.24, SD=1.76), affording (*M*=1.56, SD=1.57), supporting (*M*=0.9, SD=1.64), forcing (*M*=-1.22, SD=1.13), prohibitive (*M*=-0.98, SD=1.33), nice (*M*=0.2, SD=1.66), useful (*M*=1.46, SD=1.63), not distracting (*M*=0.8, SD=1.52), safe (*M*=1.2, SD=1.54) and comfortable (*M*=0.73, SD=1.43).

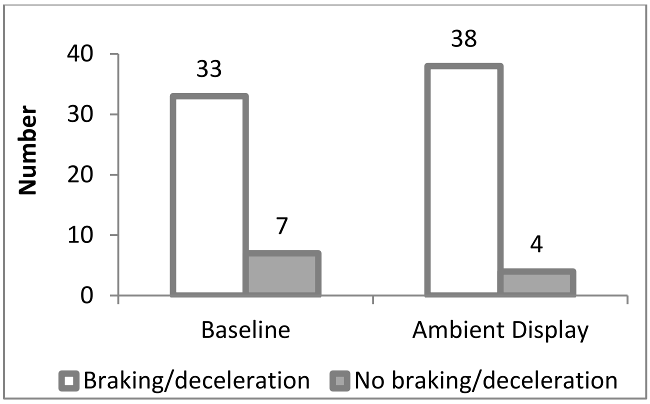


Figure 10. Number of braking/deceleration in longitudinal warning scenarios

Lateral recommendations

One participant of the ambient light condition was sorted out (only in one of the two reported scenarios) because he changed the lane before receiving a recommendation of the ambient light. In addition to warnings the ambient light communicated recommendations for driving manoeuvres. A Chi-square test showed no significance (x^2=3.59; p=.058) between the groups (see Figure 6). A further look into the scenarios showed that in the scenario with the recommendation for a lane change to the right twelve out of 20 participants driving with ambient light reacted to this recommendation and changed the lane. In the baseline condition no driver changed to the right. This led to a significant difference (x^2=17.14; p< .001) between the conditions. Participants rated the lateral recommendations by the ambient light as expanding (*M*=0.83, SD=1.43), affording (*M*=1.22, SD=1.27), supporting (*M*=1.2, SD=1.33), forcing (*M*=-0.15, SD=1.2), permissive (*M*=1.32, SD=1.35), nice (*M*=0.39, SD=1.56), useful (*M*=0.56, SD=2.01), not distracting (*M*=0.41, SD=1.77), safe (*M*=0.7, SD=1.36) and comfortable (*M*=0.61, SD=1.43).

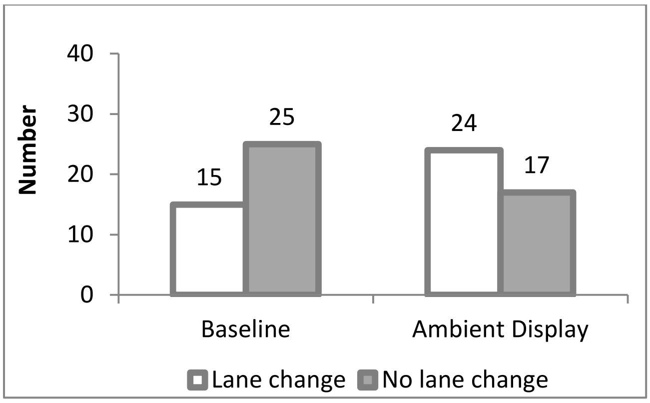


Figure 11. Number of lane changes in lateral recommendation scenarios

Longitudinal recommendations

No significant difference between the groups regarding longitudinal recommendations was found in the present study (x^2=.02; p=.886) (see Figure 7). The ambient light had no effects on the acceleration behaviour of the participants. Nevertheless, participants rated the ambient light as expanding (*M*=0.6, SD=1.54), affording (*M*=1.7, SD=1.12), supporting (*M*=1.13, SD=1.48), forcing (*M*=-0.63, SD=1.22), permissive (*M*=1.5, SD=1.61), nice (*M*=0.4, , SD=1.83), useful (*M*=1.0, SD=1.86), not distracting (*M*=0.83, SD=1.91), safe (*M*=0.33, SD=1.42) and comfortable (*M*=0.80, SD=1.61).

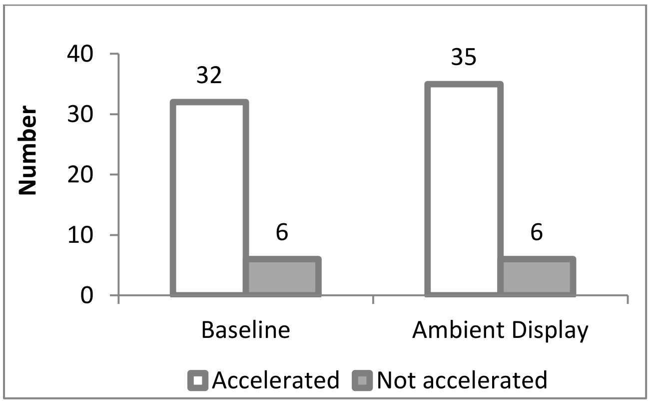


Figure 12. Number of accelerations in lateral recommendation scenarios

Subjective Data

The results for the rating of warnings and recommendation regarding the semantic differentials are displayed in Figure 8. Participants rated warnings as significantly more limiting (t_{143}= -3.074, p= .003), more avoiding (t_{144}= -2.102, p= .037), more forcing (t_{144}= -4.67, p< .001), more prohibitive (t_{144}= -12.708, p< .001), useful (t_{144}= 2.227, p= .027) and safety relevant (t_{144}= 3.732, p< .001) than recommendations. The results showed that participants could easily distinguish between a warning and a recommendation.

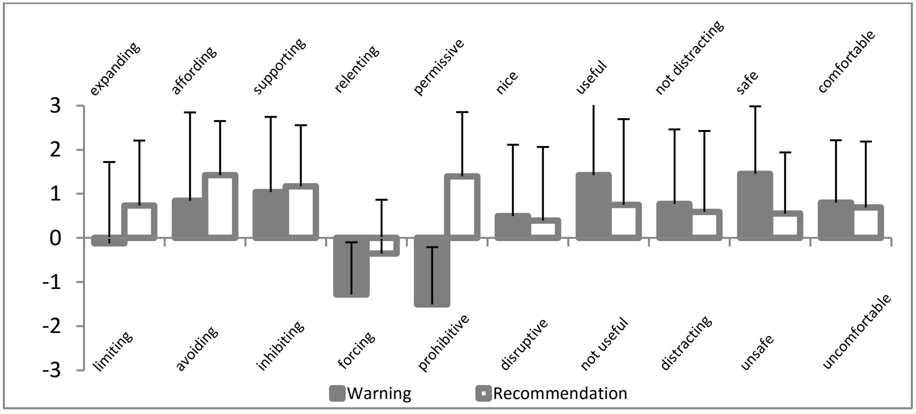


Figure 13. Subjective rating for semantic differentials

Discussion

The presented study shows the potential of a colour-coded ambient light based interaction concept for an integrative driver assistance system. Using the ambient light for warnings and recommendations reveal an ambiguous picture regarding the direction of support (lateral/longitudinal):

For lateral warnings and recommendations the ambient light had an effect on driver behaviour. More than 84% of the participants who experienced the lateral warning cancelled the lane change and avoided a collision with an approaching vehicle from behind. While the hazard was not visible for the drivers, the ambient light was able to communicate information helping the driver to anticipate future events. For lateral recommendations a significant influence on driver behaviour was found for one of the two driving scenarios in which the ambient light helped driver to change to the right lane. Compared to that, using the ambient light for longitudinal warnings and recommendations had no positive effect in this study.

From our point of view, these effects could be explained by the scenarios chosen for this study. The selected scenarios for longitudinal assistance were not sensitive enough to show any advantage of the ambient light. The needed actions were very clear and easy to understand for the drivers resulting in a high percentage of correct reactions in both the ambient light and the baseline condition (ambient light=90.5%, Baseline=82.5%). Differences between the groups could not be found because drivers did not need further support to handle these situations. Concluding, the ambient light stimuli seem to have a positive effect in all scenarios where no salient stimuli from the environment exist. Here, the ambient light helps the driver to anticipate what the correct action could be. If this positive effect is additionally dependent from the direction of the support (longitudinal/lateral) or if it is a pure effect of the scenario selection as described above could not be fully understood by the results of the study and needs further investigation.

The study also showed that participants were able to understand and distinguish between warnings and recommendations and act accordingly to it even though a generic colour coding was used for the different scenarios. The subjective rating showed that participants experienced warnings presented by the ambient light as very prohibitive, limiting and enforcing while recommendations were rated as permissive, expanding and relenting. Against our expectation, participants did not distinguish between the semantic differentials activating- preventing. It seems like participants perceive signals from the ambient light always as a recommendation to do something and not as a signal to stop an action. In depended from this interpretation, this led to the correct actions of the driver.

Outlook

The ambient light showed its potential to integrate the information of different driver assistance systems in one generic interaction design using colour codes for various recommendations and warnings. The ambient light helped to trigger appropriate actions by the driver. In further studies, the combination of the ambient light with other information channels (multimodal interaction) seems to be very promising to further enhance the effectiveness in critical situations. Furthermore, the potential of the ambient light concept for higher levels of automation should be under research to understand if the ambient light provides a generic approach to support the driver not only during manual driving but also during the automated drive. The ambient light could be used to support the driver to build up correct mode awareness by changing the colour of the light in different automation levels, to react appropriately during transitions of control or help him building up a correct situation awareness by highlighting relevant cues in the environment.

References

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*, 775-779.

Cummings, M.L., Kilgore, R.M., Wang, E., Tijerina, L., & Kochhar, D.S. (2007). Effects of Single Versus Multiple Warnings on Driver Performance. *Human Factors*, *49*, 1097–1106.

Continental. (2013). Continental Counts on LEDs as Co-pilot [homepage]. Retrieved from http://www.continental-corporation.com/www/pressportal_com_en/themes/press_releases/3_automotive_group/interior/press_releases/pr_2013_02_07_driver_focus_en.html.

Gibson, J.J. (1966). *The Senses Considered as Perceptual Systems*. Houghton Mifflin, Boston.

Henning, M.J., Kienast, H., Lindner, P., Weigel, H., Krems, J.F., Spanner-Ulmer, B. & Wanielik, G. (2008). *Multifunktionale Mensch-Maschine-Schnittstelle für Fahrerassistenzsysteme* (Multifunctional human-machine interface for driver assistance systems). In M. Maurer (Eds.), *5. Workshop Fahrerassistenzsysteme*. Karlsruhe: Fmrt.

Kelsch, J., & Dziennus, M. (2015). Gradual action-oriented ambient stimuli as means for joint driver-automation system design. *Presentation at the 6th international Conference on Applied Human Factors and Ergonomics 2015.*

Kienast, H., Lindner, P., Weigel, H., Henning, M., Krems, J., Wanielik, G. & Spanner-Ulmer, B. (2008). Aufmerksamkeitssteuerung durch räumlich gerichtete Anzeigen in Fahrerassistenzsystemen (Attention shift trough directed stimuli in driver assistance systems). *In VDI Berichte 2048*, 413-423.

Laquai, F., Chowanetz, F. & Rigoll, G. (2011, Oktober). A large-scale led array to support anticipatory driving. *2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2087 -2092.

Lee, J.D., Hoffman, J.D., & Hayes, E. (2004). Collision warning design to mitigate driver distraction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 6,* 65–72.

Löcken, A., Müller, H., Heuten, W., & Boll, S. (2015). An Experiment on Ambient Light Patterns to Support Lane Change Decisions. *Proceedings of the IEEE Intelligent Vehicles Symposium (IV),* 505–510.

Maier, K. (Eds.). (2014). Multimodaler Warnbaukasten : Entwicklung einer neuen Warnphilosophie für Fahrerassistenzsysteme zur Unterstützung unfallkritischer Bremsszenarien (Multimodal warning kit: Development of a new philosophy of warning for driver assistance systems for support in safety-critical brake scenarios). Göttingen: Cuvillier.

Pfromm, M., Cieler, S., & Bruder, R. (2013). A Driver Assistance via Optical Information with Spatial Reference. *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems*, 2006–2011.

Remington, R., Johnston, J.C. & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics*, 51, 279-290.

Scott, J.J., & Gray, R. (2008). A comparison of tactile, visual, and auditory warnings for rearend collision prevention in simulated driving. *Human Factors, 50*, 264–275.

Sivak, M. (1996). The information that drivers use: is it indeed 90% visual? *Perception, 25*, 1081-9.

Utesch, F. (2014). Unscharfe Warnungen im Kraftfahrzeug. Eignet sich eine LED-Leiste als Anzeige für Fahrerassistenzsysteme? (fuzzy warnings in vehicles. Is a LED-band usefull as a display for driver assistance systems?)(Doctoral dissertation). Retrieved from Katalog der Deutschen Nationalbibliothek. (2014).

Zarife, R. (2014). Integrative Warning Concept for Multiple Driver Assistance Systems (Doctoral dissertation). Retrieved form OPUS Würzburg. (31.07.2014).

Eco-Pedaling – Examining a highly automated eco-assistance system for pedelecs from a user perspective

Isabel Neumann, Maria Kreußlein, Peter Cocron,
Steffen Weichold, & Josef F. Krems
Technische Universität Chemnitz
Germany

Abstract

The range of motorised bicycles (i.e. pedelecs) is limited. Although conventional cycling is always a fallback-option, running out of power is an adverse event for users due to the high weight of pedelecs. In the scope of the current study, the prototype of a highly automated eco-assistance system (EAS) for pedelecs, which aimed to support users in dealing energy efficiently with the limited battery resource by automatically changing between support levels, was examined from a user-centred perspective. A field experiment was conducted with a sample of 30 experienced pedelec users investigating subjective assessments supplemented by logger data. Applying a repeated-measures design, participants drove a test track twice, once with the EAS and once by manual control (i.e. without assistance system but with electrical support). The evaluation revealed lower ratings of trust, acceptance and usability for the EAS, specifically in comparison to cycling without the EAS. Whereas users perceived a potential of the EAS to reduce workload and increase comfort while cycling, the high degree of automation was perceived as a barrier. Taken together, results indicate a rather low acceptance of highly automated assistance systems with the tested settings. Possible solutions may be a higher transparency of the system's actions or the opportunity to personalise the EAS to a higher extent.

Introduction

Besides a general increase of bicycle ownership, specifically pedelecs have become more popular during the past years in Germany and other European countries (European Cyclists' Federation, 2015). Supporting cyclists with an electric motor up to a speed of 25 kph, pedelecs combine higher driving comfort (compared to bicycles) and better sustainability (compared to cars) which in turn contributes to an increased popularity of this mean of transport. In Germany, pedelec ownership has increased to currently about 2 million pedelec-owners, even with a rising trend (Zweirad-Industrie-Verband, 2015).

Similar to battery electric vehicles (BEVs; Franke, Neumann, Bühler, Cocron, & Krems, 2012), the range of motorised bicycles (i.e. pedelecs) is limited. Although

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

conventional cycling is always a fallback-option, running out of power is an adverse event for users due to the high weight of pedelecs. Hence, especially in mountainous regions, a good management of battery ressources is essential to avoid such situations.

For BEVs, eco-driving (i.e. applying an energy-efficient driving style) has been found to be helpful in coping with critical range situations (i.e. when the remaining range is equal to or less than the distance to the destination; Franke et al., 2012; Neumann, Franke, Cocron, Bühler, & Krems, 2015). In this context, eco-assistance systems (EAS) have been shown to support the driver in reducing energy consumption (Young, Birrell, & Stanton, 2011). Therefore, it could be assumed that EAS can also support users of pedelecs in dealing efficiently with limited battery resources.

In the current study a highly automated version of such an EAS for pedelecs which aimed to reduce the energy consumption by fully automatically changing between the pedelec's support levels was evaluated. The objectives of the study were to evaluate the EAS from a user-centred perspective, also by comparing cycling with the EAS to manual control (i.e. without assistance), as well as to investigate possible related factors.

Method

Participants

A sample of N = 30 users including 8 women and 22 men with a mean age of M = 49 years (SD = 17.6) took part in the study. All participants were experienced pedelec users with 0.5 to 6.0 years of pedelec ownership. About half of them reported to experience sometimes to almost always critical range situations with the pedelec, which was the use case for the EAS. The sample was highly educated; 73% graduated from university. Due to technical problems of the EAS, the complete data sets (subjective and objective data) of 4 participants had to be excluded from the analysis.

Test pedelec and EAS

The test pedelec used in the study was a Stevens E-Caprile 25 with a Bosch Active Line 250W mid-drive electric support system. The pedelec was equipped with a prototype of an EAS for pedelecs developed by Weichold, Kriesten, Kilian, and Heinkel (2014). Depending on the actual state of charge of the battery, the driver's weight, the total length of the trip and the current slope, the EAS changed automatically between support levels (Figure 1) to ensure an energy-efficient pedalling to the destination. The interface of the EAS was displayed on a smartphone, which was fixed at the handlebar of the pedelec (Figure 1). A control unit with direct access to the CAN-BUS of the BOSCH system was connected via Bluetooth.

Figure 1. Interface of the EAS prototype (right) with support levels (upper left) and experimental setup at the pedelec (bottom left).

Study design and data collection

The study was conducted using a within-subject design testing the experimental factor *assistance* (*with EAS* versus *manual*). While participants had to change between the given support levels (turbo, sport, tour, eco and off; see also Figure 1) by themselves in condition *manual*, the EAS changed automatically between the support levels according to the algorithm in condition *EAS*.

After first obtaining information about the scope and procedure of the study, participants signed the informed consent and filled in an introductory questionnaire collecting data on sociodemographics, physical fitness and general pedelec usage. Subsequently, participants were asked to drive a predefined, mountainous route in real traffic (Figure 2) twice: once with the EAS (condition *with EAS*) and once by manual control (i.e. without assistance of the EAS but with electrical support; condition *manual*). The sequence of both experimental trials was balanced to control for order effects. For each test ride, the scenario was described as follows: “Imagine you have to cycle 80 km to your destination and the pedelec displays a remaining range of 80 km (in eco-mode).” Participants were instructed to cycle the route as energy efficiently as possible. In order to increase the compliance, all participants were told that they would receive a 20€ incentive for their participation after completing the experiment and in addition could receive extra money (5€) if they cycled particularly energy efficiently in condition *manual*. Nevertheless, all test drivers received the total 25€ for their participation upon completing the experiment.

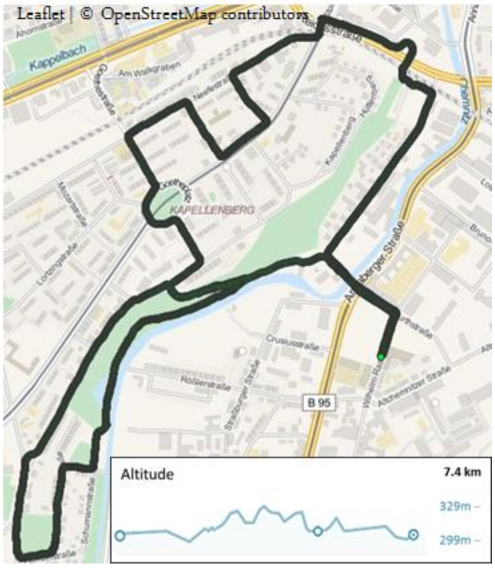


Figure 2. City of Chemnitz: test track in real traffic with elevation profile.

In the introductory questionnaire, participant’s ratings of their physical fitness were collected by answering items of the scales action, endurance and strength of the physical self-description questionnaire (Peart, Marsh, & Richards, 2005, German translation by Stiller & Alfermann, 2007). The internal consistency of all scales was good with Cronbach’s $\alpha > .87$.

Before each test ride, participants received standardised information on the EAS or cycling with manual control and filled in a pre-questionnaire about the respective system. Another questionnaire containing items on the acceptance, usability and trust in the EAS or manual control had to be answered after each of both test rides.The acceptance of both systems before and after usage was addressed by the van der Laan acceptance scale (van der Laan, Heino, & de Waard, 1997), an established instrument to evaluate attitudes towards in-vehicle devices. The scale includes two dimensions of acceptance - satisfaction and usefulness. Answers were required for nine semantic differentials (satisfaction: e.g., pleasant – unpleasant, nice – annoying; usefulness: e.g., useful – useless, bad – good). Reliability analysis revealed acceptable internal consistencies for both scales and all points of measurement (usefulness: Cronbach’s $\alpha > .80$; satisfaction: Cronbach’s $\alpha > .78$).

Trust was assessed by a scale for trust in automated systems (Jian, Bisantz, & Drury, 2000) translated by Beggiato and Krems (2013). Participants were asked to answer questions such as “I can trust the system” on a 7-point Likert scale ranging from 1 - fully disagree to 7 - fully agree. The scale was applied before and after experiencing the respective system. Reliability analysis showed acceptable Cronbach’s α ($\alpha < .76$).

Data on the usability of the EAS and manual control as well as corresponding constructs were collected via six subscales of the Technology Usage Inventory (TUI;

Kothgassner et al., 2012). They are described in Table 1. Whereas the subscales skepticism and curiosity were collected before each test ride, the constructs ease of use, utility, accessibility, and intention to use were collected after experiencing the system.

Table 1. Description of the applied sub-scales of the Technology Usage Inventory (Kothgassner et al., 2012).

Sub-scale	Description	Rating scale	Cron-bach's α
Ease of use	Perceived usability and ease of use of a specific technology. (3 items)	7-point Likert	.60 - .83
Utility	Assessment whether a specific technology is useful and appropriate and supports specific tasks in everyday life. (4 items)	7-point Likert	.83 - .90
Accessibility	Perceived opportunity to get access to a specific technology. Assessment of the person whether it is easy to get access to a technology, also from a financial point of view. (3 items)	7-point Likert	.65
Intention to use	Evaluation of the behavioural intention to use a certain technology. (3 items)	100-point Likert scale	.93 - .96
Curiosity	Evaluation of a person's curiosity about a specific technology. (4 items)	7-point Likert	.73 - .77
Scepticism	Assessment whether a specific technology is perceived as risky, severe or adverse. (4 items)	7-point Likert	.66 - .71

Note. All Likert scales ranged from agree to disagree. For the analysis, scores were calculated for each sub-scale by summarising the values for each item of the scale.

Additionally, open-ended questions were asked to assess perceived advantages and disadvantages of the EAS compared to manual control after experiencing both systems. Participants' statements were coded using thematic analysis by Braun and Clarke (2006). Furthermore, the number of participants who mentioned a specific category was identified.

Supplementary to users' assessments, the implemented data logger collected driving parameters, such as velocity, location (GPS), energy consumption and support level for each trip with a sample rate of 10 Hz. Due to missing logger data caused by technical problems of the data acquisition system, a reduced sample of $N = 17$ participants was available for the analysis of logger data.

Results

Users' assessments

Acceptance

A repeated measures ANOVA revealed that perceived usefulness for the *EAS* was significantly lower than for *manual* ($F(1, 25) = 10.50$, $p = .003$, $\eta_p^2 = .30$; Figure 3). There was neither a significant difference between *pre* and *post* ($F(1, 25) < 1$, NS, $\eta_p^2 > 0.01$) nor a significant interaction effect ($F(1, 25) = 1.52$, NS, $\eta_p^2 > 0.06$)

For perceived satisfaction, the second sub-scale of the acceptance scale (van der Laan et al., 1997), no significant differences were detected (Figure 3). There was neither a significant effect of *assistance* ($F(1, 25) = 3.59$, $p = .070$, $\eta_p^2 = 0.13$), *time* ($F(1, 25) < 1$, NS, $\eta_p^2 = 0.01$), nor a significant interaction effect ($F(1, 25) = 2.08$, NS, $\eta_p^2 > 0.08$).

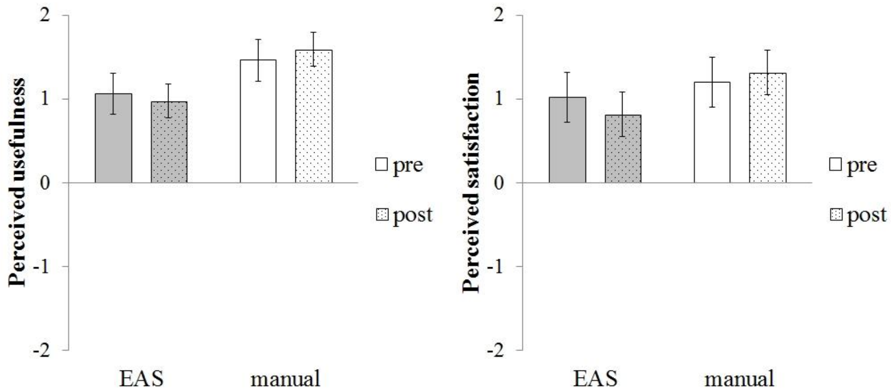


Figure 3. Perceived usefulness (left) and satisfaction scores (right) for the EAS and manual control (N = 26). Error bars represent 95% CI.

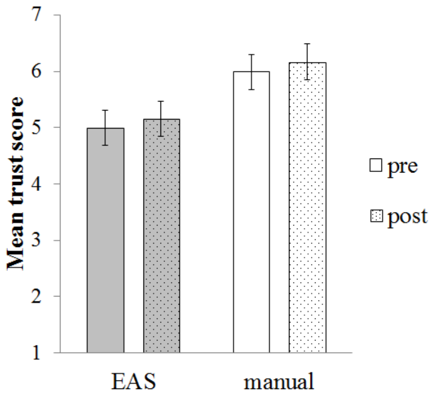


Figure 4. Mean trust score for the EAS and manual pedaling before and after usage (N = 25). Error bars represent 95% CI.

*Note.*7-point Likert scale (1 – completely disagree to 7 – completely agree).

Trust
A repeated-mesures ANOVA resulted in a significant effect of *assistance*. As can be seen in Figure 4, higher trust scores were observed for *manual* compared to *EAS* ($F(1, 24) = 26.23$, $p < .001$, $\eta_p^2 = 0.52$). No effect of *time* ($F(1, 24) = 1.55$, NS, $\eta_p^2 = 0.06$) or interaction ($F(1, 24) < 1$, NS, $\eta_p^2 < 0.01$) was observed.

Results of the scores for scepticism and curiosity collected before experiencing the system showed significantly higher scepticism values for the *EAS* compared to *manual* ($t(25) = -3.60$, $p = .001$, $d = 0.71$) but no significant differences regarding the reported curiosity about both systems were obtained ($t(25) = -1.44$, NS, $d = 0.28$).

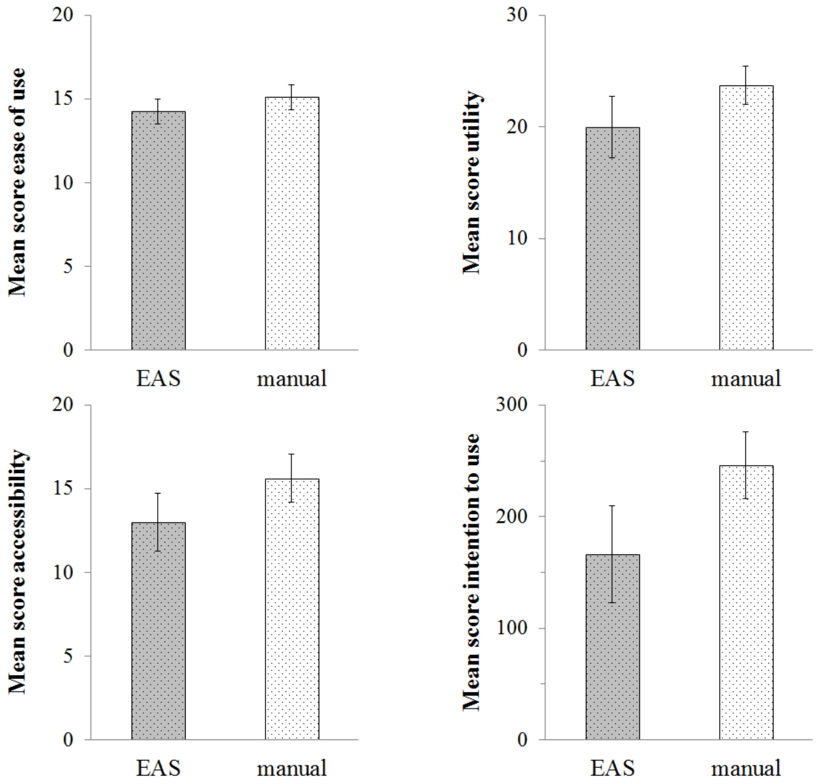


Figure 5. Scores of the TUI for the evaluation of manual and EAS after experiencing the respective system (N = 26 for all scores). Error bars represent 95% CI.

Note. Scores for ease of use and accessibility range between 3 and 21, for utility: between 4 and 28, and for intention to use: between 0 and 300. Higher values represent better evaluations.

In sum, the results of the sub-scales of the TUI consistently reflected a better assessment of *manual* control compared to the usage of the highly automated *EAS* in terms of ease of use, accessibility, utility and, arguably as a consequence, also a higher intention to use *manual* control than the *EAS*. Whereas no differences between both systems were found for curiosity about both systems, participants reported significantly more scepticism abou the *EAS* compared to *manual*.

Perceived advantages and disadvantages of the EAS

Based on participants' statements about the EAS, six categories of perceived advantages were identified (Figure 6). Amongst them the potential to reduce workload and distraction, and thus an increased opportunity to put attention on traffic, was the most often mentioned category (35% of participants) followed by an enhanced energy efficiency (31% of participants) when cycling with the EAS. Besides, six categories of potential disadvantages of the EAS emerged from participants' reports. In this context, the high degree of automation of the assistance system with very limited options to personalise was experienced as the main weakness of the system (42% of participants). This might be especially relevant in combination with the prototype status of the EAS and related problems like a delayed response of the system (e.g. delayed change of support level; 27% of participants) or the selection of an inappropriate support level (19% of participants). Figure 6 provides a brief overview of all categories of perceived advantages and disadvantages of the EAS reported by the users.

Table 2. Perceived advantages and disadvantages of the EAS with absolute frequency of reporting.

Perceived advantages (number of participants)	Perceived disadvantages (number of participants)
Reduced distraction (n = 9)	No opportunity for personalisation (n = 11)
Energy efficiency (n = 8)	Delayed response of the system (n = 7)
Comfort (n = 4)	Inappropriate support level (n = 5)
Possibility to be faster (n = 3)	System not fully developed (n = 2)
Usability & user friendliness (n = 3)	Sub-optimal energy efficiency (n = 4)
Advantage for inexperienced users (n = 2)	Distraction (n = 1)

Logger data

The analysis of logger data revealed a significantly higher energy consumption with *EAS* compared to *manual* ($t(16) = 4.07$, $p = .001$, $d = 1.00$; Figure 7) with no significant differences for mean trip speed ($t(16) = -2.55$, $p = .802$, $d = 0.05$) or mean travel time ($t(16) = 0.19$, $p = .849$, $d = 0.06$). Consistently, those support levels, which provide strong support and in turn consuming more energy, were used more often in condition *EAS* compared to *manual* (e.g., support level turbo: $t(16) = 6.11$, $p < .001$, $d = 1.48$). These results imply that the settings of the EAS could have been chosen even more energy saving and in turn less comfortable for this sample. Maybe also the opportunity to earn extra-money when cycling specifically energy saving might have lead to a specifically high energy efficiency. The possibility to adapt the level of comfort or energy efficiency could further contribute to a person's optimal energy efficiency.

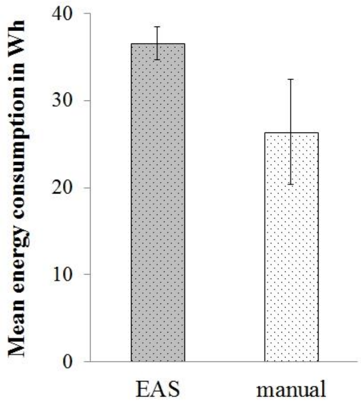


Figure 7. Mean energy consumption during experimental conditions (N = 17). Error bars represent 95% CI.

Furthermore, the analysis revealed high inter-individual differences for energy consumption during *manual* (Figure 7) which is correlated with the participants' BMI (r = .45, p = .041) and their self-assessment of physical fitness (r = -.55, p = .010). That means, not surprisingly, a low BMI and a high score of physical fitness were related to high energy efficiency during normal driving.

Discussion and conclusions

In the scope of the current study a highly automated EAS was examined and compared to manual control. Results of the subjective assessments consistently revealed significantly lower ratings of trust, acceptance, and usability for the EAS compared to manual control. Looking at the supplementary qualitative data regarding perceived disadvantages of the EAS might explain these results to some extent. Thus, the high degree of automation of the EAS combined with the missing possibility to override or personalise the system and some issues related to the prototype status of the EAS, might have led to a rather sceptical evaluation of the EAS. In this context, it should also be kept in mind that participants were rather familiar with cycling by manual control as they were all experienced pedelec users. The EAS was a completely new system which was experienced only for a short test ride. This could have influenced users' ratings; an assumption which might be also supported by the higher ratings of scepticism regarding the EAS compared to manual control. However, results of the qualitative analysis reflect a dichotomy of users' assessment of the highly automated system: On the one hand such systems bear potential such as a possible increase of traffic safety (participants' statements about reduced distraction during EAS usage) and comfort while driving. On the other hand, as stated above, findings imply that the high degree of automation of the EAS was perceived as a barrier by users as it was impossible to override the system.

Possible solutions for the reported disadvantages might be, beyond further technical fine tuning of the system, to implement opportunities for personalisation such as a personal level of comfort for the ride or a personal level of physical fitness. Additionally, an increase of the transparency of the system's actions might also be helpful, for instance by incorporating relevant feedback in the user interface of the application.

Results of the logger data showed significantly lower energy consumptions in the manual control condition compared to the EAS condition, whereas no differences for velocity or travel time were observed. This implies on the one hand that users were really motivated to drive energy efficiently, even though this meant higher physical effort. On the other hand the observed high interindividual variance which was connected to users' physical fitness could be regarded as another reason to integrate more options for customisation of the system. Thus, increased energy efficiency during EAS usage could be derived if users accept less comfort (i.e. higher physical effort) during the ride.

In sum, findings of the study reflect the need to design highly automated assistance systems in a way that the actions of the system are reasonable and transparent for the user as also discussed for systems in the automobile context (e.g. Bengler et al., 2014). Given the perceived advantages of the EAS combined with an increased popularity of pedelecs as means of transport (European Cyclists' Federation, 2015), systems like the EAS have the potential to increase comfort and safety for pedelec-owners in the future, but the present study shows that users' requirements should be taken into account to successfully introduce such systems to the market.

Acknowledgements

This research was funded by the Free State of Saxony (grant number: 100172504) in the scope of the initiative "Elektromobilität verbindet" ["electromobility connects"] as part of the German programme "Schaufenster Elektromobilität" ["showcase electromobility"]. Statements in this paper reflect the authors' views and do not necessarily reflect those of the Free State of Saxony or of other partners involved in the project.

References

Beggiato, M., & Krems, J.F. (2013). The evolution of mental model, trust and acceptance of adaptive cruise control in relation to initial information. *Transportation Research Part F: Traffic Psychology and Behaviour, 18*, 47-57.

Bengler, K., Dietmayer, K., Farber, B., Maurer, M., Stiller, C., & Winner, H. (2014). Three decades of driver assistance systems: Review and future perspectives. *Intelligent Transportation Systems Magazine, IEEE*, *6*(4), 6-22.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*, 77-101.

European Cyclists' Federation (2015). *Cycling sells: Three of EU's biggest markets see substantial increase in bike sales in 2014*. Retrieved from: http://www.ecf.com/news/cycling-sells-three-of-eus-biggest-markets-see-substantial-increase-in-bike-sales-in-2014/.

Franke, T., Neumann, I., Bühler, F., Cocron, P., & Krems, J.F. (2012). Experiencing range in an electric vehicle: Understanding psychological barriers. *Applied Psychology: An International Review*, *61*, 368-391.

Jian, J.Y., Bisantz, A.M., & Drury, C.G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*, 53-71.

Kothgassner, O.D., Felnhofer, A., Hauk, N., Kastenhofer, E., Gomm, & J., Kryspin-Exner, I. (2012). *TUI: Technology Usage Inventory Manual*. Wien: FFG.

Neumann, I., Franke, T., Cocron, P., Bühler, F., & Krems, J.F. (2015). Eco-driving strategies in battery electric vehicle use–how do drivers adapt over time? *IET Intelligent Transport Systems*, *9*, 746-753.

Peart, N., Garry, R., & Marsh, H. (2005). The Physical Self Description Questionnaire: Furthering Research Linking Physical Self-Concept, Physical Activity and Physical Education. In P. Jeffrey (Ed.) *Proceedings of the AARE Annual Conference*, Parramatta, Australia: Australian Association for Research in Education.

Stiller, J., Alfermann, D. (2007). Die deutsche Übersetzung des Physical Self-Description Questionnaire (PSDQ): Psychometrische Kennwerte und Befunde zur Reliabilität und Validität für die Altersgruppen Jugendliche und junge Erwachsene. *Zeitschrift für Sportpsychologie*, *14*, 149-161. Göttingen: Hogrefe Verlag.

Van Der Laan, J.D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies*, *5*, 1-10.

Weichold, S., Kriesten, D., Kilian, W., Heinkel, U. (2014). Energieoptimierung und Verbesserung der Reichweitenvorhersage für Pedelecs. In H. Proff (Ed.), *Entscheidungen beim Übergang in die Elektromobilität* (pp. 479-492). Duisburg, Germany: Springer Fachmedien.

Young, M. S., Birrell, S. A., & Stanton, N. A. (2011). Safe driving in a green world: A review of driver performance benchmarks and technologies to support ‘smart’ driving. *Applied Ergonomics*, *42*, 533-539.

Zweirad-Industrie-Verband (2015). *Die Fahrradbranche startet erneut mit Schwung in die Saison.* Retrieved from http://ziv-zweirad.de/de/news/detail/article/eurobike-2015-branchengespraech-mit-zahlen-und-fakten-aus-der-fahrradbranche/

Boundary conditions for safe detection of clinical alarms: An observational study to identify the cognitive and perceptual demands in an Intensive Care Unit

Kathrin Lange[1], Miriam Nowak[1], Richard Zoller[2], & Wolfgang Lauer[1]
[1]Federal Institute for Drugs and Medical Devices, Bonn, [2]University Hospital Bonn, Germany

Abstract

Many medical devices found in intensive care units (ICUs) use alarms to inform the user of critical, potentially life-threatening conditions. Urgent information is typically indicated by auditory stimuli, because audition is regarded a sentinel sense particularly suited for alerting purposes. In contrast to this general assumption, however, empirical evidence from the laboratory shows that even apparently automatic perceptual processes (such as the detection of an alarm) depend on the availability and allocation of processing resources. Aim of the present study was to characterise the work conditions in a German ICU with respect to factors that influence the perceptual and cognitive load and that may, therefore, compromise alarm detection. Seven experienced intensive care nurses were shadowed during their morning shifts to identify the time consumption and occurrence rates of individual activities, proportions of multi-tasking, and the frequency of task-switching and disruptive events. Amounts of multi-tasking (24% of the time spent on manual tasks) and task-switching (on average every two minutes) were considerable. Moreover, nurses were interrupted (e.g. by an in-room alarm) about every 3 minutes. Based on findings such as these, further studies may systematically investigate in how far the work conditions in an ICU are suited for reliably detecting alarms.

Introduction

In clinical context, device alarms are used to inform care-givers of hazardous situations such as device malfunctions or critical changes of the patient’s vital functions in order to trigger the necessary intervention. This goal can, however, only be achieved, when the care-giver detects the alarm, in the first place, and there is evidence from case reports that this may not always be the case. Notably, stimulus detection not only depends on physical features of the stimulus but also on the available processing resources. For instance, in a recent flight-simulator study, almost 40% of the participants did not report a landing gear alarm that sounded while they were dealing with a critical wind-shear situation (Dehais et al., 2014). This finding demonstrates that even salient stimuli may escape awareness when processing resources are strongly engaged otherwise (see also Simons & Chabris, 1999). This may be a serious problem in clinical work conditions like the intensive

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

care unit (ICU), where nurses have to frequently rely on device alarms to keep track of the patients' vital status and failures to detect an alarm may have fatal consequences.

The degree to which stimulus awareness is compromised by an on-going task has been shown to be associated with the task's perceptual (e.g. Cartwright-Finch & Lavie, 2007; see also Konstantinou & Lavie, 2013) and cognitive demands (e.g. Fougnie & Marois, 2007; Simons & Chabris, 1999). Hence, the risk of missing an alarm on the ICU may depend on which other activity the nurse is currently performing. Moreover, cognitive load also depends on other factors, for instance on the necessity to deal with multiple tasks in a given time (task-switching; for reviews see Kiesel et al., 2010; Monsell, 2003). Clinical work is also described as highly fragmented - determined by frequent switches between tasks (and also multi-tasking) (e.g. Cornell et al., 2010, see also Berg, Ehrenberg, Florin, Östergren, & Göransson, 2011; Berg et al., 2013; Chisholm, Collison, Nelson, & Cordell, 2000; Kalisch & Aebersold, 2010; Tucker & Spear, 2006 ; Walter, Li, Dunsmuir, & Westbrook, 2014; Westbrook, 2014). This is typically discussed in the context of interruptions, which are assumed to contribute to medical errors because of their demands on prospective memory (for reviews Grundgeiger & Sanderson, 2009; Li, Magrabi, & Coiera, 2012).

Goal of the current study

To the best of our knowledge, it has not yet been investigated, whether and how the perceptual and/or cognitive demands that follow from working in a clinical environment impair the detection of device alarms. Answering this question is important, because conscious detection may be regarded a prerequisite for adequately responding to an alarm – and hence to the alarm's effectiveness. The present observational study was a first step to reaching this goal. It aimed at describing and quantifying the work conditions on an ICU (because alarms may be particularly relevant on an ICU) by recording nurses' locations and activities, as well as external distracters. These conditions shall be mimicked in future studies to investigate their impact on alarm detection (for a similar approach see Edworthy, Meredith, Hellier, & Rose, 2013).

Methods

Study setting

The study was conducted in the surgical ICU of the University Hospital Bonn. Seven different nurses were shadowed (five nurses twice, two nurses once) during their morning shifts (between 7.00-10.00 a.m. or 10.00 a.m.-1.00 p.m.). To minimize the potential impact of the observers, participating nurses were recruited only among nurses with a special training in intensive care and at least 1.5 years of experience as an intensive care nurse. The observation setting was a standard double patient room, consisting of two compartments, which could be separated by a sliding wall. Ethics approval was obtained from the ethics committee of the University Hospital Bonn before starting the study. For confidentiality reasons, neither audio- nor video-recordings were made. All participating nurses gave their written informed consent. Other nurses and the physicians of the ICU were informed by notices posted both in the nurses' and the physicians' station. Patient-related data were not recorded. When a patient was conscious, one of the observers explained the purpose of their presence and ensured that the patient was comfortable with the situation.

Data collection

In six of the twelve observation sessions, only a single nurse was present, who was responsible for two (5 sessions) or one (1 session) patients. In the remaining six sessions, the participating intensive care nurse was accompanied by another nurse, who underwent on-the-job training for qualification reasons. These dyads were always assigned two patients. The average duration of the individual observation sessions was 165 minutes, ranging from 151 to 192 minutes. Since the goal of the current study was to assess the demands imposed on intensive care nurses *while monitoring the patient with the help of alarms*, activities occurring while the monitoring task was delegated to a colleague were not recorded. These instances included: breaks, helping a colleague in a different room, abandoning the ward for an external emergency call, and transporting a patient to the operating theatre. Whenever any of these instances occurred, the recording was suspended until the nurse returned to the patients' room. Correcting for these times yielded individual observation durations between 124 and 168 minutes (mean: 152 minutes, SD = 12).

Observational procedure and activity categories

Observational data were collected simultaneously by two of the authors, a psychologist (KL) and a medical engineer (MN). Prior to the study, the observers had visited the ICU several times (for approximately 20 hours in total) to acquaint themselves with the specific work environment. Starting with a list of nursing activities provided by the nurse management, activity categories were defined that could be identified by observable cues and that were regarded to be roughly homogeneous with respect to the involvement of perceptual, cognitive and motor processes. These categories were further refined during five pilot sessions (included in the pre-study visits to the ICU) and supplemented by additional categories for different kinds of disruptive events (i.e. alarms, telephone ringing, and being disrupted by a person). The final list of observation categories, together with the associated observable cues, is given in Table 1. For recording, a custom-made observation tool was used that was programmed as an Android application on a Samsung Galaxy tablet PC.

Table 8. Overview of observation categories and associated Kappa values

	Nurse's position	**Defining cue**	**Temporal characteristic**	**Kappa value (SD)**
		Manual activities		
Taking blood	patient room	at the patient, at IV line with syringe and sample tube	duration	0.88 (0.13)
Preparing drugs	patient room	at the medication preparation desk, handling drugs and infusion solution	duration	0.80 (0.08)
Managing lines, catheters, tubes	patient room	at the patient or device, manipulating or looking at catheters/lines or ventilation tube	duration	0.59 (0.23)
Operating a medical device via user interface (standard)	patient room	at the device, pressing buttons on user interface	duration	0.68 (0.10)
Stopping alarms	patient room	at the device, pressing buttons on user interface in response to an alarm	event	0.45 (0.21)
Other interactions with medical devices (non-standard, special devices)	patient room	at the device, involving interactions other than pressing buttons, may include additional equipment	duration	0.81 (0.10)
Fetching supplies	hall/other rooms	leaving the patient room and returning with supplies (medication, other)	duration	0.72 (0.21)
Blood gas analysis	hall	at specific device for blood gas analysis with blood sample	duration	0.90 (0.10)
Documenting	patient room	using patient data management system at specific terminal, using the PC	duration	0.90 (0.05)
Washing and changing dressings	patient room	at patient, using special equipment (wash mitt, dressing material)	duration	0.87 (0.09)
Bedding the patient and changing linens	patient room	at patient, together with a colleague; handling linens	duration	0.83 (0.13)
Tidying	patient room	anywhere in patient room, moving any kind of equipment	duration	0.57 (0.37)

Observing activity				
Observing the patient or a device	patient room	anywhere in patient room, looking at the patient or a device	duration	0.41 (0.14)
Oral activities				
Exchanging information with other nurse/physician	any location	other nurse(s) or physician(s) present, talking, listening and responding	duration	0.65 (0.14)
Teaching	any location	less experienced nurse(s) present, talking, listening and responding	duration	0.87 (0.15)
Talking to the patient	patient room	at patient, patient conscious, talking, listening and responding	duration	0.66 (0.25)
Disruptive events				
Direct interruption by a person	any location	other person entering and addressing nurse	event	0.43 (0.31)
Alarm in patient-room 1	any location	Alarm audible in current position of nurse, originating in patient room 1	event	0.74 (0.11)
Alarm in patient-room 2	any location	Alarm audible in current position of nurse, originating in patient room 2	event	0.72 (0.13)
Alarm in another patient room	any location	Alarm audible in current position of nurse, originating neither in patient room 1 nor 2	event	0.39 (0.14)
Telephone ringing	any location	Telephone ringing audible in current position of nurse	event	0.66 (0.13)

Activities were defined by location (e.g. at the patient's bed), the *effector* involved (hands, mouth, eyes) and the *tool or device that was handled*. To account for natural limitations to simultaneously performing several activities with the same effector, we grouped 'manual activities', 'oral activities', and 'observing' into separate groups and configured our observation tool to prevent the simultaneous coding of individual activities *within* these groups[1]. An active activity would automatically terminate when a new activity of the same group was started. Activities from other

[1] The only exception to this was "Stopping alarms", which could be performed one-handedly while performing a different task with the other hand, e.g. holding the patient or a syringe.

groups were not affected. For instance, when a nurse worked on the data entry terminal of the electronic patient data management system (PDMS), the manual activity “documenting” would be activated by the observer. When the nurse now read a value from a device while remaining at the PDMS terminal and then entered the value into the PDMS, “observing patient/device” would be coded simultaneously to “documenting”. By contrast, when the nurse moved to the device and pressed some buttons to extract a value, the manual activity “operating a medical device via user interface” would be coded. This would terminate “documenting”. When, in the latter case, the nurse entered the value into the PDMS system, this would trigger a second stint of “documenting”. These two examples would yield one versus two instances or occurrences of “documenting”, respectively.

Observer agreement

To obtain an estimate on the reliability, Kappa-values (Bortz & Lienert, 2008) were calculated for individual categories across 60 s time windows (see also Table 1). Most Kappa values were well above .60, indicating substantial or better observer agreement (e.g. Landis & Koch, 1977). Observer agreement was somewhat poorer for “Checking on and manipulating catheters/lines and tubes”, “Tidying”, “observing patient/device” as well as for “Stopping alarms” and “Interruptions by person”, most likely due to the elusive nature of these categories, i.e. these events may more easily be missed by at least one of the observers.

Data pre-processing

The coded activity categories, together with their start and stop times, were saved to an Excel-file. Subsequent to each observation session, the following processing steps were performed on these raw data. Firstly, each observer went through the records individually to correct entries marked as erroneous during observation. Secondly, individual recordings of observers 1 and 2 were compared. Gross deviations in the selected categories were discussed until agreeing on a common solution. The individual observation files were then changed accordingly. From these raw data, information concerning the duration of activities, the occurrence rate of activities and disruptive events, simultaneous activities and switching were extracted as detailed below.

Results

Time consumption, typical durations, and occurrence rates of activities

For each observation file, individual instances of activities were counted (i.e. each time an activity was started was counted a new instance of this activity). For each observation session, event counts were converted to hourly rates of occurrence, separately for each activity. Additionally, the total time share of each activity was calculated by summing the durations of individual instances and dividing by the observation time. For each observation session, the final values of total durations and occurrence rates were computed by averaging across the values obtained for the two observers. To obtain the most typical duration for individual occurrences of each activity, the individual instances were assigned to one of the following duration categories: 10 s or shorter, 11-30 s, 31-60 s, 61-120 s, 121-240 s, 241-360 s, 360 s or longer. Relative proportions of instances in the individual duration categories were determined and averaged across observers (see Tables 2 and 3).

The most time-consuming activities were “Information exchange”, followed by “Washing/Dressings”, “Documenting”, “Preparing drugs”, and “Linens/bedding” (Table 2). As to the most typical duration of individual instances of activities, this was less than 10 or 11 to 30 s for activities like “Lines/tubes”, “Device (standard)” or “Observing”. Other activities were typically performed for somewhat longer durations, e.g. “Documenting” or “Preparing drugs” (11- 60 s), “Linens/bedding” (31-120 s), or “Washing/dressings” (61 s-240 s). Notably, substantial proportions instances longer than 240 s were only found for “Linens/bedding”, “Washing/dressings” and “Information exchange” (Table 2).

Table 9. Mean percentages of observation time consumed by the different manual and non-manual activities (N = 12) and percentage distributions of individual instances of the activities over duration categories

	Mean (%)	SD	≤ 10 s	11-30 s	31-60 s	61-120 s	121-240 s	241-360 s	> 360 s
Information exchange	19.72	7.70	25.19	26.12	18.09	17.41	7.40	2.44	3.36
Washing/ dressings	15.77	8.92	6.35	17.52	17.31	23.54	18.65	6.53	10.09
Documenting	9.55	2.99	10.53	30.31	25.04	24.12	9.10	0.90	0.00
Preparing drugs	9.05	4.43	15.35	24.18	29.02	18.64	11.75	0.82	0.24
Linens/bedding	8.26	5.97	14.80	18.55	23.83	20.30	8.15	7.54	6.83
Talking to patient	7.09	7.63	40.56	21.13	18.09	11.81	4.56	2.73	1.13
Device (nonstandard)	5.38	3.59	3.35	22.94	24.47	36.28	10.77	1.85	0.34
Fetching supplies	5.16	4.44	25.97	24.94	24.95	15.93	4.96	2.73	0.52
Teaching	4.98	7.63	12.97	39.90	21.29	7.15	13.98	2.75	1.96
Lines/tubes	4.67	1.98	33.79	38.49	20.43	6.76	0.54	0.00	0.00
Device (standard)	3.81	1.77	49.93	29.87	15.00	3.26	1.93	0.00	0.00
Observing	3.72	1.84	61.33	26.52	8.32	2.30	1.53	0.00	0.00
Taking blood	1.93	1.48	14.53	15.67	32.84	23.10	9.32	4.55	0.00
Blood gas analysis	1.90	1.11	12.38	9.50	13.49	51.72	12.92	0.00	0.00
Tidying	1.21	1.07	32.85	41.62	14.04	11.49	0.00	0.00	0.00

The most frequent activities were “Information exchange”, “Observing”, “Device (standard)”, “Lines/tubes”, “Documenting”, and “Preparing drugs” (Table 3). Notably, the particularly frequent activities “Observing”, “Operating (standard)” and

"Lines/tubes" did not rank highly in overall time-consumption – these activities typically took only around 10 seconds. By contrast, "Washing/dressings" and "Linens/bedding", which were relatively time-consuming in total, were restricted to fewer, but longer occurrences within a shift.

Table 10. Mean numbers of occurrences/hour for the different manual and non-manual activities (N = 12)

	occurrences/hour	SD	Min	Max
Information exchange	12	6	4	21
Observing	8	4	5	15
Lines/tubes	7	2	3	11
Device (standard)	7	4	4	18
Documenting	6	2	3	11
Preparing drugs	6	3	2	9
Talking to patient	5	6	0	18
Fetching Supplies	4	2	2	7
Washing/Dressings	4	2	0	7
Device (nonstandard)	3	2	0	6
Linens/bedding	3	2	1	7
Stopping alarms	3	4	1	17
Teaching	3	4	0	10
Tidying	2	1	0	4
Taking blood	1	1	0	2
Blood gas analysis	1	1	0	3

Performing activities simultaneously

For each manual activity the percentage of time was determined that this activity was performed in parallel to observation and the communication activities, respectively. In 24% of the time spent on any of the manual tasks, observing or any form of communication was additionally performed. However, there were differences between activities (Table 4). The activities with the highest percentages of time shared with observing (patient or device) were managing lines/tubes and standard interactions with a device. Information exchange with a colleague occurred for almost one third of the time spent changing linens/bedding the patient and was also frequent while operating the blood gas analyser, managing lines/tubes or documenting. Finally, while changing linens/bedding and while washing, nurse often talked to the patient.

To determine the fragmentation of the work on the ICU counted the number of switches between the different manual activities was counted. A switch from activity A to activity B was defined, if *stopping* activity A was followed by *starting* activity B within 30 s. Otherwise a switch to "nothing" was coded (these were not further considered in the analyses). Numbers of switches were first counted separately for each observer and then averaged across observers. To determine switching rates, total numbers of switches were summed across activities and divided by the observation time, separately for each observation session. It showed that nurses switched between manual activities rather frequently: On average once every 110 s (SD = 35; between once every 53 s and once every 167 s).

Table 11. Mean percentages of time the manual activity presented on the left (rows) was accompanied by observing, information exchange with a colleague, teaching, or talking to the patient (the standard deviation is presented in parentheses).

	observin g	information exchange	teachin g	talking to patient	sum
Taking blood	1.32 (3.11)	7.50 (14.90)	5.96 (19.78)	4.55 (8.12)	19.3 3
Lines/tubes	5.60 (5.52)	15.91 (17.52)	2.92 (5.41)	5.79 (6.37)	30.2 2
Device (standard)	7.46 (7.17)	11.70 (10.58)	5.36 (8.63)	3.64 (5.20)	28.1 7
Device (nonstandard)	1.80 (3.83)	13.60 (15.20)	7.25 (15.06)	7.38 (9.46)	30.0 2
Fetching supplies	0.02 (0.06)	10.75 (10.14)	1.80 (5.26)	0.14 (0.25)	12.7 0
Documenting	3.11 (2.46)	15.61 (14.87)	1.39 (3.94)	0.98 (1.74)	21.0 8
Washing/ Dressings	0.30 (0.34)	10.74 (12.43)	4.77 (13.41)	12.27 (16.11)	28.0 8
Linens/ bedding	2.45 (4.27)	29.54 (25.74)	5.33 (14.39)	15.81 (19.00)	53.1 2
Preparing drugs	1.52 (1.79)	8.44 (9.00)	3.95 (6.41)	3.23 (5.23)	17.1 4
Blood gas analysis	0.21 (0.67)	16.30 (16.83)	0.18 (0.58)	0.00 (0.00)	16.7 0
Tidying	0.55 (0.95)	3.11 (5.15)	0.51 (1.75)	1.77 (3.36)	5.93

Task-Switching: Overall frequencies and individual transition probabilities

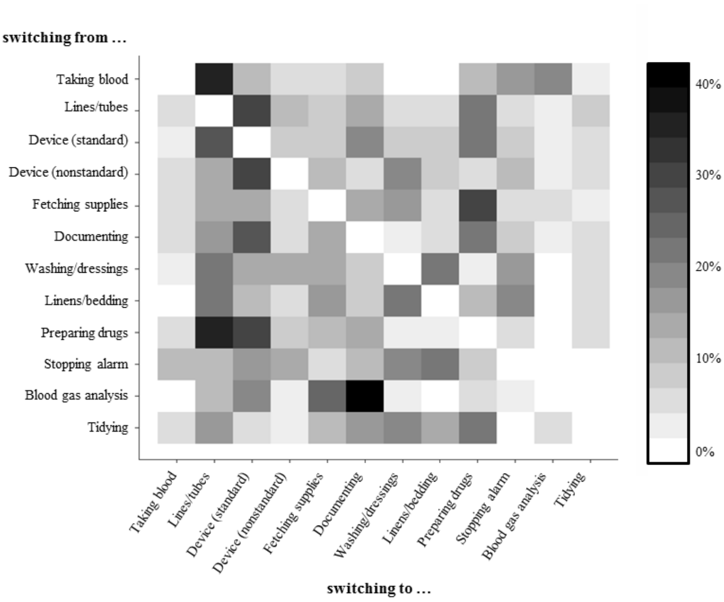


Figure 1. Distribution of switches from individual manual activities (y-axis) to any other manual activity (x-axis). Values are shown as percentages relative to the total number of switches from the activity plottet on the y-axis.

Switching from one activity to another was not arbitrary (see Figure 1, which shows the percentage of switches from any activity depicted on the y-axis that was directed to any of the activities presented on the x-axis, relative to all the switches originating from the activity on the y-axis). For instance, when the current activity was "preparing drugs", nurses most likely switched to "managing lines/tubes". Likewise, "Managing lines/tubes" was followed in most instances by a "standard interaction with a device". Assuming that particularly frequent transitions between activities reflect their procedural relatedness, an additional switching rate was computed that accounted for switches between highly related activities (i.e. with more than 15% of the switches from activity A directed at activity B). The resulting values showed that a switch occurred on average every 225 s (SD = 84), with a minimum of one switch every 91 s and a maximum of one switch every 417 s.

Disruptive events

Finally, direct interruptions by colleagues, alarms from within the double patient's room, alarms from other rooms and telephone ringing were counted. Direct interruptions by a colleague and alarms from within the double patient room were regarded "task relevant" disruptive events, because the shadowed nurse was the addressee. Alarms from other rooms and the ringing of the telephone were primarily addressed at other members of staff, hence "task irrelevant" for the observed nurse.

During observation, not every single sound was transcribed to the observation tool, but a note was taken whenever an alarm was interpreted as a new event. For describing the load associated with disruptive events, the observation time was divided into 60 s time-epochs and the proportion of time-epochs that contained at least one disruptive event was counted. At least one task-relevant disruptive event occurred in 36% of the 60-second time windows (SD = 9, range between 20% and 47% across observation sessions). Task-irrelevant disruptive events were even more frequent: These occurred in 47% of the time windows, on average (SD = 14, range between 22% and 70%).

Discussion

The present observational study was aimed at identifying exemplary work conditions on an ICU with respect to factors that may – from a theoretical perspective - impact on perceptual and cognitive load. This includes the time share of the different nursing activities, as well as their distribution over time and the presence of potentially distracting events.

Most of the activities performed by nurses may be assumed to draw on perceptual and cognitive functions (documenting, preparing drugs, device – standard and nonstandard, fetching supplies, lines/tubes, observing, taking blood, blood gas analysis). Overall, nurses spent almost half of their time on these activities. Predominantly motor tasks (washing/dressings, linens/beddings, tidying) accounted for almost 25% of the time. In about one third of the time (often overlapping the manual activities), nurses verbally communicated with colleagues or with the patient – which should also tax perceptual and cognitive processes (see also Klink, 2012), particularly in noisy environments such as the ICU (for review see Konkani & Oakley, 2012). Interestingly, although the observation categories and the coding scheme were not identical, the present findings resemble those obtained for an English ICU by Edworthy and colleagues (2013) with respect to the occurrence rates of individual activities: These authors found “Observation”, “Staff talking” and “Preparing and administering drugs” to be among the most frequent activities.

Arguably, the *individual* tasks alone may not usually exploit the nurses’ perceptual and cognitive resources to a degree that endangers alarm detection. However, the need to coordinate several tasks simultaneously (multitasking) or sequentially (task-switching) may considerably increase cognitive load. In the present study considerable degrees of multitasking and task switching were observed. Multitasking occurred in 24% of the time spent on the manual tasks - up to 30% of the tasks with perceptual and cognitive components and up to 53% of the predominantly motoric manual tasks. Moreover, nurses switched frequently between activities - between approximately once every minute and once every three minutes (see also Cornell et al., 2010). Performance decrements are to be expected for multitasking conditions, because processing resources have to be divided between tasks – even though there may be differences depending on the overlap between, for instance, processes or modalities involved (e.g. Wickens, 2008). Task-switching in clinical settings is mostly discussed as a consequence of interruptions (e.g. Chisholm et al., 2000; Chisholm, Dornfeld, Nelson, & Cordell, 2001; Grundgeiger, Sanderson, MacDougall, & Venkatesh, 2010; Li et al., 2012; Walter et al., 2014; Westbrook, Duffield, Li, & Creswick, 2011). Here, turning to a new task is assumed to load on working memory, because the intention to complete the interrupted task (together with the steps already performed) must be kept active (e.g. Parker & Coiera, 2000; see also Grundgeiger & Sanderson, 2009). However, switching from one task to another may be regarded demanding even if there is no need to resume a previous activity (e.g. when two activities are part of the same overall task), in part because of the constant need to (re-) activate current cues and task-rules (Cornell et al., 2010; see also Kiesel et al., 2010). Similar demands on cognitive control may result from the fact that attention was frequently drawn by approaching colleagues and relevant alarms (at least once every 3 minutes), while at the same time irrelevant alarms and other noise distracters had to be filtered out (more than once every 2 minutes). Note that the study did not aim at determining whether an alarm elicited an intervention, so as to get an impression on the number of potentially undetected alarms. Because of the temporal delay between individual alarms and the associated responses, this should be impossible to infer from observational data, alone.

To summarize, the present study provides examples of naturally occurring durations and frequencies of individual tasks, proportions of multitasking between individual tasks, typical switching rates and rates of task-relevant and task-irrelevant disruptive events on a German ICU. These factors may be assumed to contribute to the nurses' perceptual and cognitive load. This information may be used to more closely match conditions in future laboratory studies to those in the field in order to investigate the risk of failing to detect an alarm under controlled conditions (for a similar approach see Edworthy et al., 2013). Insights into the cognitive restrictions to alarm detection under maximally realistic situations would provide important information with regard to the implementation of medical devices in the field.

Acknowledgements

We would like to thank all the nurses of the surgical intensive care unit of the University Hospital Bonn for their collaboration, Tobias Metzger for his assistance in setting up and testing the observational protocol and Klaus Liebl for developing and programming the observation tool.

References

Berg, L.M., Ehrenberg, A., Florin, J., Östergren, J., & Göransson, K.E. (2011). An observational study of activities and multitasking performed by clinicians in two Swedish emergency departments. *European Journal of Emergency Medicine, 19,* 246-251.

Berg, L.M., Källberg, A.S., Göransson, K.E., Östergren, J., Florin, J., & Ehrenberg, A. (2013). Interruptions in emergency department work: an observational and interview study. *British Medical Journal: Quality and Safety, 22*, 656-663.

Bortz, J., & Lienert, G.A. (2008). *Kurzgefasste Statistik für die kinische Forschung. Leitfaden für die verteilungsfreie Analyse kleiner Stichproben (3. aktualisierte und bearbeitete Auflage)*. Berlin: Springer.

Cartwright-Finch, U., & Lavie, N. (2007). The role of perceptual load in inattentional blindness. *Cognition, 102*, 321-340.

Chisholm, C.D., Collison, E.K., Nelson, D.R., & Cordell, W.H. (2000). Emergency department workplace interruptions: are emergency physicians "interrupt-driven" and "multitasking"? *Academic Emergency Medicine, 7*, 1239-1243.

Chisholm, C.D., Dornfeld, A.M., Nelson, D.R., & Cordell, W.H. (2001). Work interrupted: a comparison of workplace interruptions in emergency departments and primary care offices. *Annals of Emergency Medicine, 38*, 146-151.

Cornell, P., Herrin-Griffith, D., Keim, C., Petschonek, S., Sanders, A.M., D'Mello, S., Shepherd, G. (2010). Transforming nursing workflow, part 1: the chaotic nature of nurse activities. *The Journal of Nursing Administration, 40*, 366-373.

Dehais, F., Causse, M., Vachon, F., Regis, N., Menant, E., & Tremblay, S. (2014). Failure to detect critical auditory alerts in the cockpit: evidence for inattentional deafness. *Human Factors, 56*, 631-644.

Edworthy, J., Meredith, C., Hellier, E., & Rose, D. (2013). Learning medical alarms whilst performing other tasks. *Ergonomics, 56*, 1400-1417.

Fougnie, D., & Marois, R. (2007). Executive working memory load induces inattentional blindness. *Psychonomic Bulletin and Review, 14*, 142-147.

Grundgeiger, T., & Sanderson, P. (2009). Interruptions in healthcare: theoretical views. *International Journal of Medical Informatics, 78*, 293-307.

Grundgeiger, T., Sanderson, P., MacDougall, H.G., & Venkatesh, B. (2010). Interruption management in the intensive care unit: Predicting resumption times and assessing distributed support. *Journal of Experimental Psychology: Applied, 16*, 317-334.

Kalisch, B.J., & Aebersold, M. (2010). Interruptions and multitasking in nursing care. *Joint Commission Journal on Quality and Patient Safety, 36*, 126-132.

Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A.M., & Koch, I. (2010). Control and Interference in Task Switching - A Review. *Psychological Bulletin, 136*, 846-874.

Klink, K.B., Schulte, M., & Meis, M. (2012). Measuring listening effort in the field of audiology - a literature review of methods, part 1. Messungen von Höranstrengung im Bereich der Audiologie - eine literaturgestützte Methodenübersicht, Teil 1. *Zeitschrift für Audiologie, 51*, 60-67.

Konkani, A., & Oakley, B. (2012). Noise in hospital intensive care units-a critical review of a critical topic. *Journal of Critical Care, 27*, p522.e1-522.e9.

Konstantinou, N., & Lavie, N. (2013). Dissociable roles of different types of working memory load in visual detection. *Journal of Experimental Psychology: Human Perception and Performance, 39*, 919-924.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Li, S.Y., Magrabi, F., & Coiera, E. (2012). A systematic review of the psychological literature on interruption and its patient safety implications. *Journal of the American Medical Informatics Association, 19*, 6-12.

Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences, 7*, 134-140.

Parker, J., & Coiera, E. (2000). Improving clinical communication: a view from psychology. *Journal of the American Medical Informatics Association, 7*, 453-461.

Simons, D.J., & Chabris, C.F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059-1074.
Tucker, A.L., & Spear, S.J. (2006). Operational failures and interruptions in hospital nursing. *Health Services Research, 41*, 643-662.
Walter, S.R., Li, L., Dunsmuir, W.T.M., & Westbrook, J.I. (2014). Managing competing demands through task-switching and multitasking: a multi-setting observational study of 200 clinicians over 1000 hours. British Medical Journal: Quality and Safety, 23, 231-241.
Westbrook, J.I. (2014). Interruptions and multi-tasking: moving the research agenda in new directions. *British Medical Journal: Quality and Safety, 23*, 877-879.
Westbrook, J.I., Duffield, C., Li, L., & Creswick, N.J. (2011). How much time do nurses have for patients? A longitudinal study quantifying hospital nurses' patterns of task time distribution and interactions with health professionals. *BMC Health Services Research, 11*, 319.
Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50*, 449-455.

Human factor guidelines and workload in CCTV design

David de Bruijn[1], Jorrit Jansen[2], Dick Lenior[3], Ruud Pikaar[4], & Kirsten Schreibers[1]
[1]INTERGO, [2]vhp, [3]HAN University of Applied Science, [4]Ergos, The Netherlands

Abstract

Camera surveillance (Closed Circuit Television, CCTV), is increasingly used for remote process control or surveillance. A CCTV-system is a man-machine system consisting of a chain starting at an observed situation, via cameras, and transmission towards a workstation where operator tasks and displays come together. CCTV-systems are used for traffic control, operation and surveillance of objects like bridges or locks, surveillance and security or process-control. Situations can be found that clearly lack good guidelines on human factors. The problems seen are not only related to technology such as unsharp images but also task-related, for example: a high number of images with little points of orientation and a vague target description. Such situations motivated us to study CCTV from a human factors perspective. This article describes what guidelines were developed and how this was done. As a special project, guidelines were sought for the question: how many screens can one operator handle? A new concept 'scenes' was proposed and evaluated in practice.

The guidelines project

An international guideline on CCTV exists: NEN-EN50132-2012; but it hardly addresses working with CCTV. That is why a consortium (Ergos Human factors engineering, Intergo human factors & ergonomics, VHP Human performance, University Arnhem Nijmegen) set off to develop human factors guidelines. The over-all research question was:

What should an operator be able to see, detect or read from CCTV images? What requirements should be met by the CCTV system?

Project phases were: Orientation and literature review; Situation-analyses / case-studies to describe specific CCTV tasks; Research on image quality, both explorative and experimental; Development of human factors guidelines; and Construction and evaluation of a theory on scenes as a measure for workload.

The project was funded by 13 project partners from government and several organisations from industry and security. This article aims to present some important research topics in the guidelines project. It is the intention to apply the guidelines in practice.

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Literature review

Schreibers et al. (2012) found approx. 40 publications on human factors (HF) in CCTV. An additional review was done for the theory on scenes. Highlights of the review results are:

- CCTV control centres sometimes have large video screens. According to the standard ISO-11064 they should be used mainly for information shared between operators. But in practice they are often used individually. Independent guidelines were not found on recommended use of video screens.
- The number of images an operator can follow simultaneously varies strongly with tasks and complexity as found in practice. Some sources suggest a maximum of 12 to 16 images for a simple task. Others sources mention 50 images. Clear definitions of task complexity were not found however. Because there was a great interest in this issue from our clients it was decided to start a new sub-project on it. It will be described in more detail below.
- Most publications contain lists of important issues rather than guidelines; a level of evidence is often lacking.

A separate line of research is aimed at the detection of suspect behaviour (e.g., Burghouts, 2010). Although this is interesting, its focus is less on HF. Besides, image technology was outside the scope of the guideline project.

Situation analyses for the guidelines

Following the literature review, eight situation analyses were performed in existing CCTV situations. The aim was to collect the CCTV tasks, note the CCTV technique used and the relations between these. The analyses were performed at: the bridge of a dredging vessel, a traffic control centre, two security centres for controlled entrance, a lock control, two station surveillance rooms, and a remote barge jetty supervision. Each situation was analysed systematically by HF professionals, using standardised protocols. Two different types of tasks could be distinguished:

- Type 1 – no trigger
  If there is no attention signal or sensor, then the main task is monitoring the images for unusual or specified events. If an event occurs, the operator analyses it. He then performs a follow-up task, or continues watching with a next CCTV task in the order observe – detect – recognize – identify. See figures 1 and 2. This pattern is frequently used in CCTV literature.
- Type 2 – trigger
  If there is a signal, an alarm or a telephone call, the operator selects an image to detect what is going on or starts an action himself.

An additional project was aimed at gathering guidelines about wide-angle lenses. This concerned mainly the distortions occurring at wide-angle lenses and the trade-off between the advantages of a wider field of view.

	
Observe	Detect

Figures 1 and 2. Different levels of detail in images for two observation tasks.

The findings from situation analyses and additional projects led to a set of guidelines (Pikaar, 2015). The guidelines also go into the design process and cover the following chapters:

1. Project Ergonomics – the engineering process for integrating HF in the design of CCTV-related control centres.
2. CCTV System Description – describing the elements of the CCTV system and how to achieve a complete functional description including a detailed task- and information analyses.
3. Field equipment – Cameras, camera positions, image pre-processing.
4. Control room layout and workplace design – Requirements for control room layout, workstation design and visual anthropometry for situations with a large number of information displays (i.e. hardware issues).
5. Image presentation and interaction design – for CCTV systems, the majority of information concerns images.

Perceived CCTV image quality

Experimental research has been performed on the judgement of the perceived image quality. It consisted of a set of experiments in which CCTV mediated vision was compared with direct (non-mediated) vision. The results of the experiments are published in Bennis, et al. (2014). Several test methods exist for the quality of CCTV systems such as the Rotakin test (Aldridge & Gilbert, 1996) and Vidilabs (Damjanovski, 2005). See figure 3. In this project the aim was to find a test method that could be applied in practice and that rules out subjective influences such as expectation and visual performance as much as possible. The Rotakin manikin and the Vidilabs test chart where evaluated. In addition, the Landolt C visual acuity test chart which is used by opticians to measure visual acuity, was used in the comparison. The main goal here was image judgement.

Figure 3. Rotakin test, Vidilabs test chart en Landolt C visual acuity test

The most important findings were:

The Rotakin test clearly does not provide sufficient unique distinctive characteristics to be deployed as a valid test system. Subjects appear to use ‘recognition tricks’ instead of giving a real judgement on perceived image quality.

The results from the Vidilabs test chart are difficult to interpret. The test is difficult to apply in practice because of the complexity and the sensitivity for disturbing factors.

The Landolt C test provides clear outcomes. It is a useful method for measuring the visibility of critical detail (related to viewing distance). Moreover, some important other factors involved in perceiving image quality, such as luminance/contrast and the influence of resolution, are measurable. Another important advantage is that it includes a measurement of visual acuity.

A new concept: scenes

There are questions about the number of images that can be effectively monitored. For instance Gill et al. (2005) found that operators switch to a reactive mode with more than 100 camera views. Other literature (e.g. Wallace et al., 1997) seems to suggest a maximum of 16 images. In the literature, displays or images presented to the operator are counted as a measure for workload.

To address this issue, visual information has priority and should be placed within the optimal visual range. This visual range is determined by the direction of viewing, angle of view and viewing distance. The guidelines describe a primary work area and a work area for secondary displays. Figure 4 illustrates the areas. The guidelines describe a ‘pixel area’ with a limited number of pixels, on the condition that the operator should not be able to see individual pixels. The pixel area is a general limit to the workload of the operator.

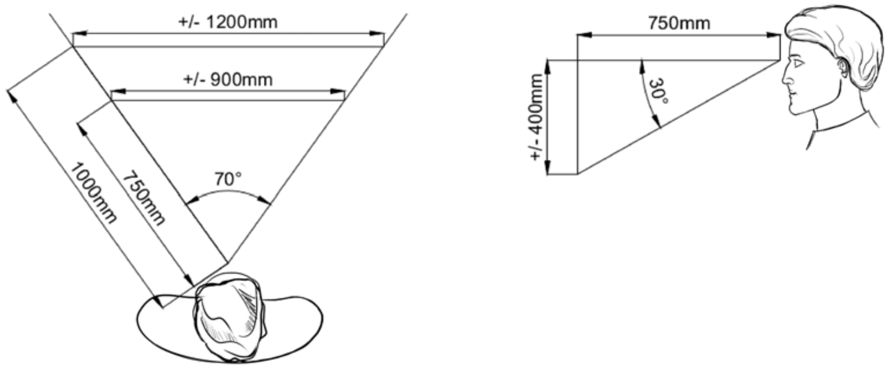


Figure 4. Illustration of the areas

A second limit is the number of ‘sets of information’ that is presented in the pixel area. For the development of the HF guidelines, the concept of scenes was introduced to address CCTV task complexity.

A scene is defined as a logical and meaningful set of visual information, to be monitored with a specific aim. The operator task determines the composition of a scene. For example: for tunnel safety monitoring a series of images representing one traffic lane could be defined as one scene. For a monitoring task, the operator could handle four tunnel tubes, i.e. four scenes. If an incident occurs, the task and thus the scene changes. Then, the operator task requires more detailed images of the incident area. For serious incidents, workload may become high and a colleague might be asked to monitor the remaining tunnel tubes. The hypothesis behind the concept of scenes is that the common goal or common area that images share, should result into less effort in monitoring than totally unrelated images. So scenes are a task-based rather than technique-based concept (De Bruijn & Burggraaf, 2015).

From scenes to workload: how much can an operator handle?

For establishing an operators’ workload, the task-based concept of scenes was elaborated. The situation analyses gave insight into CCTV-tasks. For a Type 1 task without trigger, it is supposed that the operator successively monitors a number of scenes. For certain scenes a ‘target’ may be detected and a subsequent task or ‘follow-up action’ is executed. See figure 5.

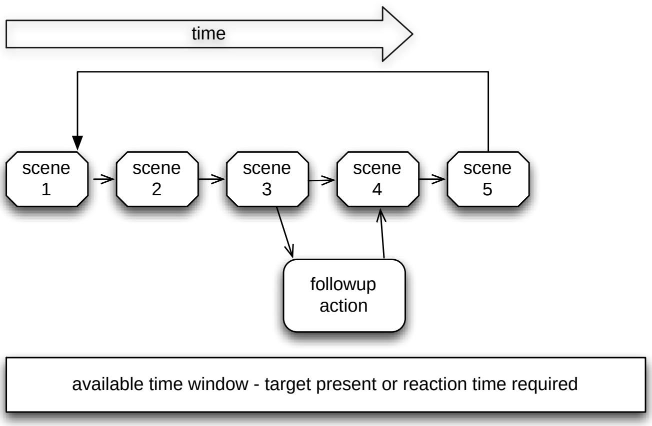


Figure 5. The monitoring cycle. Five scenes are monitored successively. For scene 3 a subsequent task is needed. Then the operator resumes monitoring.

The length of the monitoring cycle should fit within the time window in which a reaction from the operator is expected. The time window is determined by:

- the time the target is in sight. A pick pocket for instance can be recognised by his behaviour during a short period. A car thief often needs much more time and this leaves the operator more time to recognise him.
- the required reaction time. The pick pocket will vanish in the crowd in just a few seconds but more time is available if a traffic accident causes the traffic to congest and assistance is required.

Workload was operationalized as viewing time in seconds. To decide between high or lower workload, a threshold value is essential. The threshold for workload is given by the time available: the time window (see figure 4). For a Type 2 task with a trigger, the follow-up action will probably determine the time window and take most of the time. The operator can only monitor for ‘triggers’ between the subsequent actions or perhaps during the actions, if these do not require continuous attention.

Testing scenes in practice: some findings

The concept of scenes was studied in three cases: a traffic centre, a ships lock, and a railway station surveillance centre. In the traffic centre and the lock, sets of images were found that were used as a single point of attention focus during the monitoring tasks. For instance in the lock, all images of the lock are monitored if there are several ships to fit in the lock (see figure 5, in the dotted rectangle). The view of the lock consists of eight images and is regarded by the operator as being one scene for monitoring. When the operator opens a door he will focus on one image of the opening door and may even place this image on a separate display (solid rectangle in figure 6). This single image then forms one scene. The same principle of following the whole process (several images, one scene) versus inspection of a small area (one image, one scene) is found in the traffic centre.

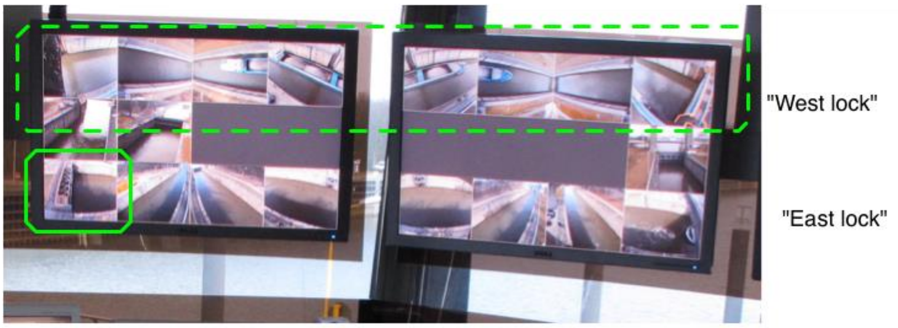


Figure 6. Example images of two locks. The operator monitors the ships in the 'West lock' (dashed rectangle) and checks on the opening of a door in the 'East lock' (solid rectangle).

In the railway station surveillance centre, scenes consist mostly of just one image. See figure 7. The eight scenes are supervised for unusual events or terrorism one after another. The images display locations within one station that are not adjacent. The focus is on one station at a time.

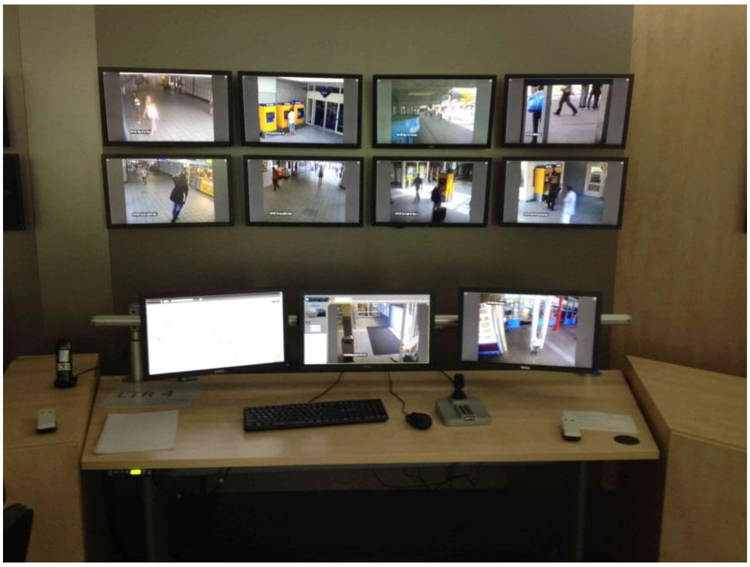


Figure 7. Railway station surveillance centre. The top eight images display separate locations in the station. Each image represents a separate scene.

An interesting finding is that viewing time per scene is in practice 2-3 seconds in all three cases (traffic, lock, railway), even for very complex images. Scene complexity was rated by the researchers with a list of complexity factors. The factors were derived from the literature (see table 1).

Table 1. Image complexity factors derived from the literature, as used for classifying scenes in the cases

Complexity factor	*Categories*	*Example*
'Behaviour' of the target	None, simple, complex	None = target absent. Flow of traffic=simple. Human behaviour=complex
Movement of the target	Biological, mechanical	Example: aggressive behaviour is biological. It is complex but can be recognised instantly.
Paced or non-paced	Paced, non-paced	Is continuous watching required? For instance, closing the barriers before opening a bridge, ensuring nobody passes them
Crowding		
• Similarity target and distractors	Dimensions may be: colour, direction, size, shape	A cork screw in a drawer full of cutlery may be hard to find, a red pencil isn't
• Number of distractors	None to many	some or many pieces of cutlery
• Distance target - distractors	Close or distant	The cork screw is between the cutlery or in a separate compartment
Image distortion	Glitter, noise, shadow, resolution etc.	A person in a scene is harder to detect between dark shadows
Significant detail in the image	Size of the detail to be found in relation to the total image	A child's bicycle on a railway crossing is harder to recognise than a truck

In practice, the operators appear to be hardly affected by the factors listed in Table 1. Due to their experience, even complex targets are recognized quickly and with little effort (e.g. recognizing a person in a crowd by the way they walk, even when only the outline of the person is visible yet, among many other persons). In order to understand this discrepancy between literature and practice, it was noticed that for instance research into suspicious behaviour or the influence of screen layout is mostly done with students or university staff as participants. Other research on threat-related intentional actions was done with naïve, paid, subjects. Indeed, an operator in the traffic centre confirmed that novice operators did not even come close to skilled operators in speed and accuracy. This suggests the image complexity factors may play a role in the learning time required for the job, and are not (or less) relevant to experienced operators.

Another explanation for the discrepancy is that some complexity factors found in experiments do reflect the task rather than the image. Examples of task factors are: the allowable number of missed targets and whether the task is concise with clear targets to be met, or vague with general instructions to prevent terrorism.

Follow-up tasks, such as controlling a lock door or a dynamic road information panel, have a much larger impact on workload (measured in time spent on the task) in the three cases than the actual viewing tasks. This is illustrated in the monitoring and control of a lock: detection of a situation (a ship has left the lock) takes seconds but guarding the closing lock door may take a minute.

The allowable number of scenes was estimated for the railway station surveillance centre, on the basis of the constant 2 - 3 seconds viewing time per scene. The task chosen was surveillance for crowding in the central hall area. Crowding can be expected during rush hours and the operator should alert service personnel within one minute. This means an allowable time window (see figure 5) of 60 seconds to view all scenes and contact personnel. The latter task will take approx. 20 seconds. This leaves 40 seconds for monitoring scenes. Viewing time was estimated to be 4 seconds, a bit longer than standard because the default scenes presented by the CCTV may be more dense in rush hours than usual. This should give the operator time to monitor 40 / 4 = 10 scenes. Although it is difficult to obtain exact figures, this suggests that the operator can handle the current 8 scenes per station.

Designing scenes

Two important factors in the design of scenes are:

- Choosing which images are the best representation of the task area. In a railway station it may prove necessary to have more than one camera for an area in order to avoid any blind spots. A minimum of images is sometimes not best: in the case of the traffic supervision centre, the images are overlapping each other slightly. This raises the likelihood of the operator seeing a decrease in distance between cars as early as possible.
- Selecting order and layout of scenes for optimal situation awareness (SA): correct perception of a situation, understanding it and being able to predict what is likely to happen. One guideline is to place the images or scenes in a logical order. The level of situation awareness required depends on the task. If the task is to detect unattended luggage and if each camera image has a unique code, then a low SA is required. Tracing a moving person in a complex environment requires a high SA however.

There was a wish from the project sponsors to work out the principle of the monitoring cycle into a visual representation. Two flow-charts are made, containing guidelines and choices to be made for estimating the allowable number of scenes in a particular situation.

Discussion

The experts from the sponsors (present in the workshops) judged scenes to be a useful concept for the monitoring of a large number of images, each in their own specific applications of CCTV. Although this is a first step, the idea of calculating the monitoring cycle was attractive.

An unexpected finding is that viewing time is a few seconds for each scene, and is hardly affected by complexity. This viewing time was obtained by asking the operators. They were also asked what images were making up the ‘total’ picture (scene). An interesting research method would be eye-tracking. This analysis technique may reveal different ways of viewing the images within a scene. For instance, how images are scanned, whether some images receive more attention, or any differences in viewing for each sub-task. Eye tracking may also reveal situations of high workload where scanning patterns change or become non-systematic.

Experienced operators were hardly 'affected' by the complexity factors. It is still unclear how consistent the complexity factors in the table are over a range of different situations. These factors are based on just a few experiments and cases, so a broader investigation may reveal new factors or a difference in the weight of factors. Another question is whether novice operators do actually recognise these factors during their training. Also, training takes longer if scenes involve more complexity factors.

The concept 'scene' has been explored in three situations in practice. Extension to other areas can refine the concept. It could also be approached experimentally by investigating whether scenes composed of several images take the same viewing time as scenes with a single image.

Acknowledgement

Copies of the Guidelines and the document on scenes can be requested via www.cctvguidelines.com. We ask a small favour in return: do give feedback on the material. The authors thank the professional colleagues of the Human Factors Research Group (vhp hp, INTERGO human factors & ergonomics, ErgoS Human Factors Engineering, and HAN University of Applied Sciences) for their valuable participation in the research project. This research would not have been possible without the funding by Saab-HITT Traffic, IHC Dredgers/IHC Beaver Dredgers, Nedap Security Management, Rijkswaterstaat, Rijksbelastingdienst, NS (Dutch Railways), ProRail ICT Services, Royal Haskoning/DHV, Total E&P, Vopak, Waterschap Hollandse Delta, DGOrganisatie Bedrijfsvoering Rijk, Province of North-Holland.

References

Aldridge, J. & Gilbert, C. (1996). *Performance testing of CCTV. Perimeter surveillance systems. Using the Rotakin Standard Test Target.* Police Scientific Development Branch. Publication 14/95. Home Office - Police policy directorate; Hertfordshire, United Kingdom.

Bennis, A., Landman, R. & Lenior, D. (2014). CCTV mediated observation versus non-mediated observation: investigating perceived image quality with different test systems. In O. Broberg, et.al (Eds.), Proceedings of the XI conference on Human Factors in Organizational Design and Management. Copenhagen, 2014.

Burghouts, G.J. & Marck, J.W. (2010). Reasoning about threats: from observables to situation assessment. *IEEE Transactions on System, Man and Cybernetics. Special issue on Pattern Recognition for Anti-terrorism Applications., 2010).*

De Bruijn, D.W., & Burggraaf, J. (2015). *The concept of Scenes: related to task complexity and workload.* Revision 1.2 May 18th 2015. Human Factors Research Group NL, p/a Enschede, The Netherlands: ErgoS Human Factors Engineering.

NEN-EN-50132-7 (2012). *Alarm systems - CCTV Surveillance systems for use in security applications.* Part 7: application guidelines.

Damjanovski, V. (2005). *CCTV Networking and digital technology.* Burlington: Elsevier Butterworth – Heineman.

Gill, M., Spriggs, A., Allen, J., Hemming, M., Jessiman, P, Kara, D., Kilworth, J., Little, R. & Swain, D. (2005). Control room operation: findings from control room observations. Home office - Research, Development and Statistics Directorate. Online report 14/05 (2005).

Pikaar, R.N. (editor) (2015). *Human Factors Guidelines for the design of CCTV-systems* (revision 1). Human Factors Research Group NL, p/a Enschede, The Netherlands: ErgoS Human Factors Engineering.

Schreibers, K.B.J., Landman, R.B. & Pikaar, R.N. (2012). *Human Factors of CCTV; Part 1. Technology and Literature review.* Human Factors Research Group NL, p/a Enschede, The Netherlands: ErgoS Human Factors Engineering.

Wallace, E., Diffley, C., Baines, E., Aldridge, J. (1997). Ergonomic design considerations for public area CCTV safety and security applications. In: Seppälä, P., Luopajärvi, T., Nygård, C., Mattila, M. (Eds.). From Experience to Innovation. Proceedings of the International Ergonomics Association congress. Sandridge, St Albans, United Kingdom.

Automation surprise looked at from a Demands-Resources Perspective

Karel Hurts & Robert Jan de Boer
Amsterdam University of Applied Sciences,
The Netherlands

Automation surprise (AS) is usually seen as a sign of the breakdown of pilot-aircraft interaction. In attempt to resolve several conflicting findings with respect to the precise relationship between pilot workload, degree of automation (DoA), and the frequency of experiencing AS, it was hypothesized that the average AS-rate (number of AS-occurrences per flight - or per unit time – and per pilot) depends on the specific way in which Elapsed Flight Duty Period (seen as a type of "demands") combines with DoA (seen as a type of "resources"), rather than on each of these two factors considered on their own. Specifically, the average AS-rate was expected to be higher for non-matching than for matching combinations (both being high or both being low) of DoA and Elapsed FDP. This hypothesis was based on psychological arousal theory, signal-detection theory, and general research findings pertaining to the development of automation trust during human interaction with automated systems. Data collected in a survey held among 200 airline pilots just failed to confirm the hypothesis. However, the average AS-rates that were observed were in the expected direction. In the discussion, the theoretical implications of this finding will be addressed.

Introduction

Automation surprise is a phrase that first appeared in the aviation literature in the 1990's (Woods et al., 1994; Sarter et al., 1997). Dekker (2009) defines automation surprises as those cases where:

a) "The automation does something …
b) ... without immediately preceding crew input ...
c) … related to the automation's action, …
d) … and in which that automation action is inconsistent with crew expectations."

Note that the discrepancy to which this definition refers may have been present already for a while before the pilot becomes aware of it. This is similar to the phenomena of inattentional blindness, automation-related complacency, and automation bias (De Boer, 2012; De Boer et al., 2014; Parasuraman & Manzey, 2010).

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

In the existing literature, automation surprise is often associated with loss of situation awareness under conditions of high cockpit automation (Operator's Guide to Human Factors in Aviation**,** 2014; Optimum Use of Automation, 2006). From this point of view, automation surprise is considered an undesirable phenomenon because of the risk of losing aircraft and flight control and, ultimately, the risk of operational safety hazards.

However, available research shows that the phenomenon of automation surprise (AS) cannot be explained or functionally understood in a simple way, involving only a single or a few factors. The following list of research findings illustrates the ambiguity that surrounds attempts to understand the relationship between amount of workload, degree of cockpit automation, and behavioural phenomena such as automation surprise, complacent pilot behaviour, and pilot situation awareness.

a) Complacent pilot behaviour (i.e., missing important signals from the environment and from the cockpit instruments due to inattention) may be associated with high workload (Parasuraman & Manzey 2010), but also with low workload (Sarter, 2008; Norman, 1990; Matthews & Desmond, 2001).
b) With *higher* degrees of automation, often poorer situation awareness is observed, but *superior* situation awareness has also been observed, compared to lower degrees of automation (Kaber & Endsley, 2004; Onnasch et al., 2014).
c) In a previously conducted AS-study (Hurts & De Boer, 2014) it was found that *higher amounts of* external workload are sometimes associated with *lower* (i.e., not-expected) frequencies of experiencing AS.
d) In the same study, it was found that degree of cockpit automation was not significantly correlated with the frequency of experiencing AS, despite the fact that higher degrees of automation seem to offer more opportunities for experiencing AS.

In an attempt to understand the seemingly conflicting findings regarding the relationship between degree of automation, amount of external workload, and the frequency of experiencing AS (see points c and d above), a different perspective on the nature and function of AS was developed. As will be seen below, this perspective is based on psychological arousal theory, as well as on signal detection theory. It is also based on existing research concerning the way in which pilot trust and mistrust in automation develops.

Problem statement and hypothesis

Step 1: the goal of optimizing psychological arousal

One theory that combines the notions of amount of external workload and degree of automation in a single construct is psychological arousal theory. From the research that has been devoted to this theory, it follows that the pilot does not just attempt to minimize his effective workload (or arousal level), but rather tries to *optimize* it

(Young & Stanton, 1997; Wilson & Rajan, 1995; Matthews & Desmond, 1997). In the present study, pilot arousal level is seen as being determined by the combination of current degree of cockpit automation - seen as a type of "resources" -, and Elapsed Flight Duty Period (FDP) – seen as a type of "demands". (In this article, degree of automation – or DoA - will be defined as *the complexity of the flight control mode*, see Table 1 for further details.) Specifically, if the determining factors are both high or both low (are matching), the arousal level can be considered to be optimal. Otherwise (if these factors are not matching), it can be said that there is overarousal or underarousal.

Usually, the pilot has no direct control over Elapsed FDP (i.e., the number of hours he/she has been working without interruption). Therefore, under conditions of over- or underarousal he/she can influence his current arousal level only by adjusting the current DoA (see step 3 below for the details).

Step 2: detecting an automation-pilot conflict as trigger for testing automation trust

On a different note, it is likely that DoA is also used by the pilot to calibrate his/her current level of trust in the cockpit automation. From the literature on the importance of trust in semi-automated working environments (see, e.g., Bass & Pritchett, 2008), it can be expected that automation trust must occasionally be tested in order to build and maintain it, or, if unavoidable and necessary, to (temporarily) reduce it. For example, automation distrust may arise due to the automation being intransparent to the pilot. This may, in turn, cause him/her to (temporarily) reduce DoA. It is proposed that an obvious trigger for conducting such tests is formed by the detection and conscious experience of a conflict between expected and actual automation behaviour. Specifically, during a test phase the pilot attempts to identify the cause of the conflict, and, if necessary, adjusts the current DoA accordingly (i.e., choose more automation or less automation, depending on the relative amount of trust the pilot has in him-/herself as pilot and the automation).

Step 3: increasing the importance of arousal considerations during the test phase

It is at this point during the test phase that arousal considerations come into play. Obviously, these considerations have to be somehow reconciled with performance-related and safety-related considerations. It is proposed that a natural way for the pilot to increase the importance of arousal considerations under conditions of over- or underarousal is *to lower the treshold for detecting a conflict between expected and actual automation behaviour*[2]. This proposition is based on the general logic of

---

[2] "Detecting a conflict" should be compared to detecting a signal, as described by signal detection theory (SDT). As is the case for signals in SDT, it is assumed that conflicts occur in a noisy environment, containing many other types of events that may suggest that there is a conflict. Though the pilot cannot perceive a conflict directly, he/she can statistically weigh the evidence supporting the existence of a "true" conflict. As in SDT, the pilot can make two types of error regarding the detection of a conflict. A *false alarm* occurs if the pilot only *believes* that there is a conflict, whereas, in fact, there is none. For example, the pilot mistakes some piece of – innocent - automation status information for an alert

signal-detection theory as follows: as a result of lowering this threshold, more conflicts will be detected (on the average) in a fixed time period, compared to when the threshold remains unchanged. This, in turn, has the effect that more tests will be conducted in the same time period, and, eventually, that more opportunities will be created for changing the current DoA.

Implications of the three-step process

Note that there is no guarantee that this strategy for influencing DoA will always result in an improvement of the pilot's arousal level. Nonetheless, in cases of over- or underarousal lowering the conflict detection threshold seems to be an effective strategy for influencing the probability that pilot arousal level will shift in the direction of optimal arousal. This expectation is supported by studies that show that problematic pilot-automation interactions may occur if DoA remains - too - high during an extended period of flight time, as illustrated by the phenomenon of *automation overreliance* (also referred to as automation bias or complacent pilot behaviour). In terms of our model, overreliance becomes a risk if a high DoA is combined with a low Elapsed FDP (signaling underarousal). The reversed combination of a high Elapsed FDP with a low DoA (signaling overarousal) is also known to be associated with problematic interactions, as illustrated by the phenomena of *automation underreliance,* automation disuse or non-conforming pilot behaviour (Parasuraman, 1997).

Finally, note that the detection of a conflict between expected and actual automation behaviour may also have other causes and implications than those discussed above. For example, learning by the pilot from previously resolved conflicts is likely to affect future pilot-automation interaction. Also, it is likely that the detection of a conflict is influenced by the size of the discrepancy, as well as by the frequency with which conflicts have occurred in the past (see also De Boer, 2012). However, in this article, the role played by these additional factors will not be further discussed.

Assuming that each detected conflict results in an automation surprise (AS), the following hypothesis can be derived from the previous discussion:

Hypothesis

If DoA at the time of the last AS and the Elapsed FDP at the same time do not match, the frequency of experiencing AS is higher compared to when they match (interaction between DoA and and Elapsed FDP with respect to AS-frequency).

In order to test this hypothesis, the survey data that were described in the study of Hurts and De Boer (2014) were re-analysed. It was assumed that Elapsed FDP

---

signaling unexpected danger. A *miss* occurs if the pilot ignores or somehow fails to detect a "true" conflict. Such errors might be due to inattentional blindness.

would (partly) reflect the build-up of pilot fatigue, which, according to the literature (Stanton & Young, 2000), can be considered to be one aspect of mental workload.

Method

Participants and procedures

For this study, the data were used that were collected in the 22-question survey described by Hurts and De Boer (2014). Twohundred pilots participated in the survey, most of whom were recruited through Crew Center of KLM, the VNV (Dutch Association of Airline Pilots), and National Aerospace Laboratory NLR. Most respondents filled in the web version of the survey. It took them from 20 to 30 minutes. Though the survey was filled in anonymously, a few questions were included in order to verify that the respondents were really airline pilots (e.g., respondents were asked about the various aircraft they had been flying and they had to indicate how they had been approached for their participation). An automation surprise (AS) was briefly explained to the participants in terms of a few typical pilot reactions to automation behaviour. Participants were required to describe their last AS, as well as provide information about several (predefined) accompanying circumstances. Only a subset of the twenty-two questions were of direct interest to the present study, as will be explained below.

Design

Dependent variable
The frequency with which an AS occurred for any participant was measured in two different ways:

AS-frequency score 1 (flight-based frequency measure): this score (one per participant) was defined as the fraction of the total number of operated flights on which an AS was experienced by a participant. It was estimated on the basis of the answers given by the participants to the survey question "How many flights ago was your last automation surprise?", as follows[3]:

---

[3] *Number of flights since last AS* can be seen as somehow estimating the *period* of the frequency with which AS is experienced during a flight (here, *period* is defined as the number of consecutive non-AS-flights separating two AS-flights). Let's call this estimator *#NASF*. However, because the time at which the survey was filled in was assumingly chosen at random by the pilot and/or researcher, the period must, on the average, have passed only for 50% at the time of the measurement of *#NASF*. Therefore, and because frequency is the inverse of the period, the term $2\times(\#NASF)$ appears in the denominator of the formula. The constant 0.5 is added to this term as a means to correct for the fact that the participants most likely *included* the last AS-flight in their count of the number of flights since their last AS-flight. This has resulted in overestimations of the values for period. This "counting error" can occur only once in a period: hence, the value of $(2\times(0.5))$ in the denominator of the formula.

$$f(AS\_1) = \frac{1}{2 \times ((Number\ of\ flights\ since\ last\ AS) - 0.5)}$$

AS-frequency score 2 (time-based frequency measure): this score (again, one per participant) was defined as the average number of AS-flights experienced by a participant per month. It was computed by multiplying the outcome of the above-mentioned formula by the answer given by the pilot to the question: "How many flight do you operate in a month, on the average?".

It should be noted that both AS-frequency scores yield slight overestimations of the "real" AS-rate because no scores could be computed for participants who never had experienced an AS (see under *Results* for more details).

Independent variables

Degree of automation (DoA) was assessed using a seven-point scale, with scoring categories ranging from "No automation" to "Full automation" (see Table 1). The categories were designed in a post-hoc fashion by an experienced flight instructor based on the open-answers given by the participant to the question "What flight mode were you in at the moment of the last automation surprise?" (one score was computed for each participant).

Elapsed Flight Duty Period (FDP) refers to the number of hours the participant had been working without interruption at the time of his/her last AS. (Participants were asked to choose one among several numerical time intervals.)

Analyses

Data were analysed using multiple regression analyses in which DoA and Elapsed FDP were entered as predictors, and frequency of experiencing AS was used as dependent variable.

The interaction between DoA and Elapsed FDP was entered into the regression analyses as a third predictor - and was used for testing the hypothesis. This was done as follows:

a) Elapsed FDP was first dichotomized, giving the values high (1) and low (-1), depending on whether the participant's score was above or below the average Elapsed FDP of 5.5 hours. This was done in order to end up with an easy-to-interpret interaction.
b) For each participant, the product between DoA (in mean-centered form) and Elapsed FDP was computed and this term was entered as a separate predictor in the regression model. Note that the product term was an ordinal scale variable with both negative values (corresponding to non-matching combinations) and positive values (corresponding to matching combinations).

c) The hypothesis would be confirmed if the effect of DoA × Elapsed FDP on the dependent variable was statistically significant ($p < 0.05$) *and* if *higher* product values (corresponding to matching combinations of DoA and Elapsed FDP) were associated with *lower* AS-frequency values (on the average) than lower product values (non-matching combinations).

Table 1. Seven scoring categories, and their associated frequencies, for measuring complexity of flight control mode (degree of automation). Based on answers to open survey question and measured on an ordinal scale. 1= lowest, 7 = highest degree of automation.

	Complexity of flight control mode (degree of automation)	*% of valid*
1	FD ON, MANUAL FLIGHT	4.8
2	AP OFF, AT ON , FD ON MANUAL FLIGHT	0.6
3	AP ON, AT OFF, MANUAL SELECT	0.6
4	AP/AT ON, MANUAL SELECT (HDG, VOR/LOC, VS)	18.7
5	AP/AT ON, FMS GUIDANCE SINGLE (HOR./VERT.)	6.6
6	AP/AT ON, FMS GUIDANCE DUAL, APPR. MODE	66.9
7	AUTOLAND	1.8
Total valid		100.0

Results

Some demographics

Of all respondents, 96% was male, 54% was in the rank of captain, and 42% was in the rank of first officer (the balance is in the rank of second officer). With regard to aircraft type currently operated, respondents mentioned Boeing 737NG, Airbus A330, Boeing 777, Embraer 170/190, and Fokker 70/100 as the aircraft types flown most frequently. This reflects the fact that most respondents were employed by KLM, the fleet of which is primarily composed of the above-mentioned planes.

The average age of the respondents was 38 years, with a range from 23 to 58 years, sd = 9.63 years. Moreover, the mean value for amount of flying experience was 8867 hr, sd = 5480 hr, with a range from 750 hr to 27500 hr. Finally, the average number of flights per month was 22.8, with a range from 3 to 43 flights, sd = 15.09 flights.

In the analyses mentioned below, DoA was treated as an interval-level variable, even though, strictly speaking, it was only an ordinal scale variable.

Frequency of experiencing AS

The frequency of experiencing AS could only be computed for 186 (93%) respondents. These were the respondents who had indicated that they had at least one AS-experience. Therefore, both frequency scores provided slight overestimations of the "real" frequency with which AS's occurred. The average value for AS-frequency score 1 was 0.08 - or 8% AS-flights -, median = 0.03, sd = 0.13. This was based on an average value for number of flights since the last AS of 71, median = 20, sd = 170. The average value for AS-frequency score 2 was 1.44 flights per month, median = 0.40, sd = 2.95.

It turned out that neither AS-frequency score was normally distributed (both were skewed to the right). Therefore, in the analyses mentioned below, both scores were first subjected to a log10-transformation. After transformation, both transformed scores passed the K-S-normality test at a 0.05 significance level.

Interaction between DoA and Elapsed FDP

Figure 1 shows the average values of AS-frequency score 1, broken down by Elapsed FDP (high versus low) and by DoA. Regression analysis showed that the interaction between DoA and Elapsed FDP just failed to reach the level of significance, $t(159) = 1.84$, $p = 0.07$, but was in the expected direction. The two main effects (one for DoA, the other for Elapsed FDP[4]) were not significant either, $p > 0.10$.

Figure 1 suggests that the expected interaction was stronger for low degrees of automation. A post-hoc analysis revealed that, for the lowest four degrees of automation (less than or equal to the median rank of 4), the difference between the low and high Elapsed FDP-groups was significant, $F(1,38) = 5.00$, $p < 0.05$, partial $\eta^2 = 0.12$ (one-way analysis of variance).

The regression analysis belonging to AS-frequency score 2 revealed similar results: the interaction between DoA and Elapsed FDP was again almost significant, $t(158) = 1.93$, $p = 0.06$. The mean frequency scores generally followed the same (expected) pattern as in Figure 1.

[4] With regard to the main effect of Elapsed FDP, it should be noted that in Figure 1 the FDP-base rate was not taken into account (this is the probability with which the various original FDP-values occur, irrespective of whether or not an AS is observed). Follow-up analyses show that, after being corrected for this base rate, the average probability of experiencing an AS for the two highest FDP-intervals (8-11 hours, that is) becomes significantly higher than that for the remaining FDP-intervals (Hurts & De Boer, 2014).

It is concluded that, though the hypothesis could not be confirmed, the average values for AS-frequency scores 1 and 2 showed a trend in the expected direction: scores were lower if DoA and Elapsed FDP were matching than if they were not matching. The difference between matching and non-matching combinations was particularly salient for the four lowest degrees of automation.

Conclusions and Discussion

In this study, the hypothesis was tested that the frequency of automation surprises depends on the extent to which the current degree of cockpit automation (DoA), seen as a type of "resources", matches the current value for Elapsed FDP, seen as a type of "demands". It turned out that the average frequency of experiencing an automation surprise (AS) was lower if DoA (assessed at the time of the last AS) matched Elapsed FDP (assessed at the same time), compared to a non-matching combination. Though this effect was expected, it just failed to reach the level of statistical significance.

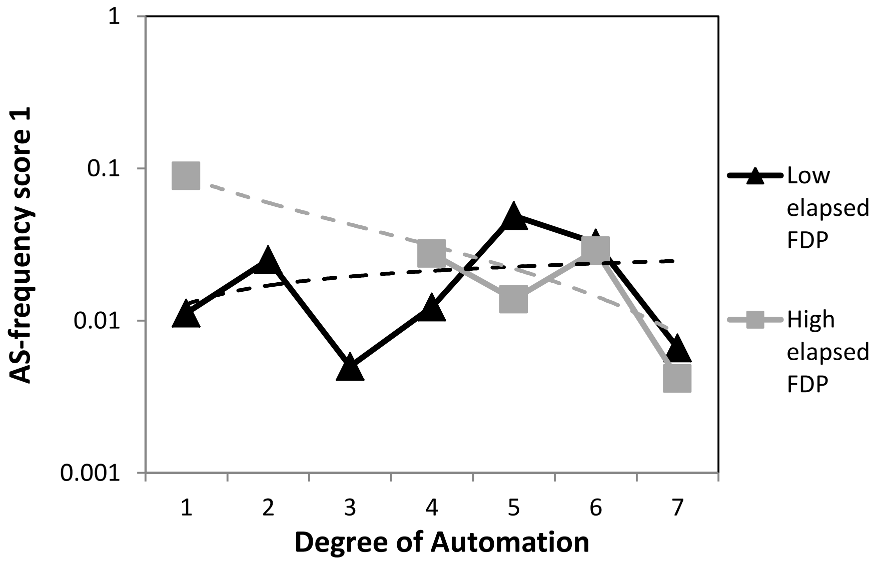


Figure 1. Average values for AS-frequency score 1, broken down by DoA and Elapsed FDP. Dashed lines represent best-fitting regression lines. Y-axis values deliberately shown on a logarithmic (base 10) scale.

The absence of any difference in average AS-frequency between low Elapsed FDP-pilots and high Elapsed FDP-pilots under conditions of high cockpit automation needs explanation. Perhaps, on some short-haul flights (i.e., flights with a duration of less than 6 hours – precisely the duration that corresponds to shorter-than-average values for Elapsed FDP), there are operational constraints that require the pilot to continuously fly with high cockpit automation. In that case, and following the rationale of this study's hypothesis, it is not likely that the pilot will detect and resolve more automation-pilot discrepancies than under conditions of low automation. This would explain the pattern of results observed in Figure 1. However, this post-hoc explanation should be treated with care and further research is needed to investigate it.

Severity of AS-consequences

In terms of signal-detection theory, lowering the threshold for detecting a conflict can be expected to result (in the long run) in more false alarms: cases where it is incorrectly assumed by the pilot that a conflict has been detected (more liberal response bias). Therefore, the strategy of lowering the confict detection threshold is expected to generate less severe AS-consequences (non-matching combinations of DoA and Elapsed FDP), on the average, compared to the situation where the threshold is higher (matching combinations, more conservative response bias). This latter expectation was tested in a post-hoc analysis of the statistical interaction between DoA and Elapsed FDP with respect to the self-reported severity of the consequences of the last AS. *Severity of AS-consequences* was assessed using a six-point scale, with scoring categories ranging from "No consequences" to "Damaged aircraft". In other respects, the expectation was tested in a way similar to that used for testing the main hypothesis of this article.

Figure 2 shows the average values of Severity of AS-consequences, broken down by Elapsed FDP (high versus low) and by DoA. It turned out that the interaction between DoA and Elapsed FDP was not significant, $p > 0.10$. As can be seen, the contrast between high Elapsed FDP and low Elapsed FDP was only in the expected direction for the highest degree of automation (higher Severity of AS-consequences-scores for the high Elapsed FDP-group). The two main effects (one for DoA, the other for Elapsed FDP) were not statistically significant either, $p > 0.10$.

It is concluded that the data on Severity of AS-consequences do not provide evidence for the expectation that the effects of DoA and Elapsed FDP can be understood in terms of more or less incorrectly detected conflicts (false alarms) - depending on whether the response bias for detecting conflicts took on a more liberal or a more conservative value, respectively.

Implications

Regardless of the fact that the findings of this study do not allow very strong conclusions, future researchers should not discard more traditional ways of looking at AS. For example, it is still likely (as others have suggested) that experiencing AS is a sign of a vulnerability in automation-pilot interaction. At the same time, researchers are encouraged to investigate more fully the possibility of additional purposes being served by the experience of AS – an issue that has only been touched upon in this study.

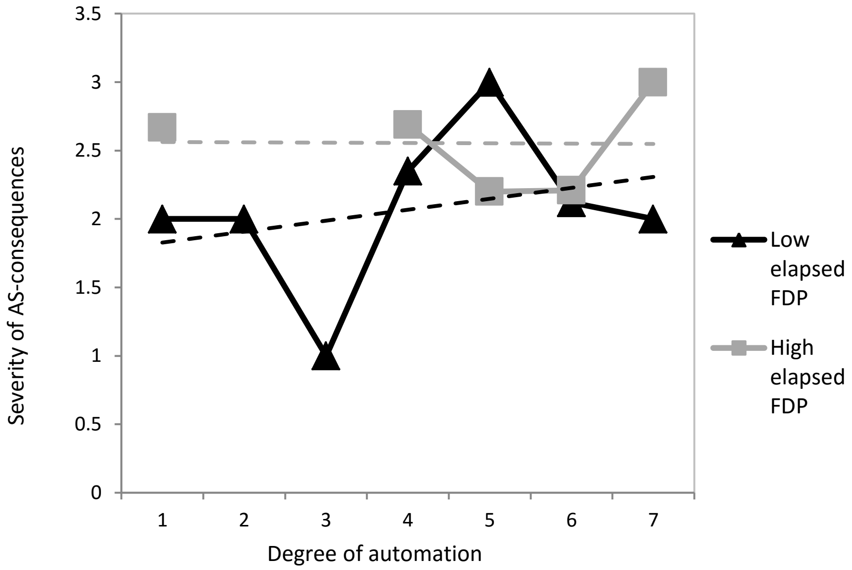


Figure 2. Average values for Severity of AS-consequences, broken down by DoA and Elapsed FDP. Dashed lines represent best-fitting regression lines.

In addition to attempting to validate the central ideas of this study under different and better controlled circumstances, future research should address the following questions:

1. Will the main hypothesis be confirmed if the "demands" affecting the pilot (now assessed by means of Elapsed FDP) are assessed in different ways?
2. How precisely does automation surprise affect the current mode of cooperation between pilot and automation?
3. How does learning from previously explained and resolved conflicts affect future pilot-automation interaction?
4. Are there other ways (besides lowering the conflict detection threshold under conditions of over- or underarousal) in which the pilot can improve his arousal level?
5. What role will automation surprise play during pilot-automation interaction if it is considered as an experience with variable intensity (i.e., a pilot can be more or less surprised about automation behaviour)?

References

Bass, E.J. & Pritchett, A.R. (2008). Human-automated judge learning: a methodology for examining human interaction with information analysis automation. IEEE *Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 38*, 759-775.

De Boer (2012). *Seneca's Error: An Affective Model of Cognitive Resistance*. Doctoral thesis. Delft, The Netherlands: TU Delft.

De Boer, R.J., Heems, W., & Hurts, K. (2014). The duration of automation bias in a realistic setting. *The International Journal of Aviation Psychology, 24*, 287-299.

Dekker, S. (2009). *Report of the flight crew human factors investigation conducted for the Dutch safety board into the accident of TK1951, Boeing 737-800 near Amsterdam Schiphol Airport, February 25, 2009*. Lund: Lund University, School of Aviation.

Hurts, K. & De Boer, R.J. (2014). What's it doing now? Results of a survey into automation surprise. In A. Droog (ed.), *Proceedings of 31st EAAP Conference* (pp. 197 – 210). Valetta, Malta: EAAP.

Kaber, D.B. & Endsley, M.R. (2004).The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. *Theoretical Issues in Ergonomics Science, 5*, 113–153.

Matthews, G. & Desmond, P.A. (1997). Underload and performance impairment: evidence from studies of stress and simulated driving. In D. Harris (ed.), *Engineering Psychology and Cognitive Ergonomics* (pp. 355 – 361). Aldershot, UK: Ashgate.

Matthews, G. & Desmond, P.A. (2001). Stress and driving performance: Implications for design and training. In P. Hancock and P. Desmond (eds.), *Stress, Workload and Fatigue* (pp. 211-231). Mahwah: Erlbaum.

Norman, D. A. (1990). The 'problem' with automation: inappropriate feedback and interaction, not 'over-automation'. *Philosophical Transactions of the Royal Society of London B*, 327, 585 - 593.

Onnasch, L., Wickens, C.D., Li, H., & Manzey, D. (2014). Human performance consequences of stages and levels of automation: an integrated meta-analysis. *Human Factors*, 56, 476-488.

Operator's Guide to Human Factors in Aviation (2014). *Unexpected Events Training (OGHFA BN).* Briefing note. Downloaded on February 2, 2015, from www.skybrary.aero.

Optimum Use of Automation (2006). *Airbus Flight Operation.* Briefing Note. FOBN Reference : FLT_OPS – SOP – SEQ 02 – REV 03 – JUL. 2006. Airbus: Blagnac Cedex, France.

Parasuraman, R. (1997). Humans and automation: use, misuse, disuse, abuse. *Human Factors*, *39*, 230-253.

Parasuraman, R. & Manzey, D.H. (2010). Complacency and bias in human use of automation: an attentional integration. *Human Factors, 52*, 381 - 410.

Sarter, N.B. (2008). Investigating mode errors on automated flight decks: Illustrating the problem-driven, cumulative, and interdisciplinary nature of human factors research. *Human Factors*, *50*, 506 – 510.

Sarter, N.B., Woods, D.D., & Billings, C.E. (1997). Automation surprises. In G. Salvendy (ed.), *Handbook of Human Factors & Ergonomics, second edition* (pp. 1926 - 1943*)*. New York, NY: Wiley.

Stanton, N.A. & Young, M.S. (2000). A proposed psychological model of driving automation. *Theoretical Issues in Ergonomics Science, 1*, 315-331.

Wilson, J.R. & Rajan, J.A. (1995). Human-machine interfaces for systems control. In J. R. Wilson and E. N. Corlett (eds.), *Evaluation of Human Work: a Practical Ergonomics Methodology* (pp. 357- 405). London: Taylor & Francis.

Woods D.D., Johannesen, L.J., Cook, R.I., & Sarter, N.B. (1994). *Behind human error: cognitive systems, computers, and hindsight.* Wright Patterson Air Force Base, Dayton, OH: CSERIAC.

Young, M.S. & Stanton, N.A. (1997). Automotive automation: investigating the impact on drivers' mental workload. *International Journal of Cognitive Ergonomics*, *1*, 325 - 336.

Mental workload assessment using eye-tracking glasses in a simulated maritime scenario

F. Di Nocera[1], S. Mastrangelo[2], S. Proietti Colonna[1], A. Steinhage[3], M. Baldauf[4], & A. Kataria[4]
[1]Sapienza University of Rome, Italy - [2]Ergoproject s.r.l., Rome, Italy - [3]Future-Shape GmbH, Germany - [4]World Maritime University, Malmö, Sweden

Abstract

The primary goal of this study was the assessment of maritime operators' mental workload in simulation scenarios designed for containing typical traffic situations to be handled by officers during usual routine ship navigation. Taskload was manipulated throughout the sessions to analyse changes in ocular activity during complexity peaks. Specifically, it was tested the viability of implementing a dynamic analysis of eye movements collected through eye tracking glasses (while the operator was freely moving on the bridge) and using the distribution of eye fixations as an indicator of mental workload. Another objective of the present study was to assess the relation between attentional control and the mental workload as perceived by the operators. Results showed that the distribution of eye fixations changed with taskload and that individuals showing high attentional control reported low workload. Furthermore, frequent eye movement transitions have been found between the instruments monitored, suggesting that the information they provide could most usefully be integrated for improving the operators' performance.

Introduction

According to the International Maritime Organisation (IMO) "Shortcomings in human performance at all levels in the chain of responsibility are a major cause of incidents" (IMO Resolution A.1060(28)). Starting from this consideration and assuming that enhancing human performance should be one of the primary goals of a "safer, more secure and environmentally friendly shipping", it becomes imperative to find a way to objectively measure it. This quantification should be even more necessary given the growing proliferation of new technologies (e.g. e-Navigation systems, Unified Bridges, Remote monitoring systems, wearable technologies) whose impact should be carefully evaluated before their introduction. At this stage, institutional initiatives on this topic seem to be still insufficient. This is quite surprising if we consider that in the maritime domain any single on board material and equipment is tested from an engineering standpoint to match precise safety requirements (e.g. breaking load, tensile strength). Using a metaphor: we measure any single part of an anchor chain without appropriately considering its weakest

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

link. In this very case, the weakest link is the interaction between humans, technology, procedures and work environment.

This inconsistency was also noticed by the European Commission that -from 2012 to 2015- consequently funded three Human Factors research projects under the FP7 Programme Transport (SST.2012.4.1-1: Human Element Factors in Shipping Safety):

- FAROS - Human Factors in Risk-based Design Methodology (http://faros-project.eu);
- CASCADe - Model-based Cooperative and Adaptive Ship-based Context-Aware Design (http://www.cascadeproject.eu);
- CyCLaDes - Crew-centered Design and Operations of ships and ship systems. The latter one has been the shell of this study (http://www.cyclades-project.eu).

The present report describes part of the activities carried out in the last project. Mainly, our primary goal was the assessment of Officers Of the Watch (OOW) -who represent the human operators in the socio-technical system of an ocean-going ship- mental workload using eye-movement data collected through wearable eye-tracking glasses. Particularly, we tested the viability of implementing a dynamic analysis of eye movement data collected using eye tracking glasses while the operator was freely moving on the bridge using distribution of eye fixations as an indicator of mental workload.

Moreover, given that the lack of active attentional control can undoubtedly play a role in routine activities carried out on the bridge (where event pacing is slow most of the time, thus facilitating the operator distraction), another objective of the present study was to assess the relation between “absent-mindedness” (Cheyne et al., 2006) and the mental workload as perceived by the operators.

Eye-tracking in maritime HF/E research

The use of eye tracking technology in HF/E research and practice has considerably increased over the years. However, eye-tracking studies on the operator functional state carried out in the maritime context are few and cannot compare to others (e.g. aviation and automotive). There are several reasons for this gap. First, the focus on fatigue rather than mental workload didn’t help. Fatigue is an elusive concept to deal with and does not address very well the complexity of the interaction with technology. Moreover, the maritime context has only recently been adopting technological changes that need mental workload to be addressed. Nevertheless, expression of interest towards the use of ocular measures is increasing among researchers working in the maritime domain (Bjørneseth et al., 2012) and some authors have already employed the technique in several simulation studies. Lützhöft and Dukic (2007), for example, attempted an analysis using eye-tracking data in a pilot study aimed at investigating the ocular strategies adopted by experienced and novice navigators looking at the different instruments on the bridge. They reported that experienced navigators tended to show a lower ratio of glances per minute (a strategy that is probably related to their knowledge and practice with instrumentation). This effect was interpreted as an indication of lower mental workload, but neither behavioural nor subjective measures of mental load were used in that study.

More recently, Forsman et al. (2012) investigated gaze behaviour from both experienced and novice boat drivers during high-speed navigation at sea. They found that novice drivers tended to look at objects close to themselves to a larger extent than experienced drivers, whereas experienced drivers fixated objects in the far distance. Experienced drivers tend to rely more on environmental cues than novices who fixate more on the navigational aids. Results also showed that novices' visual search strategies become less flexible as the workload increases. However, also in this case neither subjective nor behavioural measures of mental workload were used, and changes in workload were inferred only on the basis of the differences in difficulty between tasks or parts of a task.

Besides the scarcity of studies, it is worth noting the applied potential to maritime industry of what has been reported so far. Shipyards are increasingly using both simulations and eye-tracking technologies for their bridge consoles' design/assessment processes. They also consider mental workload evaluation as a relevant part of these processes, even if there is still a lack of consensus on protocols and measures to be used to get meaningful results (see Bjørneseth, 2012).

One technical difficulty with eye-tracking in the maritime domain is that the operator frequently moves from one tool to the other, and only during specific phases of his/her activity sits in front of the bridge. Table-mounted eye-trackers are of little or no use in this context and different strategies for data collection should be employed. One of the aims of the present simulation study is to estimate the usefulness of wearable eye-tracking systems in this field.

Wearable eye-tracking systems shaped as glasses are relatively new devices that have seldom been used in experimental research, whereas there are plenty of marketing-related studies on consumers' scanning strategies. Tonkin et al. (2011), for example, compared consumers' visual behavior when searching for an item located on virtual or physical shelves and Gomes et al. (2014) investigated consumers' attention to labels on beverages. Other studies (e.g. Mormann et al., 2014, Varela et al., 2014 & Wästlund et al., 2015) showed that eye movements between packaged products on a shelf may provide valuable information for making considerations about buying behavior. Nevertheless, the interest towards this type of instruments has been growing and wearable eye-trackers have been used for investigating the gaze patterns of students attending classes (Rosengrant et al., 2012), for evaluating the ease of reading with different page layouts (Chanijani et al., 2015), and for gathering valuable information to improve nurses' medicine-tray preparation task (Rodriguez et al., 2014). An increasing interest towards this technology is also emerging in HF/E research. For example, Gable et al., (2013) recorded eye movements using eye-tracking glasses while performing a primary driving task and secondary list-search on a mobile phone and their results showed differences in fixation data between workload conditions. Other studies have investigated the potential of these instruments for developing automated systems in the field of information security (Neupane et al., 2015) and for studying the effect of the curved display on visual performance and user experience (Choi et al., 2015). However, some limitations exist. For example, in a study involving harvester operators in routine field operations, Häggström et al. (2015) reported that large head movements, lightening changes and vibrations may affect the use of these systems in real-world settings. Another issue is what indicators should be used for gathering valuable information from eye-movement data. Several ocular indicators exist and could be utilized in any domain: from the usability of consumer products to operator functional state assessment (see Duchowski, 2007 for a detailed description of standard metrics). However, many indicators require eye movement recordings in controlled settings, allowing little or no movement to the user/operator. Of course, this is a substantial limitation in those settings where the individual is required to move around a room for accomplishing a task.

Based on previous research on the distribution of eye fixations and its relation with mental workload (see Di Nocera, Camilli & Terenzi, 2007; Proietti Colonna et al., 2014) we wanted to test the feasabily of implementing a dynamic analysis of eye movement data collected using eye tracking glasses while the operator was freely moving around the bridge. For this reason, we monitored OOW interacting with a simulated bridge during a typical coastal traffic scenario: navigation through a sea area with established traffic separation scheme, overlaid by Vessel Traffic Service (VTS) monitoring, with medium to dense traffic, including crossing ferries and selected off-shore activities. Taskload was manipulated throughout the sessions to analyse changes in ocular activity during complexity peaks.

Method

Participants. Twenty seafarers (2 females, mean age = 34.61 years, std dev = 6.25) participated in this study. Participants' experience in the maritime domain ranged from 1 to 25 years. All of the participants were qualified navigation officers with Certificates of Competency in accordance with their designated rank.

Simulator and scenario design. A comprehensive study on mental workload assessment of OOW was performed using a series of real-time simulation trials. In our trials we studied the overall ship's navigation process, which includes many sub-tasks all related to ensure the safe and efficient execution of a ship's voyage from a port of departure to a port of destination. When standardizing minimum requirements for technical equipment for integrated ship navigation, the IMO mentions the following processes and tasks that an integrated navigation system shall contain to support the human operator, the OOW on a ship's navigational bridge:

- Route planning
- Route monitoring
- Collision Avoidance
- Navigation Control (manual and automatic control of ship’s movement)
- Navigation Status (a.o. monitor environmental conditions) and
- Alert Management

All those tasks needs to be simultaneously and uninterruptedly performed by the responsible OOW under all possible prevailing circumstances of environmental conditions, characterized by sea area, water depth, visibility, wind (force, direction) and waves (heights, direction), etc. as well as of the ship status (actual draught, loading conditions, speed/course through water and over ground etc.). The overall aim of the OOW is to ensure, that any risk (of a collision, grounding, etc.) is at least at an acceptable level, and any danger of an accident is avoided. For this objective the human operator is obliged to use all his senses (e.g. for visual look out for collision avoidance) but also “all available means” appropriate to the prevailing circumstances (as e.g. Radar/ARPA, Automatic Identification System for collision avoidance or ECDIS or paper charts, position and depth sensors for grounding avoidance).

The Maritime Risk and Safety (MaRiSa) Simulation laboratory at the World Maritime University (WMU) provides a unique combined ship-handling (SHS) and safety and security training (SST) desktop simulator specifically developed for training and research purposes.

With the bi-directional system, complex simulation scenarios can be provided for various research purposes. The entire complex ship is available to replicate the ship-handling process on the navigational bridge (SHS) combined and connected with engine processes from the engine room (ER) and engine control room (ECR). Moreover, emergency management situations and procedures even inside the vessel as e.g. for fire-fighting, water inrush can be integrated into a simulation scenario with the help of the SST simulator. Figure 1 depicts the configuration of the combined SHS and SST simulator at the MaRiSa simulator laboratory at WMU.

research. As and example, the desktop SHS bridge configuration used for the workload assessment studies corresponds to the SOLAS (International Convention for the Safety of Life at Sea) conformant bridge equipment of a RoRo (Roll-on/roll-off: named according to the way vehicles board and leave the vessel) and RoPAX (RoRo with passenger capacity) ferry. Besides the variable 135° view, it features the elements of the basic navigational equipment including ECDIS, Radar/ARPA with integrated overlay of AIS information, interfaces of GPS; DOLOG (tool that measures vessel’s speed), echo sounder and further devices as well as the handles for rudder, engine and thrusters.

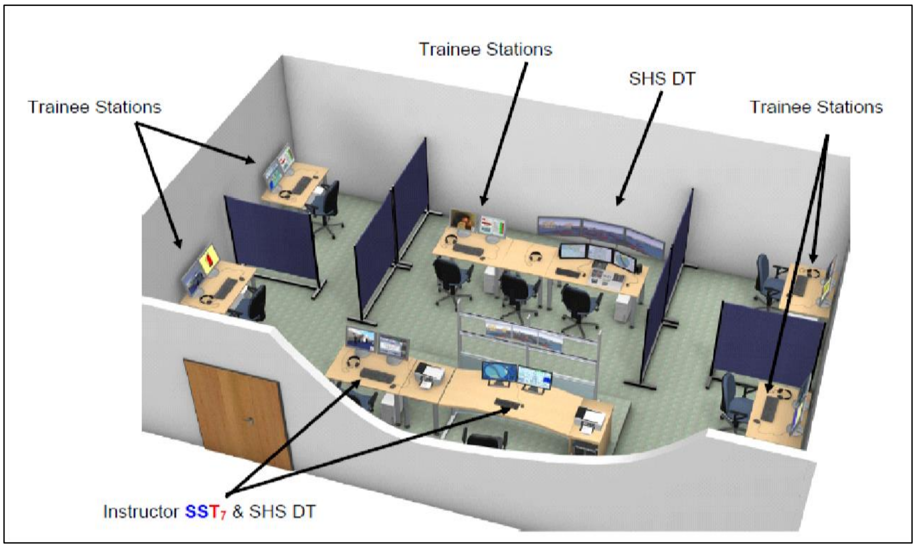


Figure 1. Configuration of the combined Ship-handling and Safety and Security Training Simulator at the MaRiSa-Simulation Lab of World Maritime University, Malmo, Sweden.

This unique configuration allowed for a wide range of simulation scenarios for

This desktop simulation environment has been proven sufficient during series of simulation experiments e.g. investigating potential effects of newly developed applications in the frame of IMO's and IALA's e-navigation initiative and a number of European projects on research and technological development on its impact on behaviour of human operators, like CyClaDes and ACCSEAS (e.g. Baldauf, Benedict & Krüger, 2014; Holder et. al., 2014).

For the purpose of the workload investigations in this research, simulation scenarios have been designed for a series of simulation runs containing typical traffic situations to be handled by navigating officers during usual routine ship navigation. The simulation was run at the MaRiSa facility, providing a typical traffic situation for coastal traffic. In order to create varying workload conditions, challenging situations have been integrated into a typical navigation scenario of 20 min length and requiring careful navigation when coming from an open sea-like area, and then following the approach and passage in the coastal sea area with a traffic separation scheme with a typical medium to dense traffic load situation. Moreover, shore-based VTS monitoring was also added and integrated with the simulation runs, therefore contributing to the realism therein. Through VTS communication, all participants were provided with an update on the local weather conditions and the traffic situation in the area. Furthermore, the VTS doubled up as other communicating entities, such as when participants called other vessels in the vicinity on their Very High-Frequency radio (VHF). The sea area was Northern Öresund (close to Helsingborg – Helsingør).

Bridge instrumentation. During this study, participants interacted with simulated on-board instrumentation. Familiarity with the features of these instruments was assessed using an *ad-hoc* five-point Likert scale. Only 3 participants reported a substantial difference with the type of devices they have previously used. The following is a brief description of the simulated tools and parts of the bridge that were taken into consideration in this study as Areas of Interest (AOIs; figure 2) and for which the transitions of eye movements (between instrument pairs) were examined:

- View of the sea, other ships, land, etc. (OUT)
- Electronic Chart Display & Information System (ECDIS): a computer-based navigation information system that complies with IMO regulations and can be used as an alternative to paper nautical charts.
- Global Positioning System (GPS).
- Automatic radar plotting aid (ARPA) that can calculate the tracked object's course, speed and closest point of approach, thereby knowing if there is a danger of collision with the other ship or landmass.
- Steering and engines controls (CONSOLE).

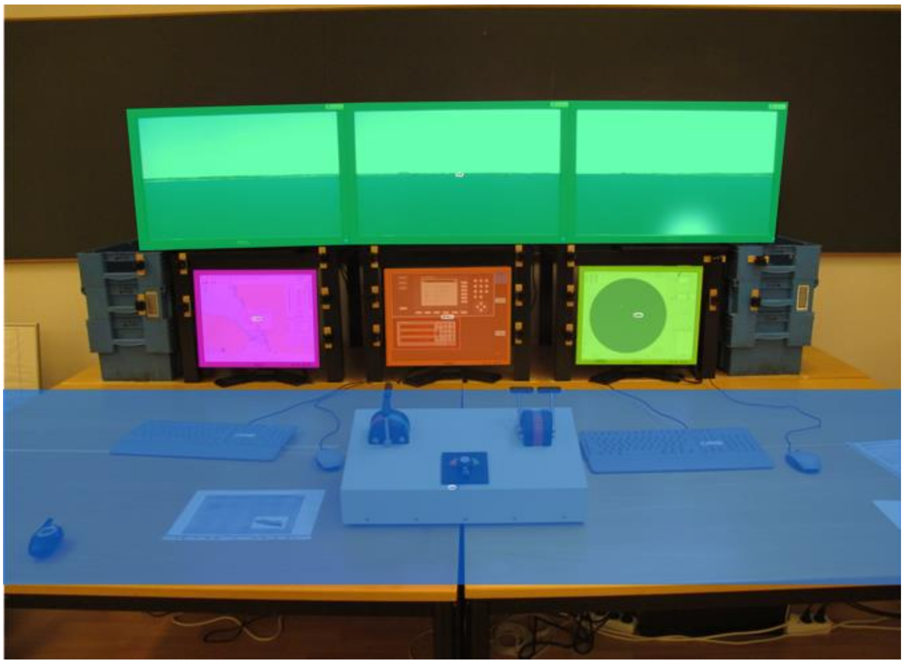


Figure 2. Areas Of Interest (AOIs) from up to down and from left to right: OUT, EDCIS, GPS, ARPA, CONSOLE.

Design. The taskload for the acting OOW's was varied by including crossing ferries and selected off-shore activities (dredging, fishery, maintenance operation). Activity levels were consequently divided into three phases:

- 5 minutes of navigation in easy condition (Easy);
- 10 minutes with increased level of difficulty (Hard);
- 5 minutes of decreased level of complexity (Easy).

The level of complexity was controlled by situations with increasing risk of collisions and additional activities like communication requests via VHF from VTS and/or other targets. The VTS called the participant's ship upon perceiving the development of a potentially dangerous situation, such as the possibility of the participant entering the opposing traffic lane of the Traffic Separation Scheme to keep clear of buoys and vessels in the area such as crossing ferries and dredgers.

This design allowed to test the sensitivity of the Nearest Neighbour Index (NNI: Di Nocera, Camilli & Terenzi, 2007) to changes in taskload by using the phase change from a low taskload to a high taskload condition (Easy to Hard) and *vice versa* (Hard to Easy) as factor in the successive analyses. This was needed because the end of the Hard condition did not necessarily coincide with the end of the evolving situation created by the increased difficulty. The phase (Easy vs. Hard vs. Easy) was instead used as factor for analysing the eye movement transitions between Areas of Interest (AOIs, see figure 2) coinciding with the instrumentation described in the "bridge instrumentation" subsection.

Measures. Measures collected included workload indicators derived from the eye-tracking data and subjective reports, as well as a subjective measure of attentional control. Eye movements were collected using the Tobii Eye Tracking Glasses (version 1): a plastic and rubber pair of glasses with lenses made of hot mirror glass and infrared (IR) reflecting coating. IR markers were placed around the displays to compute eye position. The sampling rate was 30Hz. Time series of the NNI values were successively computed using a newer Matlab version of the ASTEF package (Camilli et al., 2008). The index provides a single value that is indicative of the type of spatial distribution of the data by comparing the mean distance between pairs of (nearest) neighbors in the data to that expected by chance (random distribution). The actual mean distances can be smaller (point are clustered; NNI<1), larger (points are regularly distributed; NNI>1), or not different from the expected distances (points are randomly distributed; NNI=1). A total of 20 values (one each minute) was obtained from the raw gaze data recorded. Transitions between instrument pairs (transitions between each AOI and all the others) have been computed using the Tobii Studio software. The perceived level of workload was assessed using the NASA – Task Load indeX (NASA-TLX: Hart & Staveland, 1988). The level of attentional control has been evaluated using the Cognitive Failures Questionnaire (Broadbent et al., 1982) an inventory of everyday errors providing information on the respondent's attentional control (the higher the score, the lower the attentional control: see Di Nocera, Ferlazzo & D'Olimpio, 2014).

Procedure. Participants filled a short ad-hoc questionnaire about the familiarity with the type of instruments that were simulated into the MaRiSa facility and the CFQ.

Prior starting the simulation they underwent a short calibration procedure and were briefed about the route to sail. The NASA-TLX was administered at the end of the simulation.

Data analysis and results

Concerning the VTS interventions (barring the standard VHF calls to provide weather and traffic updates to all vessels in the vicinity), nearly 80% of the VHF Calls coincided with the central 10 minutes of the simulation runs that pertained to high levels of complexity. That required the OOW to take note of the information and take requisite action as necessary in the form of exercising caution, maintaining fairway, altering speed, making evasive maneuvers, as the case may be.

Eye tracking data recordings were previously examined for completeness and validity. Six participants with incomplete eye-tracking recordings were excluded from the analysis. NNI values were used as dependent variables in a repeated measures ANOVA design using Phase Change (Easy to Hard vs. Hard to Easy) as factor. Results showed a tendency towards statistical significance ($F_{1,13}=3.97$, $p=.07$): NNI values were higher for the passage from Easy to Hard rather than for the passage from Hard to Easy.

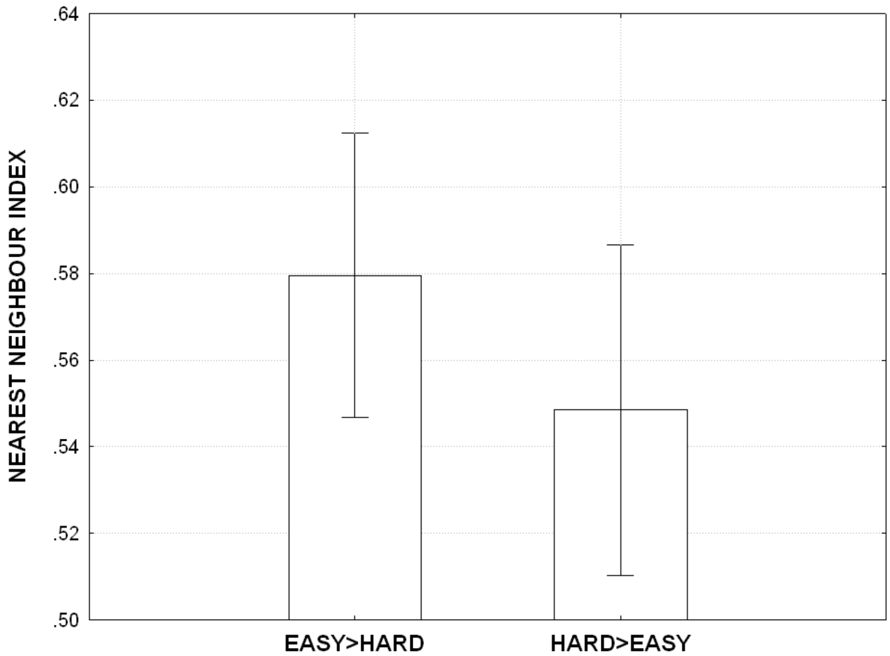


Figure 3. Nearest Neighbour Index by Phase Change.

To assess the role of attentional control on the perceived mental workload the Cognitive Failures Questionnaire median score was used to divide participants into two categories according to the median score: High Attentional Control and Low Attentional Control. Total NASA-TLX scores were employed as dependent variables in an ANOVA design using Attentional Control as factor. Results showed a statistically significant main effect (F1,18=6.38, p<.05).

The proportion of transitions between all instrument were analysed in an ANOVA design Phase (Easy vs. Hard vs. Easy) x Instrument Pair (transitions from each AOI to the others). Results showed a significant interaction between these two factors ($F_{38,722}$=2.12, p<.001). Post-hoc Duncan testing showed that this interaction was due to the following comparisons:

- ARPA-CONSOLE, ARPA-ECDIS, ECDIS-ARPA: In all cases a significantly higher proportion of transitions was found in the second easy condition than the other two phases.
- ECDIS-OUT and OUT-ECDIS: In all cases a significantly lower proportion of transitions was found in the second easy condition than the other two phases.
- ECDIS-CONSOLE and CONSOLE-ECDIS: In all cases a significantly lower proportion of transitions was found in the first easy condition than the other two phases.
- ARPA-OUT: In all cases a significantly higher proportion of transitions was found in the hard condition than the other two phases.

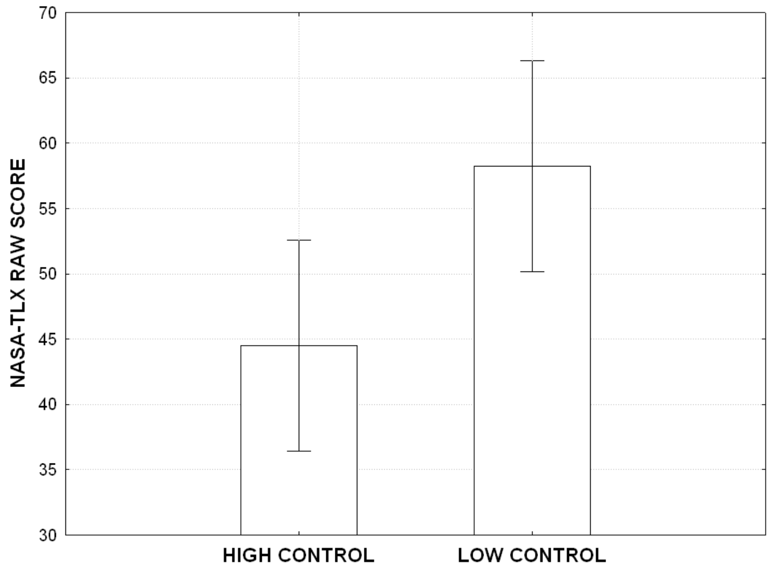


Figure 4. NASA-TLX raw scores by Attentional Control.

Discussion and Conclusions

The use of eye-tracking in the maritime domain is still in its early stage. Nevertheless, the introduction of automated system (and the technological change in general), as well as the consequent reduction of personnel, are boosting the investigation on mental workload and the quest for a good measure to be employed for gathering information on the on-going functional state of the maritime operator.

Results of the analyses carried out on the eye movement transitions between AOIs contributed to better understand the operators' behaviour and could support the ongoing discussions about future ship bridge concept layouts. For example, the frequent ARPA-ECDIS-ARPA transitions indicate of how these instruments are used comparatively and that the information they provide needs to be integrated by the operator. According to the proximity-compatibility principle (Wickens & Carswell, 1995), the displays relevant to a common task or mental operation (close task or mental proximity) should be rendered close together in perceptual space (close display proximity).

Most importantly, results of this simulation study provide a contribution to the development of a measure of mental workload based on the distribution of eye fixations. Indeed, this was a first attempt to employ the NNI index in such a complex setting and, more important, using eye-tracking glasses. Albeit not statistically significant (p=.07), most participants consistently showed the NNI discrimination pattern in phase change (10 out of 14). Many factors may account for this imperfect outcome: from individual differences to the fixation identification algorithm. Nonetheless, the fact that even in such a complex scenario and with a completely different technology a pattern is consistent it is quite encouraging. Future studies will be run for addressing the best way to identify fixations (and consequently their distribution) with eye-tracking glasses.

Interestingly, results obtained from the present study also showed that attentional control (here measured in terms of more or less absent-mindedness) may be involved in operators' workload perception. This is a quite interesting finding that it is worth to be further explored. Indeed, operators' distraction may be crucial in those environments as the bridge of a ship, where event pacing is slow most of the time. Operators who tend to be more distracted (and one could argue that this is not much different from what is usually considered a loss of situation awareness) are those who perceive the task as most loading. Likely, this represents a cost of re-engaging a previously disengaged task, which is the cost of (internal) interruptions that has been reported elsewhere (see Foroughi et al., 2014), but not yet studied in the maritime domain.

Acknowledgements

This study was funded by the European Commission through the FP7 project "CyClaDes - Crew-centered Design and Operations of ships and ship systems".

References

Baldauf, M., Benedict, K., & Krüger, C. (2014): Potentials of e-Navigation – Enhanced Support for Collision Avoidance. *TransNav - the International Journal on Marine Navigation and Safety of Sea Transportation*, 8, 613-617.

Broadbent, D.E., Cooper, P.F., FitzGerald, P., & Parkes, K.R. (1982). The Cognitive Failure Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology, 21*, 1-16.

Bjørneseth, F. B., Renganayagalu, S. K., Dunlop, M. D., Homecker, E., & Komandur, S. (2012). Towards an experimental design framework for evaluation of dynamic workload and situational awareness in safety critical maritime settings. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers* (pp. 309-314). British Computer Society.

Camilli, M., Nacchia, R., Terenzi, M., & Di Nocera, F. (2008). ASTEF: A simple tool for examining fixations. *Behavior research method*s, *40*, 373-382.

Chanijani, S.S.M., Bukhari, S.S., & Dengel, A. (2015). Analysis of text layout quality using wearable eye trackers. In *Multimedia & Expo Workshops (ICMEW), 2015 IEEE International Conference on* (pp. 1-6). IEEE.

Choi, K., Bae, H., Ju, S. W., & Suk, H.J. (2015). 53.2: Visual Search and Attention: What Eye-Tracking Reveals about Visual Performance in the Curved Display. In *SID Symposium Digest of Technical Papers* (Vol. 46, No. 1, pp. 798-801).

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making, 1*, 271-285.

Di Nocera, F., Ferlazzo, F., & D'Olimpio, F. (2014). Is 'error proneness' specific to errors? Predictive validity of the cognitive failures questionnaire. *International Journal of Human Factors and Ergonomics, 3*, 208–225.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373). Springer Science & Business Media.

Foroughi, C.K., Werner, N.E., Nelson, E.T., & Boehm-Davis, D.A. (2014). Do interruptions affect quality of work?. *Human Factors, 56*, 1262-1271.

Forsman, F., Sjors-Dahlman, A., Dahlman, J., Falkmer, T., Lee, H.C. (2012). Eye Tracking during High Speed Navigation at Sea. *Journal of Transportation Technologies*, *2*, 277-283.

Gable, T.M., Walker, B.N., Moses, H.R., & Chitloor, R.D. (2013). Advanced auditory cues on mobile phones help keep drivers' eyes on the road. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 66-73). ACM.

Gomes, T., Hurley, R.A., Duchowski, A., Darby, D., & Ouzts, A. (2014). The effect of full body versus partial body graphic labelling on beverage packaging. *Packaging Technology and Science*, *27*, 933-943.

Häggström, C., Englund, M., & Lindroos, O. (2015). Examining the gaze behaviors of harvester operators: an eye-tracking study. *International Journal of Forest Engineering*, *26*, 96-113.

Hart, S.G. & Staveland, L.E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati (Eds.) *Human Mental Workload*. Amsterdam: North Holland Press.

Holder, E., Motz, F., Horoufchin, H., & Baldauf, M. (2014). Concept for the Integration of Information Received Via Communication Equipment with Onboard Navigational Systems. In N. Stanton, G. Di Bucchianico, A. Vallicelli, and S. Landry (Eds.), *Advances in Human Aspects of Transportation: Part I* (pp. 203-215). AHFE Conference.

Mormann, M., Pieters, R., Wedel, M., Kovacheva, A., Inman, J.J., Brasel, S.A., & Gips, J. (2014). The Eyes Have It: New Insights From Tracking Attention Patterns in Consumer Research. *Advances in Consumer Research*, *42*, 146-149.

International Maritime Organisation (2014). Strategic Plan for the Organization (for the six-year period 2014 to 2019. IMO Assembly Resolution A.1060(28).

Lützhöft, M., Dukic, T. (2007). Show me where you look and I'll tell you if you're safe: Eye tracking of maritime watch-keepers. *Paper presented at the 39th Nordic Ergonomics Society Conference*. October 1-3, Lysekil, Sweden.

Neupane, A., Rahman, M. L., Saxena, N., & Hirshfield, L. (2015, October). A Multi-Modal Neuro-Physiological Study of Phishing Detection and Malware Warnings. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 479-491). ACM.

Proietti Colonna, S., Capobianco, C., Mastrangelo, S., Di Nocera, F. (2014). Implementing dynamic changes in automation support using ocular-based metrics of mental workload: a laboratory study. In D. de Waard, J. Sauer, S. Röttger, A. Kluge, D. Manzey, C. Weikert, A. Toffetti, R. Wiczorek, K. Brookhuis, and H. Hoonhout (Eds.) (2015). *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2014 Annual Conference*. (pp. 91-97). Downloaded from hfes-europe.org. ISSN 2333-4959 (online).

Rodriguez, N. D., Lilius, J., Bjorklund, S., Majors, J., Rautanen, K., Danielsson-Ojala, R., & Tuominen, I. (2014). Can IT health-care applications improve the medication tray-filling process at hospital wards? An exploratory study using eye-tracking and stress response. In *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on* (pp. 423-428). IEEE.

Rosengrant, D., Hearrington, D., Alvarado, K., Keeble, D., Rebello, N.S., Engelhardt, P.V., & Singh, C. (2012). Following student gaze patterns in physical science lectures. In *AIP Conference Proceedings-American Institute of Physics*, *1413*, 323.

Tonkin, C., Ouzts, A.D., & Duchowski, A.T. (2011). Eye tracking within the packaging design workflow: interaction with physical and virtual shelves. In: *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications* (p. 3). ACM.

Varela, P., Antúnez, L., Cadena, R. S., Giménez, A., & Ares, G. (2014). Attentional capture and importance of package attributes for consumers' perceived similarities and differences among products: A case study with breakfast cereal packages. *Food Research International*, *64*, 701-710.

Wästlund, E., Otterbring, T., Gustafsson, A., & Shams, P. (2015). Heuristics and resource depletion: eye-tracking customers' in situ gaze behavior in the field. *Journal of Business Research*, *68*, 95-101.

Gestures while driving: A guessability approach for a surface gestures taxonomy for in-vehicle indirect interaction

Joana Vieira[1], Rosane Sampaio[1], Rafael Nascimento[2], João Pedro Ferreira[2], Sara Machado[3], Nuno Ribeiro[3], Estevão Silva[3], & Sandra Mouta[1]
[1]Centro de Computação Gráfica, [2]Universidade do Minho, [3]BOSCH Car Multimedia, SA
Portugal

Abstract

Surface gesture interaction in the automotive context is still exploratory and lacking guidelines. To address this issue, a guessability study was developed to associate end-user gestures with functionalities of an in-vehicle HMI system. Interaction with the system was performed indirectly, with the use of surface gestures. Participants were presented with instructions, followed by a static interface image of a menu (e.g.: music list, contact list), and prompted to create a gesture that would allow them to respond to the instruction (e.g.: "Select previous" or "Make a call"). Results demonstrated that the gestures proposed in the concept phase were simple and familiar, and allowed the creation of a gestures' taxonomy for adjustment, acceptance, refusal, and navigation actions. The guessability methodology proved to be useful and demonstrated how user-centered design can improve the usability of an interaction even at an advanced stage of the design and development process.

Introduction

Surface gesture interaction bloomed with the growing applications of smartphone and tablet technology. With the commercialization of motion-sensitive controllers, gesture interaction has slowly become a must-have in current devices. Under the umbrella of "natural" gestures, some interactions are defined without previous testing, or even disregarding standards of interaction design. In some devices, users trigger functions unwittingly, because interaction with systems lack a general interaction consistency (Norman, 2010).
Following the need to develop gestures established over solid usability guidelines, guessability or user-elicitation studies are becoming more frequent along the development process of interaction devices.

Guessability or user-elicitation studies

Guessability is the cost associated with the use of a new interface for the first time, and it can be translated into time, errors or effort in performing a given task (Moyes,

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

& Jordan, 1993). As an example, a fire extinguisher should imply less time and effort in its usage, which means it should have high guessability (Jordan, Draper, MacFarlane, & McNulty, 1991) as it will only be used in an emergency context. Systems should crave high guessability in order that a user's first interaction is immediately successful without consulting support documentation.

User-elicitation or guessability methodologies have been mainly used to gather useful information for the design phase of gesture-based interfaces. In a guessability study, participants are prompted to create a gesture for a particular action, with no given training or feedback. It is assumed that all input is accepted by the interface.

Lee and Wong (2015) made a guessability study for gesture interaction with public information displays using augmented reality technologies. The goal was to identify intuitive gestures for the interaction with the displays, asking the users to define and act the gestures they thought more adequate to achieve the interaction tasks. A sequence of animations was shown on the screen and participants had to act out a type of gesture that would perform the task presented in the animation.
In another study, Vatavu (2012) conducted an experiment to elicit user input for free-hand gestures for frequent interactions with a TV, resulting in gestures for free-hand TV controls. Instructions, usually named *referents*, were presented to the participants with a short description and by running a video demonstration of the referent effect (e.g.: volume going up). Participants were then asked to propose a gesture that would trigger that effect. After an agreement analysis, the results obtained allowed the creation of twelve proposed gestures for each referent, along with several guidelines for further application in smart environments.

Similar procedures were used for surface computing (Wobbrock, Morris, & Wilson, 2009) and mobile interaction (Ruiz, Li, & Lank, 2011). In all cases, very simple interfaces were used, either static or animated, explaining or demonstrating the expected effect after a given interaction. With this methodology, the end user informs the designers about the most immediate and intuitive gesture for a given action. The results are useful as they allow building general taxonomies or specific gestures when a high agreement rate is observed among participants.

Context and goals of current study

The current guessability study was developed in the context of HMIEXCEL project, a project of national interest co-funded by the Portuguese government (COMPETE; FEDER) and the European Commission, based on the co-promotion between Bosch Car Multimedia Portugal and University of Minho. HMIEXCEL intended to develop advanced in-vehicle multimedia solutions.

One of the developed concepts for the in-vehicle infotainment system interface involved indirect interaction with a touchscreen in the central stack area. The basis of the concept is that by using a limited set of gestures, a driver can interact with this system without taking the eyes off the road. The system consists of two displays presented vertically, one atop the other in the central stack area (Figure 14).

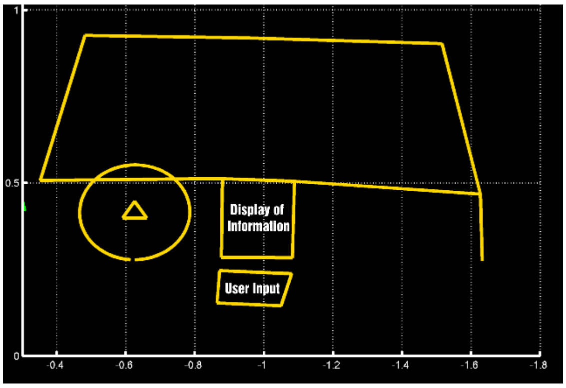


Figure 14. In-vehicle interaction concept: a top display is used to present information and a bottom display receives user inputs

The top display presented images and animations like a regular screen, while the bottom display worked as a touchpad and received all user input. The interaction was made on the bottom display using touch interaction, which will henceforth be referred to as *surface gesture*.

The testing goals included finding proximity between surface gestures from different users, creating a taxonomy for the selected gestures; and identifying potential design concerns in order to improve the efficiency, effectiveness, and end-user satisfaction of the interface. In work process terms, this would mean an analysis of the recorded videos, the search for recurrent patterns, building a general taxonomy that would translate this recurrence and, finally, suggesting surface gestures for particular actions.
Gestures were elicited from participants by presenting a static instruction to perform an action – a *referent* – followed by a picture of the HMI system's interface. The participants proposed a surface gesture that would perform that action, according with the interface.

Method

Participants and apparatus

Thirteen participants holding a Portuguese driver's licence were recruited (8 male, 5 female). All were technology savvy, with ages ranging from 23 to 45 years (M=29, SD =6). All participants had normal or corrected-to-normal vision, and no upper-body impairments limiting the interaction with the device.

Participants were seated in a Logitech Racing Kit G27 equipped with a Play Seat Evolution steering wheel and three pedals. There was no driving task involved, but participants should have a similar range of movements to that of the driving position.

Two displays were placed on the right side of the participant, at a comfortable distance for interaction. Both had a visible and functional area of 9.7”. The bottom display was an iPad© Air (2048x1536, 264 ppi) and it was the only with a touch surface area. The displays were placed vertically on top of each other, with the bottom display having 60º degrees of inclination relatively to the top display (Figure 2).

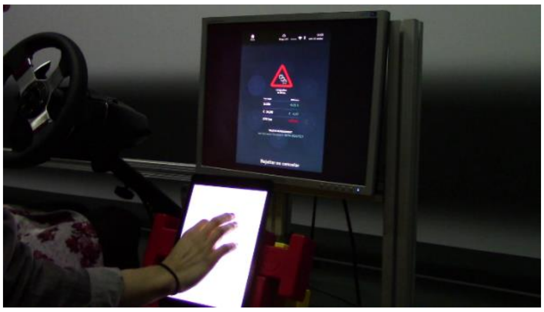
(a)  (b) 

Figure 2. Set-up used for the study: (a) view from camera and (b) animated capture from the bottom display’s screen using TUIOpad.

All interactions were recorded with a video camera placed behind the participant. The bottom display’s screen capture was registered both with video and an open-source mobile application to record touch data. The iPad© screen was recorded using Quicktime (Apple, 2015) running on an external computer with OS X Yosemite (Apple, 2014). Data related with surface gesture interaction was collected with TUIOpad (Akten & Kaltenbrunner, 2010) an open Source App for iOS used to capture multi-touch data and send it to a client via a TUIO-based communication protocol. An external computer, running a custom made application, received data from the TUIO server and saved it on a file.

The bottom display registered the number of fingers used, the direction and location of each surface gesture. The referents on the top display were presented using E-prime (Psychology Software Tools, Inc., 2012).

Referents

The referents are the commands or actions that the participant is instructed to perform. Twenty-one referents were selected. They were presented using written instructions followed by static images of the graphic interface of the interaction concept, such as a music or contact list. Figure 3 depicts three examples of the twenty-one referents used, as well as their respective graphical user interface.

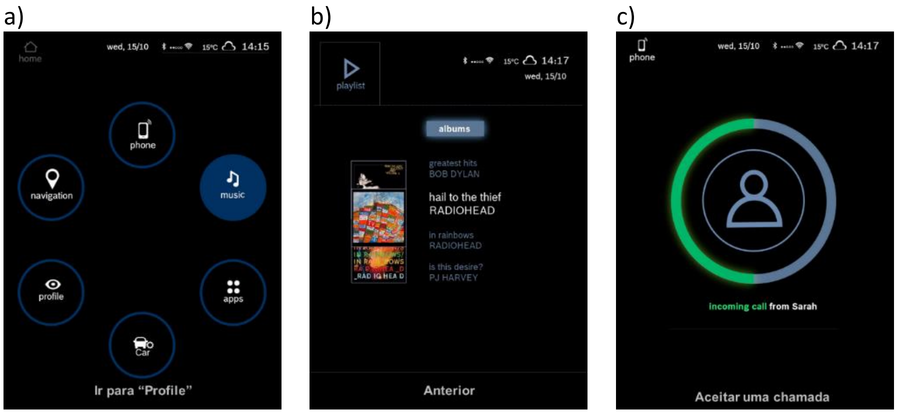


Figure 3. Three examples of referents. From left to right: a) Go to “profile”; b) “Previous”; c) Accept Call.

The selection of the referents followed the stakeholders’ priorities for the interaction concept, focusing more on general navigation actions, and navigation through phone and music menus. The study also served as a test to the visual interface created up to then, as the context allowed seeing whether it was suggestive enough or needed further indications. As used in this study, the interface had a permanent black background, with a six-item main menu arranged in a circle. Once inside a menu, the navigation was mainly list oriented, presenting items vertically. No explicit cues were presented suggesting directions of movement.

Some referents were of generic navigation in the HMI system, and others were specific functionalities of the “Phone” and “Music” sub-menus. All referents used in the experimental session are summarized in Table 1.

Table 12. Twenty-one referents used in the guessability study, grouped by type of functionality.

General HMI Navigation	***Phone Menu***	***Music Menu***
Go back to Main Menu	*Show contact info*	*Next*
Select an option	*End a call*	*Previous*
After selected, choose an option	*Send an SMS*	*Increase volume*
Back to previous menu	*Make a call*	*Decrease volume*
Go inside a sub-menu	*Silence a call*	
Go to “Profile”	*Accept a call*	
To accept or agree	*Open an SMS*	
To reject or to cancel	*Cancel a call*	
	Reject a call	

Procedure

All participants read and signed an informed consent and filled the questionnaire about their previous use of technology. Participants were seated to the left of both displays. The instructions to the participants referred that they would be presented written commands on the top display, and they should perform a gesture to execute that action. They were also informed that the input to the system on their right side was indirect, meaning that there were no buttons or touch areas to be looked at.

After a brief period of training with different referents, the first of two similar blocks was presented. The referents inside the blocks were organized in sets of two images: the first image was visible on the top display for 3 seconds and it showed a black canvas with the written referent (e.g. “Previous”); the following image represented a static image of the HMI graphical interface and was visible for 10 seconds, during which period the participant should perform one surface gesture on the bottom display. This gesture should allow one to execute the action described by the referent. After performing each gesture no feedback was given, and the experiment advanced to the next referent. The bottom display remained blank.

The experiment was divided in two blocks with one interval between, managed by the participant. In each block, twenty-one referents were presented twice. Thus, each participant performed all prompted gestures four times, in a randomized manner. This organization gave information about the evolution and internal consistency of the gestures the participants made. The entire procedure lasted approximately 30 minutes.

Guessability and agreement measures

Two types of measures regularly used in guessability studies were also applied in the current study, the Guessability and Agreement measures.

To analyze the guessability of a referent, Wobbrock, Aung, Rothrock, and Myers (2005) proposed a measure of guessability (*G*) that intends to infer the guessability of a proposed referent, which was here adapted for surface gestures:

$$G = \frac{\sum_{sg \in SG} |P_{sg}|}{|P|} . 100\%$$

where *P* is the set of proposed gestures for all referents, *SG* is the set with all surface gestures made for a given referent, and P_{sg} is the set of proposed surface gestures in set *SG* using gesture *sg*. For instance, if the “one finger tap” gesture was used 250 times, the guessability measure for the “Select an option” referent would be calculated by: *SG* = {38} and *G* = 38/250*100% = 15,2%. Of all the 250 times “one finger tap” was used, 38 times was for the “Select an option” referent, and this means this referent was able to accommodate 15,2% of the “one finger tap” gestures proposed by the participants.

The agreement score (*A*) also proposed by Wobbrock and colleagues (2005), reflects the degree of consensus among participants, is defined by:

$$A = \frac{\sum r \in R \sum P_i \subseteq P_r \left(|P_i| / |P_r| \right)^2}{|R|}$$

where *r* is a referent in the set of all referents *R*, P_r is the set of proposed gestures for referent *r*, and P_i is a subset of identical gestures from P_r. For example, referent "*Back to Main Menu*" had 8 different gestures proposed (39 in total), one with 1 occurrence and the others with 17, 1, 4, 4, 1, 4 and 7 occurrences. These were considered different groups, and the calculation was the following, resulting in a score of 0.26:

$$A_{back\ to\ Main\ Menu} = \left(\frac{1}{39}\right)^2 + \left(\frac{17}{39}\right)^2 + \left(\frac{1}{39}\right)^2 + \left(\frac{4}{39}\right)^2 + \left(\frac{4}{39}\right)^2 + \left(\frac{1}{39}\right)^2 + \left(\frac{4}{39}\right)^2 + \left(\frac{7}{39}\right)^2 = 0.26$$

The agreement score ranges between $|P_r|^{-1}$ and 1 (absolute agreement), and it is higher when a small number of different gestures is proposed.

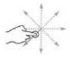
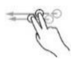
Results

The technology use questionnaire demonstrated that most participants used Android systems (54%), although 38% interacted daily with both Android and iOS systems. Only 8% used exclusively iOS. All were regular smartphone users and interacted daily with a laptop or desktop. Most users also interacted with tablets (77%), but inside a car most interaction was made indirectly, using buttons (62%).

A total of 1092 gestures were captured (13 participants x 21 referents x 4 repetitions). For the analysis of the gestures, all videos were replayed and analyzed. Both the video registered by the camera and the screen capture were played side by side. All gestures were analyzed according to the following criteria: Number of fingers (one, two, three); Type of touch (tap, long press, slide, scroll, pinch); Direction (left to right, right to left, diagonal). Out of several surface gestures collected, Table 2 depicts and describes the gestures performed more often.

The frequency, guessability and agreement of the gestures were analyzed.

Table 2. Description of surface gestures performed more frequently

	Frequent surface gestures			
 **One Finger Compass Rose**	Index finger slides in one of six possible directions, starting at the center of the display		Index finger taps once, regardless of spatial coordinates	 **One Finger Tap**
 **One Finger Double Tap**	Index finger taps twice, regardless of spatial coordinates		Thumb and index or medium fingers slide inwards	 **Pinch in**
 **One Finger Long Press**	Index finger presses the display for a longer period of time, regardless of spatial coordinates		Thumb and index or middle fingers slide outwards	 **Pinch out**
	Frequent surface gestures (cont.)			
 **One Finger Slide Down-Up**	Index finger slides upwards		Index and middle finger slide upwards	 **Two Fingers Slide Down-Up**
 **One Finger Slide Up-Down**	Index finger slides downwards		Index and middle finger slide downwards	 **Two Fingers Slide Up-Down**
 **One-Finger Slide Right-Left**	Index finger slides from right to left		Index and middle finger slide from right to left	 **Two Fingers Slide Right-Left**
 **One Finger slide Left-Right**	Index finger slides from left to right		Index and middle finger slide from left to right	 **Two Fingers Slide Left-Right**

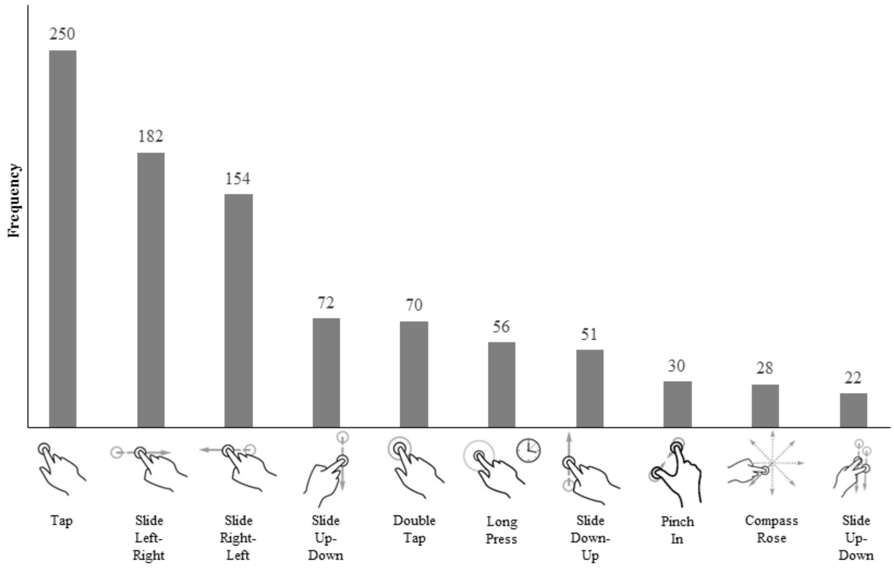


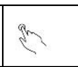
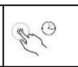
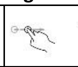
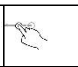
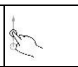
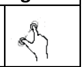
Figure 4. Surface gestures organized by frequency

In some cases, the first and second most frequent gestures were antagonistic. For instance, to answer a call, almost half of the participants slided one finger to the right, but about 20% made the slide on the opposite direction.

Of all 1092 gestures collected, most were distributed among ten categories depicted in Figure 4. Of these, 915 answers were selected for further analysis, illustrating the gestures made more than 20 times by participants. The most frequent surface gestures used one finger and consisted in simple taps and slides.

Table 3 depicts the guessability measure for the referents whose gestures were made more than 20 times (84% of all proposed gestures).

Table 3. Guessability results. Percentage of proposed gestures accommodated by each referent. The highest value for each referent has a grey background.

	One Finger								Two Fingers	
										
	Compass Rose	Tap	Double Tap	Long Press	Slide Left Right	Slide Right Left	Slide Up Down	Slide Down Up	Slide Up Down	Pinch In
	%	%	%	%	%	%	%	%	%	%
Back to Main Menu		0,4	5,7	7,1	9,3		1,4	2,0	31,8	13,3
Select an option		15,2	2,9	7,1				5,9		
Selected-choose option		12,0	8,6	14,3	1,1					
Previous menu		0,4	2,9		15,4	4,5	4,2		4,5	
Go inside a sub-menu	39,3	11,6	5,7	5,4						
"Driver Profile"	60,7	8,8	5,7							
To accept or agree		7,6	1,4		7,1	3,2				
To reject or to cancel					6,0	11,7	5,6			13,3
Show contact info		6,8	8,6	37,5	1,1	0,6				
End a call		2,8	2,9		4,4	11,7	8,3			10,0
Send an SMS		5,6	10,0	16,1	2,7	9,1		2,0		
Make a call		7,6	8,6	3,6	8,2	6,5				
Silence a call		2,0	12,9	3,6	2,2	1,3	15,3		9,1	23,3
Accept a call		2,4	1,4		13,2	5,8		3,9		
Open an SMS		12,8	14,3	3,6		3,2				
Cancel a call			1,4	1,8	4,9	9,7	4,2			13,3
Reject a call					5,5	15,6				13,3
Next		0,8	7,1		4,4	11,0	1,4	27,5		
Previous					12,1	3,2	20,8	3,9		
Increase volume		1,6			2,2			54,9		
Decrease volume		1,6				2,6	38,9		54,5	13,3

According to Table 3, there are some conflicts as the same surface gestures were proposed for different referents. Nevertheless, the gesture with the highest value in each referent, for all referents, was selected (value in grey background for the highest value in the row). Some gestures are evident for some referents, like navigation in menus with 61% plus 39% for the compass rose, or increasing and decreasing volume with 55% of guessability for the slide up and down gesture. These values mean that the referents are either suggestive enough for a gesture, or make use of more semantic knowledge, like increasing and decreasing volume. Others are not so attribuitable to a given referent, like double tapping to "Make a Call" with only 9% of guessability.

Further agreement analyses were made to better fundament any selection. These values reflect the conflicts observed previously. Lower agreement scores corresponded to a higher number of proposed gestures by referent. General agreement score ranged between 0.17 and 0.67 (Figure 5).

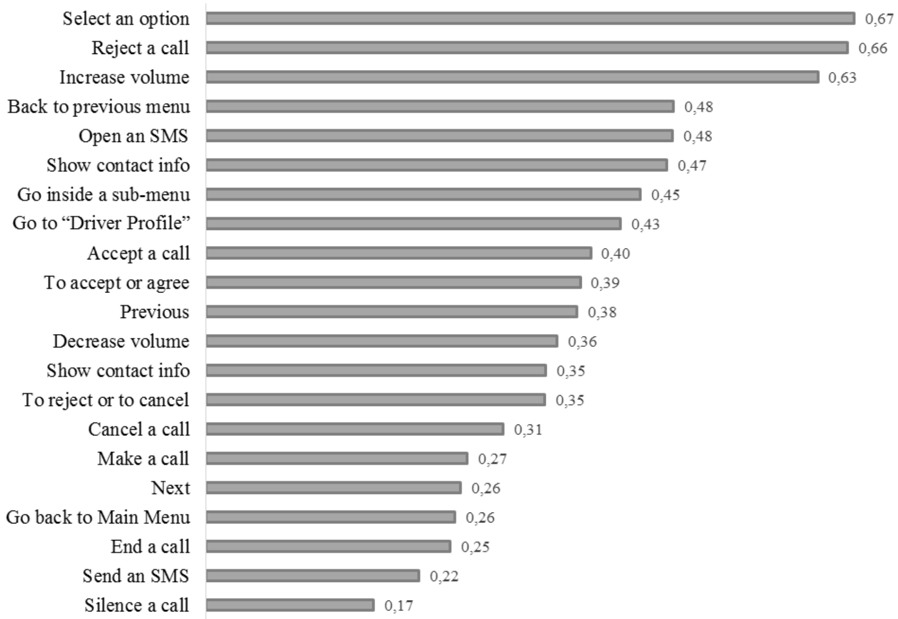


Figure 5. Agreement Scores for surface gestures used for referents. The score ranges between $|P_r|^{-1}$ *and 1.*

Figure 5 demonstrates that the referents with the highest agreement score between participants were "Select an option", "Reject a call" and "Increase volume". All other referents had a score of under 0.5, which indicates some variability among participants, pointing towards the need for a more suggestive interface for the HMI system.

Taxonomy of Gestures

The creation of a taxonomy of gestures intended to have practical effects on the development process of the HMI system, and looked for macro patterns. This allowed verifying patterns that were organized into the proposed gesture taxonomy for indirect input (Table 4). In this exercise, three main actions were identified among all referents: To adjust, to make acceptance or refusal actions, and to navigate through menus.

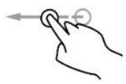
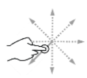
Table 4. Proposed taxonomy of gestures

ADJUSTMENTS/CHANGE MODE	Adjust intensity (volume)
	Change position (music)
	Change status (on/off)
	Change size (zoom)
ACCEPTANCE ACTIONS	OK, Accept, Agree, Open, Make, Send…
REFUSAL ACTIONS	Reject, Cancel, Finish…
NAVIGATION	Move through list
	Move inside or through menus

Some referents were not included in the taxonomy because they need specific and recognizable gestures in any context, such as the “Home”, and “Return” functions.

After analyzing frequency, guessability and agreement measures, it was possible to associate a surface gesture for 12 of the 21 referents (Table 5), making a total of eight different surface gestures. These were based in the referents that gathered higher guessability and agreement scores.

Table 5. Library of surface gestures proposed for 12 of the 21 referents

Taxonomy	Referent	Proposed Gesture	
ADJUSTMENTS/ CHANGE MODE	**CHANGE VOLUME** Increase /Decrease Volume		Slide Down-Up/ Up-Down
ACTION: ACCEPTANCE	Select an Option		One Finger Tap
	Open SMS		One Finger Double Tap
	Make / Accept a Call OK/Agree		One Finger Slide Left-Right
ACTION: REFUSAL	End /Cancel/ Reject a Call Reject/Cancel		One Finger Slide Right-Left
NAVIGATION	**NAVIGATION IN MENUS** Go to “Driver Profile” Go to a sub-menu		One Finger Compass Rose
GENERAL/ OTHERS	**PREVIEW** Show contact info		One Finger Long Press

The proposed taxonomy allowed to group surface gestures according to their function or semantic meaning (acceptance/refusal).

Discussion

This study has provided important results regarding preferences and expectations of the participants when interacting with an HMI system, which should be integrated in the surface gesture design process.

The fact that the interaction was indirect influenced mapped gestures. For instance, about 58% “Tap” gestures were made in the corresponding place where the object was presented on the top display. This indicated that participants transferred the top spatial representation to the bottom display, and touched it accordingly. Although the location of the gesture has no influence on the resulting feedback of the concept, it is an interesting pattern. This result demonstrates how the proposed gestures were economic, in the sense that the simplest surface gesture could be used in several contexts if there was a correspondence between the presented interface and the place where the input was made.

It was also evident that participants used very simple and familiar gestures, and most gestures used only one finger, mainly the index finger. The fact that all users interacted daily with smartphones and/or tablets might have influenced the surface gestures made throughout the experiment, sometimes with direct transposition. Another explaining factor could be the 10s time limit imposed, favoring simple and quick surface gestures. This result indicates that the number of fingers should never be a differentiating factor when activating frequent functions.

Having opposite gestures for opposite actions was also common, which demonstrated general coherence and semantics of the proposed gestures. For instance, participants either slided down-up or slided up-down to increase and decrease volume, respectively. It was also interesting to verify the tendency to associate positive actions with the left to right movement, and refusal or negative actions with the right to left movement.

The results suggest that the system would benefit from more guidance cues to facilitate interaction. Although the main menu seems to be explanatory enough, for it gathered one of the highest guessability values with the compass rose gesture, most gestures were grouped in the “Tap” and “Slide” categories, and some referents shared a high number of inputs from the same surface gestures. Nevertheless, the twelve referents represented in the final proposal of gestures managed to gather higher levels of agreement and guessability.

The guessability of the interface under study can be improved, and this first exploratory analysis allowed to identify where the focus of the next development phase should be: in placing cues for users indicating the direction of the gestures, and in making all navigation contextual, in order to have the possibility to use the same gesture in different contexts.

Conclusion

A guessability study was developed to propose a library of gestures for an indirect interaction concept. The outputs of the study indicate that all participants, independently of their experience with mobile devices, preferred simple and familiar gestures using the index finger. It was possible to propose eight gestures for a total of twelve referents, out of twenty-one. Some referents needed different gestures due to their specificity (e.g. “Return Home”), and it became clear that participants converged towards the same simple and familiar surface gestures, avoiding distinctive input. The study also suggested that the HMI interface could be improved with more cues for the user, indicating, for instance, the direction of the surface gesture.

Further stages of this research will include the validation of the gestures, incorporating the selected gestures in a real interactive context. This is foreseen in guessability methodology, and would provide further information on the learnability of the gestures in a real context. The guessability methodology demonstrated how user-centered design can be applied even at an advanced stage of the design and development process, using simple set-ups and analysis. Nevertheless, it should be emphasized how significant changes might result from these studies, and how valuable these methods can be in initial stages of design and development, leading to a reduced amount of time spent testing and correcting concepts.

Acknowledgements

This research was supported by: Operational Competitiveness Program - COMPETE, QREN (Quadro de Referência Estratégico Nacional), European Regional Development Funds (European Union); BOSCH for the project HMIEXCEL - I&D crítica em torno do ciclo de desenvolvimento e produção de soluções multimédia avançadas para automóvel; Strategic program FCT-UID/EEA/00066/2013.

All illustrations provided and adapted from GestureWorks® (www.gestureworks.com)

References

Akten, M., Kaltenbrunner, M. (2010). TuioPad. Martin Kaltenbrunner.

Apple. (2014). OS X Yosemite.

Apple. (2015). QuickTime.

Jordan, P.W., Draper, S.W., MacFarlane, K.K., & McNulty, S.A. (1991). Guessability, learnability, and experienced user performance. In N. Diaper, D., Hammond (Ed.), HCI’91 People and Computers VI: Usability Now (pp. 237–245). Cambridge University Press.

Lee, G.A., & Wong, J. (2015). User Defined Gestures for Augmented Virtual Mirrors : A Guessability Study. In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (pp. 959–964). New York, USA: ACM.

Moyes, J., & Jordan, P.W. (1993). Icon design and its effect on guessability, learnability, and experienced user performance. *People and Computers*, *8*, 49–60.

Norman, D. (2010). Designing For People Gestural Interfaces : A Step Backwards In Usability Visibility. Interactions.acm.org, 17, 46-49. Retrieved April 24, 2015.

Psychology Software Tools, I. (2012). E-prime. Psychology Software Tools, Inc.

Ruiz, J., Li, Y., & Lank, E. (2011). User-defined motion gestures for mobile interaction. Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems - CHI '11, 197.

Vatavu, R.-D. (2012). User-defined gestures for free-hand TV control. In Proceedings of the 10th European conference on Interactive tv and video - EuroiTV '12 (pp. 45–48). New York, USA: ACM Press.

Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005). Maximizing the guessability of symbolic input. In CHI '05 extended abstracts on Human factors in computing systems - CHI '05 (pp. 1869–1872). New York, New York, USA: ACM Press.

Wobbrock, J. O., Morris, M. R., & Wilson, A. D. (2009). User-defined gestures for surface computing. Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09, 1083.

Comparison of an old and a new Head-Up Display design concept for urban driving

Martin Götze, Christian Schweiger, Johannes Eisner, & Klaus Bengler
Institute of Ergonomics, Technische Universität München
Germany

Abstract

In the future, full use of advanced driving assistance systems will move from highways and freeways to urban areas. This additional assistance systems use case may require the communication of additional information and warnings. Consequently, how information is currently presented in head-up displays will need to change in order to not visually overload the driver. Current head-up design concepts are usually arranged in three clusters: the driven speed, navigation information, and various assistance systems. The new design concept needs to be more generic, driver-focused, and action-orientated. For this experiment, an old design concept and a potential new one were compared in terms of reaction time, response accuracy, and a subjective evaluation. Thirty participants (N =30) performed an occlusion task and a choice reaction time task to find possible differences between the designs. Statistical tests were performed to examine global and specific accuracy. The occlusion task showed no significant difference between the two designs. However, the new design yielded better CRT performance in terms of response accuracy and reaction times. Then again, the subjective measures showed an advantage of the old HUD design. In conclusion, this paper shows the benefits and downsides of a new urban HMI concept.

Introduction

Driving a vehicle is a visual and cognitive burden for the driver, especially in urban areas (Recarte & Nunes, 2003; Rumar, 1990). Nevertheless, looking at the number of accidents in recent years the number of road fatalities has fallen sharply (Statistisches Bundesamt Wiesbaden, 2014). First and foremost, with 93%, is human error which has made a major contribution (Winner et al., 2012). Analysing the types of human error involving personal injury, it can be seen that most of them could be avoided, or at least assisted, by a suitable human-machine-interface (HMI) design providing the right information in the right way. For this purpose, a generic and integrative HMI design concept for head-up displays (HUD) is proposed, with warnings and information presented in a driver-focused and action-orientated manner (Petermann-Stock & Rhede, 2013).

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

The scenarios and characteristics of urban areas are quite different from rural areas or highways with a higher frequency of necessary manoeuvres (Gevatter & Grünhaupt, 2006). In addition, the city environment is characterised by a much higher complexity (Schröder, 2012) and multiple road users like other cars, trains, bikes and even weaker traffic participants (Winner et al., 2012), see Figure 1. Thus, about 68% of all accidents in 2010 involving personal injury occurred in these areas. Some of the reasons were also seen in the increased decision making, the much shorter time window and the fast sequence of traffic notifications or road signs (Götze et al., 2014).

Figure 1. Typical urban area scenarios with multiple road user and a much higher complexity compared to rural areas. © iStock.com/Bim

The head-up display is well known from aviation and was first introduced in an automotive context by Bubb (1978). In comparison with head-down displays, like the instrument cluster or any other in-vehicle displays, the main advantage is the virtual image (2 - 2.5 m infront) in the line of sight of the driver (see Figure 2), which leads to significantly faster reaction times to the presented information (Wittmann et al., 2006) and still about 40% - 50% visual acuity on the road or driving scene (Schmidtke & Bernotat, 1993). In addition, a significantly lower workload presenting the same warnings as in a head-down display has previously been found (Götze & Bengler, 2015). Nonetheless, there is a risk of overloading the display with too much unnecessary information, especially in urban areas where this new display concept approach is presented and compared to presently available solutions.

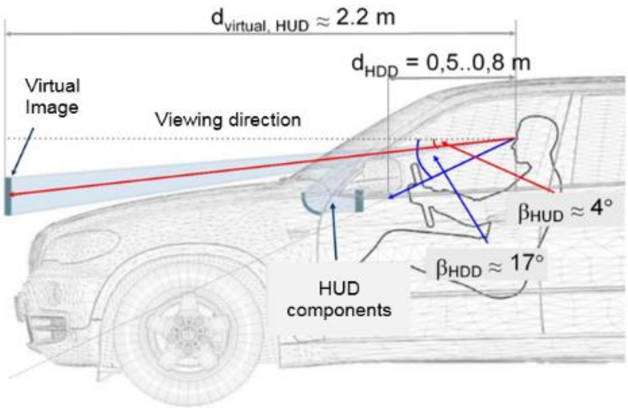


Figure 2. Comparison of different angles and display distances between the head-up and head-down display (Milicic, 2010).

Method

One part of the experiment was conducted using the occlusion technique (ISO 16673:2007). Several studies have showed the validity, reliability and suitability of this method for evaluating or comparing different in-vehicle display design concepts (Baumann et al., 2004; Gelau et al., 2009; Horst, 2004; Noy et al., 2004). Occlusion is classified as a universal tool for measuring the time and accuracy when reading information on HMI displays. The method is used to assess the visual demands of the tasks. The independent variables for this tool are the time period in which a certain stimulus is presented and the time frame between two of those stimuli (Baumann et al., 2004). This study was done by limiting the presentation time of the stimulus on the screen.

The other part of the experiment consisted of two parallel tasks; a choice reaction time task (CRT) and a continuous tracking task (CTT). The aim of this experimental part was to observe the head-up display and to respond to different stimuli using a keypad, while simultaneously doing the CTT on a projection screen. The CRT technique is used to measure reaction times and accuracy when responding to presented stimuli. The method was applied in such a way that the stimulus disappeared after 1500 ms, or after the participant pressed a button to indicate whether or not the answer was correct. Additionally, the time between the stimuli varied between 3000, 5000 and 7000 ms. The CTT is a visual-manual task requiring continuous control by the participants (Eichinger, 2011). The task in this type of CTT was to control the position of a vertically and horizontally moving cross-hair towards a central point using a joystick. The task was used to represent and simulate the driving task and was therefore presented on a projection screen in front of the vehicle.

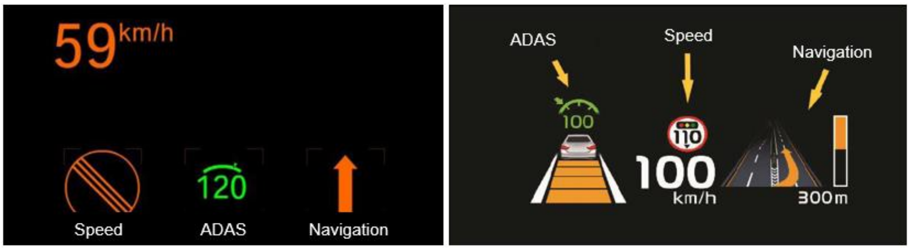


Figure 3. Different current state head-up display designs divided into three clusters; a less detailed version left from Breisinger (2007) and a more detailed one from a Kia K9 Quoris.

Stimuli

Current state head-up display design concepts were researched and categorised. As a result, it was found that the majority are divided into three different clusters. One cluster showed the current speed combined with traffic sign recognition, one displayed navigation information, and the last one provided advanced driving assistance systems information. The order of the three clusters might have differed from brand to brand, but the content was still the same. The one from Breisinger (2007) in Figure 3 was used as an "old" design pattern and is further referred to as "cluster design". The solution for a "new" HUD design concept might be a generic and integrative approach, presenting the relevant information centred and packed so that the driver can gather all the important information at a first glance without first needing to scan the whole head-up display.

Figure 4 shows examples of both of the compared concepts with the categories of speed, navigation, and warning. The question raised for the "speed" category was always about the driven speed being below or above the shown speed limit. The second category, "navigation", was used to ask about the direction of the navigation arrow or the distance to the next turn which resulted in two different targets at the design concept. The last category, "warning", showed two different warning signs (small and large). Not all font and symbol sizes match in both design concepts due to the different approaches of presenting the information. Nevertheless, all the font sizes met the requirements of ISO 15008 (2009) which recommends a minimum size of 0.2°. As an additional measure, a coloured and monochrome version were used for the generic design to examine the influence of colour on reaction times and accuracy.

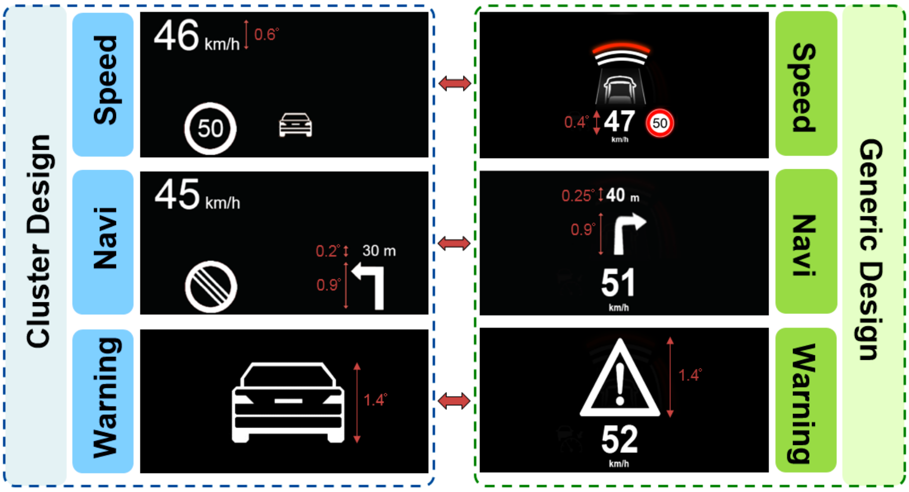


Figure 4. Examples of the different categories for each design concept. The cluster design (left) was inspired by Breisinger (2007). The generic design (right) is a new approach.

Framework Conditions of the Study

The study took place in a stationary 2008 BMW X5 (E70) with a pre-installed series head-up display. The resolution of the HUD was 480x240 px (approx. 20 x 10 cm) and it was able to display the colours red, yellow and orange. In the vehicle, a joystick (right-hand) and a keypad (left-hand) were installed to perform the CRT and CTT. The vehicle had been parked and centred 4.50 m in front of a projection screen (see Figure 5). The distance complied with the 85^{th} percentile of the length of compact-class vehicles from 2010 (Schuster et al., 2011) and simulated the fixation of the brake lights one vehicle length ahead. The light level was always kept in the scotopic range.

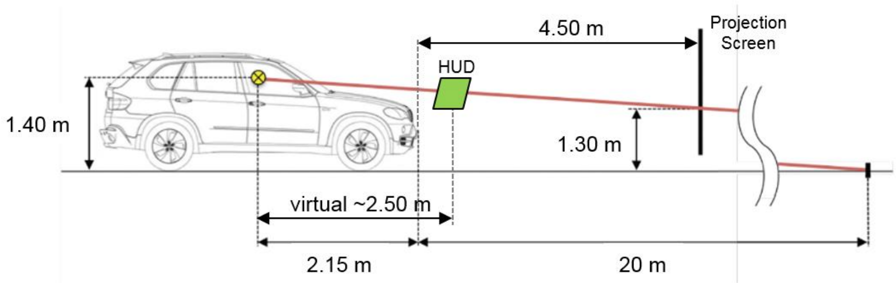


Figure 5. Framework of the study showing the different (virtual) distances between the driver, the vehicle, head-up display and projection screen

The occlusion task consisted of three different stimuli (navigation, speed, and distance) each with four variations. In addition, three cycles were performed resulting in 36 stimuli for each HUD design concept (= 72 stimuli each participant).

For the CRT task, a warning stimuli was added and the distance stimuli removed to fit with the used keypad. Altogether, three different stimuli were included (navigation, speed, and warning symbols) with four variations again for the first two stimuli, along with two variations for the warning symbols. Participants had to do eight cycles of the ten stimuli with an overall 80 CRT stimuli for each HMI design concept (= 160 stimuli each participant).

Procedure

When the participants arrived, all of them performed a visual acuity test (Landolt ring test) to ensure that each participant met the minimum acuity for driving (0.5 according to Colenbrander and De Laey [2005]). This was especially important considering the nature of the experiment in which two design concepts are compared in terms of reading information. Immediately after, participants filled out a demographical questionnaire and were introduced to the experiment and its objective. One of the experimental parts then started in a permuted order. The individual instructions for each part were given at the particular time the task started, always followed by a training session to practice each task. Additionally, the Post Study System Usability Questionnaire (PSSUQ) was filled out during the CRT part. At the end, each participant answered a final questionnaire, rating each concept.

For the occlusion task (see Figure 6), participants read the instructions and question (italic) for each stimuli on the projection screen. It was shown for 3000 ms. Immediately after, a fixation cross appeared for 2000 ms to ensure the eyes were focused on the screen and not yet on the HUD. Finally, the occlusion stimuli was presented in the head-up display for either 200, 250, or 300 ms followed by the question again (bold) until the response was given. Only the accuracy was measured.

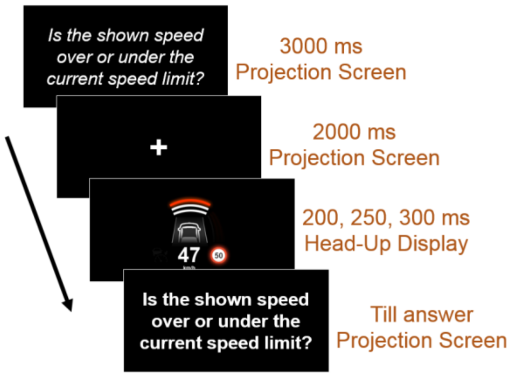


Figure 6. Procedure and duration of the occlusion task on each display.

During the CRT task, participants operated the joystick and simultaneously, stimuli were presented on the HUD for 1500 ms or until a button was pressed. The keypad consisted of four buttons: left, right, up, and down. Participants answered the speed stimuli (to high ↑, to low ↓), the navigation stimuli (arrow left ←, arrow right →), and the warning stimuli (either key is correct).

The PSSUQ is a questionnaire used to measure the usability of a system (Lewis, 2002). The PSSUQ can be classified as a psychometric tool to initially evaluate computer systems. It was developed by IBM (Lewis, 1992) and has a reliability of 0.83-0.96. To match the questionnaire with this study, the phrase “system” was replaced by “display concept” to avoid flawed ratings of different systems.

All displayed questions, instructions or stimuli in this study were prepared and executed with E-Prime 2.0 (Psychology Software Tools, Inc.). The performance metrics (accuracy and reaction time) were also recorded with E-Prime.

Results

Thirty-one healthy volunteers participated in this study (11 women, 20 men). The participants had between 18 and 52 years of age (M = 27.1, SD = 9.6). All of the participants passed the visus test (M = 0.9 visus). One of the participants was excluded from the occlusion task due to technical problems.

Occlusion Task

The mean global accuracy was calculated for 30 participants and all three time-frame conditions. No response, or reporting an inaccurate answer, were considered as errors. A paired-sampled t-test was executed to examine any difference in global accuracy rate. The mean and SDs for each design concept can be found in Table 1. No significant difference was found.

Table 1. Mean global response accuracy for two design concepts using occlusion (N = 30).

	Cluster Design	*Generic Design*
Mean	77.7 %	78.1 %
SD	9.9 %	10.2 %

The mean accuracy for all three different presentation times and both design concepts was calculated. A one-way repeated measure ANOVA was conducted. Mauchly’s test indicated that the assumption of sphericity had not been violated. The results show that there was no significant effect. These results suggest that the design of the display concept has no effect on the accuracy of reading and reporting the information from the HUD. Accuracy percentages and standard deviations are given in Figure 7. In addition, there was a significant main effect of presentation time, $F(2, 28) = 10.945$, $p \leq .001$, suggesting that presentation time significantly affected mean accuracy. This effect is well-known from previous studies (Götze et al., 2013).

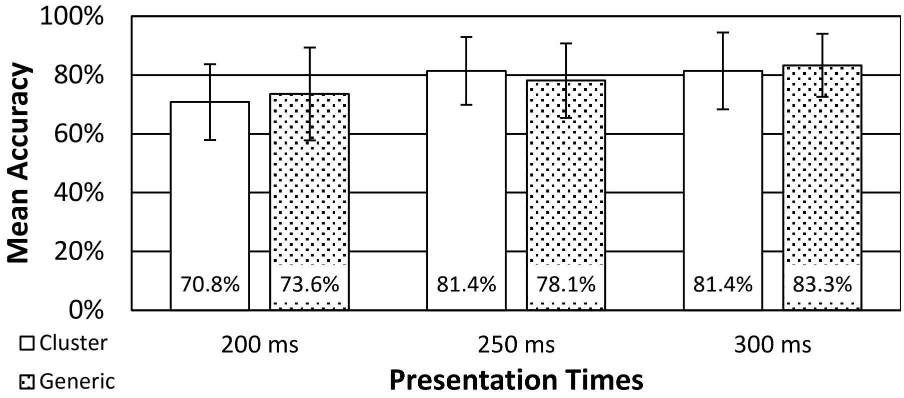


Figure 7. Mean accuracy for each presentation time with each design concept (N = 30).

Choice Reaction Time Task

The mean global accuracy was calculated for all 31 participants and all three stimuli categories. Only correct reaction times were calculated. A paired-sampled t-test was executed to examine any difference in global accuracy rate. The mean and SDs for each design concept can be found in Table 2. Again, no significant difference was found.

Table 2. Mean global response accuracy for two design concepts using a CRT task (N = 31).

	Cluster Design	*Generic Design*
Mean	89.24 %	91.37 %
SD	7.1 %	5.5 %

The reaction times were calculated for all participants and all three stimuli categories. A paired-sampled t-test was executed to examine any difference in global mean reaction time. The mean and SDs for each design concept can be found in Table 3. A significant difference was found between the two designs; $t(30) = 6.5071$, $p \leq .001$.

Table 3. Mean global reaction times for two design concepts using a CRT task (N = 31).

	Cluster Design	*Generic Design*
Mean	954.9 ms	907.9 ms
SD	69.7 ms	64.9 ms

Additionally, all the accuracies were individually calculated for the three different stimuli categories. A one-way repeated measure ANOVA was conducted. Mauchly’s test indicated that the assumption of sphericity had been violated, $X^2(2) = 6.591$, $p = 0.037$; the degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = 0.831$). The results show that there was no significant effect of the design concept on mean accuracy for the different stimuli categories. The means and standard deviation can be found in Table 4. These results suggest, again, that the design of the concept has no effect on the accuracy of reading and reacting to information from a head-up display.

Table 4. Mean accuracy (SD) and mean reaction times (SD) for all three categories (N = 31).

	Speed	*Navigation*	*Warning*
Cluster Acc	80.1 % (14.7)	95.3 % (6.5)	95.8 % (7.2)
Generic Acc	82.3 % (12.0)	96.2 % (4.2)	98.4 % (6.8)

Furthermore, all the reaction times were individually calculated for the three different stimuli categories. Here, again, a one-way repeated measure ANOVA was conducted. Mauchly's test indicated that the assumption of sphericity had not been violated. The test found a significant effect of the design on mean reaction times for the three different categories $F(1, 30) = 45.495$, $p \leq .001$.

This finding made it necessary to execute paired-sampled t-tests comparing each design concept for each category. The results were corrected by the Holm-Bonferroni Sequential Correction (Holm, 1979). All means and the standard deviations are shown in Figure 8. A significant difference was found when comparing the two design concepts in the speed category; $t(30) = 4.278$, $p \leq .001$. In addition, a significant difference was found for the warning category; $t(30) = 6.189$, $p \leq .001$. No significant difference was found for the navigation task. These findings suggest that the generic design concept is beneficial in terms of reaction times compared to the cluster concept.

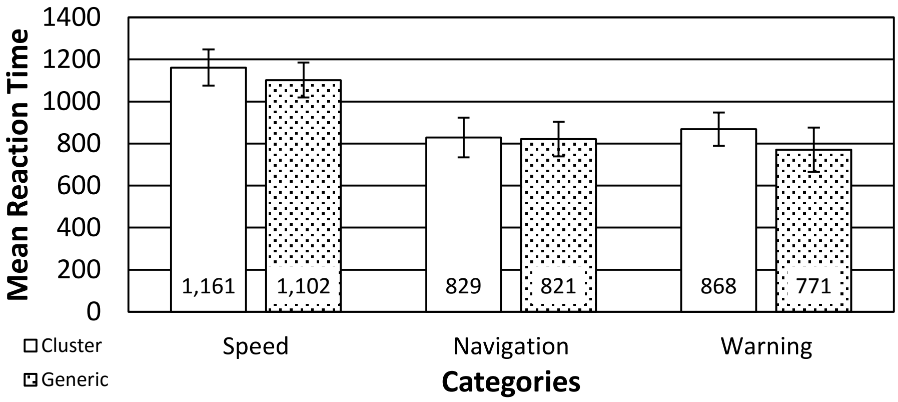


Figure 8. Mean reaction times with standard deviation for all three categories (N = 31).

Questionnaires

For the PSSUQ, the mean of all 19 items was calculated. The overall score for both design concepts can be found in Table 5. A lower score implies a better rating. Executing a pair-sampled t-test found no significant differences.

Table 5. Mean value and standard deviation of the PSSUQ (N = 30).

	Cluster Design	*Generic Design*
Mean	2.33	2.70
SD	0.83	1.23

The final questionnaire included a question about the satisfaction with each design concept, rating it on a seven-point Likert scale. The question was scaled from “I do fully agree” (1) to “I do not agree at all” (7) and therefore a lower score being a better rating. The cluster design was rated 2.77 (1.6) and the generic design with 3.13 (1.7). No significant difference was found.

Discussion

The main aim of this study was to evaluate a new approach to a head-up display design concept for urban driving with a current state version. No difference in terms of colour was found. Looking at the occlusion results, no significant difference was found either in respect to the global accuracy, or for the specific accuracies of the different presentation times. Based on those results, neither of the two design concepts are superior to each other. However, looking at the choice reaction time task, a significant difference in the global reaction time, as well as in the individual reaction time in the three categories has been seen. In detail, responses to the navigation question and to the warning signals resulted in a significantly better result for the generic version. This is particularly interesting, since the font and traffic sign size is smaller in the generic version. Despite that, again, no difference has been found in all three categories in respect to the accuracy between the two design concepts. Participants were equally accurate, but with different reaction times.

Conclusion

Neither design concept seems to be ultimately superior when compared with each other. Still, both HUD designs do show benefits in individual places. Further research needs to be done to assist the driver with urban driving in the best possible way. Moreover, additional ADAS need to be added and considered when evaluating the concept. Furthermore, in order to follow the integrative approach (Götze et al., 2015), the adding of more in-vehicle components is necessary which might then result in different design options.

In conclusion, the study shows the potential of a new generic design approach for head-up displays in urban driving. While the reaction times were significantly shorter with the new concept compared to the cluster design, the subjective rating showed some contrary effects. In a future study, the benefits of both design concepts will be considered to build an even more suitable approach to urban driving.

References

Baumann M., Keinath, A., Krems, J.F., & Bengler, K. (2004). Evaluation of invehicle HMI using occlusion techniques: experimental results and practical implications, *Applied Ergonomics, 35*, 197-205.

Breisinger, M. (2007). *Anzeigekonzept für Head-Up Displays im Automotive Bereich* [HMI Concepts for Head-Up Displays in the Automotive Sector]. Diploma thesis, Ludwig-Maximilians Universität. Munich, Germany: Department Institute for Informatics.

Bubb, H. (1978). *Optical signaller to denote safe distance of car - has light strip which is projected on windscreen, width of which corresponds to safe distance.* Patent Identifier No. DE 2633067 A1.

Colenbrander, A., De Laey, J.-J. (2005). *Vision Requirements for Driving Safety – International Councill of Ophthalmology*. Report prepared for the International Council of Ophthalmology at the 30th World Ophthalmology Congress. Sao Paulo, Brazil, February 2006.

Eichinger, A. (2011). *Bewertung von Benutzerschnittstellen für Cockpits hochagiler Flugzeuge* [Evaluation of Man-Machine-Systems in highly agile Aircraft Cockpits]. PhD thesis, Universität Regensburg. Regensburg, Germany: Fakultät II Psychologie, Pädagogik und Sportwissenschaft.

Gelau, C., Krems, J.F., Henning M.J. (2009). On the realiability of the occlusion technique as a tool for the assessment of the HMI of in-vehicle information and communication systems, *Applied Ergonomics*, *40*, 181-184.

Gevatter, H.-J., & Grünhaupt, U. (2006). *Handbuch der Mess- und Automatisierungstechnik in der Produktion* [Handbook of Measurement Technology and Automation Technology in Production] (2., vollständig bearbeitete Aufl). *VDI-Buch*. Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.

Götze, M., Conti, A.S., Keinath, A., Said, T., & Bengler, K. (2013). Evaluation of a New Cockpit Color Concept under Mesopic Lighting for Urban Driving. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Noar, O. Nierstrasz, C. P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, G. Weikum (Eds.), *Lecture Notes in Computer Science. Design, User Experience, and Usability. Web, Mobile, and Product Design* (pp. 359–366). Berlin, Heidelberg: Springer Berlin Heidelberg.

Götze, M., Bißbort, F., Petermann-Stock, I., & Bengler, K. (2014). “A Careful Driver is One Who Looks in Both Directions When He Passes a Red Light” – Increased Demands in Urban Traffic. In D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Noar, O. Nierstrasz, C. P. Rangan, B. Steffen, M. Sudan, D. Terzopoulos, D. Tygar, G. Weikum (Eds.), *Lecture Notes in Computer Science. Human Interface and the Management of Information. Information and Knowledge in Applications and Services* (pp. 229–240). Cham: Springer International Publishing.

Götze, M., & Bengler, K. (2015). Urban Driving: Where to Present What Types of Information – Comparison of Head-Down and Head-Up Displays. In S.

Yamamoto (Ed.), *Lecture Notes in Computer Science. Human Interface and the Management of Information. Information and Knowledge in Context* (pp. 190–200). Cham: Springer International Publishing.

Götze, M., Ruff, F., & Bengler, K. (2015). Optimal Information Output in Urban Traffic Scenarios: An Evaluation of Different HMI Concepts. *Procedia Manufacturing*, *3*, 739–746.

Holm, S. (1979). A simple sequential rejective method procedure. *Scandinavian Journal of Statistics, 6,* 65-70.

Horst, R. van der (2004). Occlusion as a measure for visual workload: an overview of TNO occlusion research in car driving, *Applied Ergonomics, 35*, 189–196.

ISO 15008 (2009). *Road vehicles -- Ergonomic aspects of transport information and control systems -- Specifications and test procedures for in-vehicle visual presentation*. Berlin: DIN Deutsches Institut für Normung e. V.

ISO 16673 (2007). *Road vehicles -- Ergonomic aspects of transport information and control systems -- Occlusion method to assess visual demand due to the use of in-vehicle systems.* Berlin: DIN Deutsches Institut für Normung e. V.

Lewis, J.R. (1992). Psychometric Evaluation of the Post-Study System Usability Questionnaire: The PSSUQ, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 36*, 1259–1263.

Lewis, J.R. (2002). Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *International Journal of Human-Computer Interaction, 14*, 463–488, Lawrence Erlbaum Associates, Inc.

Miličić, N. (2010). *Safe and ergonomic use of Head-Up Displays in the vehicle. Dissertation*, PhD thesis, Technische Universität München. Munich, Germany: Department of Electrical and Computer Engineering.

Hortst, Y.I., Lemoine, T.L., Klachan, C., & Burns, P.C. (2004). Task interrupability and duration as measures of visual distraction, *Applied Ergonomics*, *35*, 207–213.

Petermann-Stock, I., & Rhede, J. (2013). Intelligente Strategien für nutzerzentrierte MMI Konzepte im urbanen Raum [Intelligent Strategies for user-centered HMI Concepts in Urban Space]. In VDI Gesellschaft Fahrzeug- und Verkehrstechnik (Hrsg.), *Der Fahrer im 21. Jahrhundert* (VDI-Berichte, Vol. 2205, pp. 263-286).

Recarte, M.A., & Nunes, L.M. (2003). Mental Workload While Driving: Effects on Visual Search, Discrimination, and Decision Making, *Journal of Experimental Psychology: Applied, 9*, 119–137.

Rumar, K. (1990). The basic driver error: Late detection. *Ergonomics, 33*, 1281-1290.

Schmidtke, H., & Bernotat, R. (1993). *Ergonomie* (3., neubearb. und erw. Aufl). München [u.a.]: Hanser.

Schröder, T. (2012). *Analytische Betrachtung der Auswirkungen komplexer Verkehrssituationen* [Analytical Review of the Impact of complex Traffic Situations]. Diploma thesis, Technische Universität München. Munich, Germany: Department of Mechanical Engineering.

Schuster, A., Stattler, J., & Hoffmann S. (2011). *Bestimmen der aktuellen Abmessungen differenzierter Personen-Bemessungsfahrzeuge* [Determining the Dimensions of differentiated Measuring Vehicles]. Westsächsische

Hochschule Zwickau. Retrieved October 30, 2015, from: http://www.fh-zwickau.de/fileadmin/ugroups/kt/iev/projekte/FoBemPkw.pdf

Statistisches Bundesamt Wiesbaden (2014). *Fehlverhalten der Fahrzeugführer bei Unfällen mit Personenschaden* [Miscoduct of Drivers in Accidents involving Personal Injury]. Retrieved October 30, 2015, from: https://www.destatis.de/DE/ZahlenFakten/Wirtschaftsbereiche/TransportVerkehr/Verkehrsunfaelle/Tabellen/FehlverhaltenFahrzeugfuehrer.html

Winner, H., Hakuli, S., & Wolf, G. (2012): *Handbuch Fahrerassistenzsysteme: Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort* [Handbook of Driver Assistance Systems: Basics, Components and Systems for Active Safety and Comfort] (Vol. 2), Wiesbaden: Vieweg + Teubner.

Wittmann, M., Kiss, M., Gugg, P., Steffen, A., Fink, M., Pöppel, E., & Kamiya, H. (2006). Effects of display position of a visual in-vehicle task on simulated driving. *Applied Ergonomics, 37,* 187-199.

Towards developing a head-up display warning system - How to support older drivers in urban areas?

Juela Kazazi, Susann Winkler, & Mark Vollrath
Technische Universität Braunschweig,
Germany

Abstract

Driver warning systems are essential, when pursuing safer urban areas. The complexity here is very high, which is a problem especially for older drivers as they are over-represented in urban crashes. The aim of this driving simulator study (two experimental groups and one control group with respectively 12 older drivers, aged > 65 years) was to determine what kind of head-up warnings (between-subjects factor: Stop sign warning (SW), Caution sign warning (CW)) might have the best effect on the driving performance of older drivers (compared to control group) in scenarios with different criticalities. The results show that in most scenarios, the brake reaction times in the SW group were significantly shorter compared to the CW and control group. Furthermore, the SW led to the highest maximum braking value, whereas the CW group led only to somewhat higher maximum braking values as compared to the control group. The SW warning is recommendable for critical scenarios, which demand an immediate driver reaction. In less critical situations, it might be sufficient to raise the drivers' attention, which is why the CW should be triggered. Accordingly, a two stage warning system combining both strategies (warning and acute warning) is being tested in further studies.

Introduction

In contrast to rural areas, the interaction of drivers with other road users is much more frequent in urban areas. Here the density of repeatingly crossing road users is at its maximum. In these areas, drivers have to continuously divide their attention between various objects of interest (e.g., oncoming vehicles, vulnerable road users). Overall, accident statistics indicate that more critical traffic situations, like intersections, lead to a higher death rate (Statistisches Bundesamt, 2012; Morgenroth et al., 2009). Furthermore, accidents in urban areas are relevant for older drivers (> 65 years) as they are over-represented in these crashes (McGwin & Brown, 1999; Evans, 2004). This group of drivers seems to have the greatest difficulty negotiating in highly critical situations, as indicated by their high percentage of crashes, which

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

might be due to a weaker driving performance as to factors such as declines in vision, hearing, reaction time and cognitive functions (Marshall et al., 2010).

As far as in-vehicle technology is concerned, warning systems are being developed to improve safety in driving. They especially aid drivers who are at greater risk of crashes and significantly reduce the number of fatalities (Marshall et al., 2010). These systems are of great benefit in notifying the driver with regard to own lack of attention, guiding the driver's attention to critical objects and in supporting the driver to keep safe distances to cars in front (Staubach, 2009; Alm & Nilsson, 1995). The extent of changes in behaviour due to such systems is dependent on how drivers detect, understand and particularly on the design of the human–machine interface (Weller & Schlag, 2004). Unfortunately, human factors which may limit system performance have not been taken into consideration (Kantowitz, 2000; Hancock & Verwey, 1997), since system effectiveness does not only depend on the design of the system, but also on the joint performance of system and driver. Until now, most of the studies on supporting drivers with warning systems are looking for a unique function, valid for all drivers, in all possible situations, leaving the age of the driver as well as different complexities of situations disregarded. Yet, when looking at urban areas it becomes clear that not all situations are of the same criticality. Situations with suddenly crossing pedestrians are more critical than for instance a sudden braking lead vehicle, when the distance between the vehicles is large.

From this perspective, different critical situations demand drivers to either slow down or stop their vehicle to avoid a collision. For example in highly critical situations the best reaction may be an emergency braking, which might be induced by a stop sign, since this sign is coupled with coming to a stop. However, less critical situations might merely require the driver to slow down gradually, which might be achieved through a caution sign, since this sign suggests being attentive. Through this, the question arises what kind of warning type might have the best effect on driver performance in situations with different criticalities. As part of the research project UR:BAN (Manstetten et al., 2013; www.urban-online.org), the aim of this study was to first create scenarios with various difficulties in order to find out what warning types have positive effects on the driving performance of older drivers in different critical situations and second which warning type is best suited for highly critical, critical and moderately, critical scenarios.

For the current study the warning types were presented in a Head-up display (HUD) since many advantages of using a HUD for that (e. g. shorter glance duration to the HUD, drivers are not forced to taking their eyes off of the road) have been outlined by previous research (Ablassmeier et al., 2007). Traffic signs were used as warning symbols, as these are known to all drivers from driving license training and everyday traffic (Alves et al., 2013; George et al., 2012; Plavsic et al., 2009). It was expected that a stop sign warning (SW) would encourage the driver to an emergency braking action, whereas a caution sign warning (CW) would lead to slowing down.

The current study was conducted in a driving simulator including four scenarios, varied by the criticality (highly critical, critical and moderately critical) as well as the characteristics of the critical object (e.g., pedestrian, vehicle and obstacle). Variations of scenarios as well as the types of warning will be described in more detail in the method section (*Driving Scenarios and Simulation*).

Method

Participants

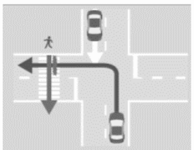
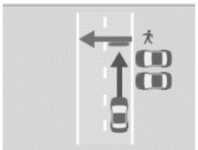
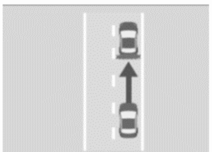
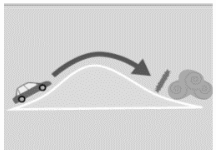
A total of 36 subjects (29 male, 7 female) were divided into two experimental conditions (SW, CW) and a control condition. Each condition consisted of 12 subjects. The mean age of participants in this study was 71.9 years (SD = 4.4 years), owning their driving license on average for 49.7 years (SD = 7.7 years) and driving about 12000 km annually. Subjects were recruited from a database of older drivers of the Technische Universität Braunschweig. Subjects were trained (on the same day as the experiment to increase compliance) in the driving simulator of the Technische Universität Braunschweig. The training contained three driving roads. In the first, the participants drove on a straight rural road, followed by an urban road, to enable participants to train acceleration and deceleration. The last urban road contained intersections where the participants had to turn left or right, in order to get familiar with this behaviour. Participants who developed simulator sickness had to be excluded from the study (n = 16 out of 52 trained participants overall). In the study, all remaining participants (n = 36) had normal or corrected-to-normal vision. Participants were compensated for their time by either choosing to receive a box of chocolates or eight Euros an hour.

Driving Scenarios and Simulation

Table 1 gives an overview of all scenarios and their criticality. In each scenario the ego vehicle, driven by the participant, travelled through a simulated urban road and had to make a turn when it was indicated by a voice output and an arrow in the speedometer. In the *Pedestrian 1* scenario, the driver had to make a left turn at an intersection. While turning, a pedestrian crossed the ego vehicle´s road. In the *Pedestrian 2* scenario as a pedestrian, who was hidden by parking cars, suddenly crossed the ego vehicles path. In the *Vehicle* scenario, a lead vehicle came to a sudden stop. In the *Obstacle* scenario, the driver was confronted with a hay bale hidden behind a hill. The different scenarios, as seen in Table 1, were created using the driving simulation software SILAB (Krüger, Grein, Kaussner & Mark, 2005; see www.wivw.de).

This study was conducted in the fixed base driving simulator of the Department of Engineering and Traffic Psychology at the Technische Universität Braunschweig. It consists of a seat box with a steering wheel with force feedback, accelerator and brake pedals and two LCD screens serving as rear-view mirrors. The virtual scenery is projected onto three screens (left, ahead, right), providing the drivers with a 180° field of view at about 2.1 m distance from the driver’s seat. Using driving simulators makes it possible to control variables in the scenery as well as to accurately measure driving performances, which are difficult to survey in the field. Furthermore, an evaluation of warning systems in critical situations cannot easily be achieved in on-road studies, since drivers might be harmed.

Table 13: Description of scenarios used in the study

Scenario	Pedestrian 1	Pedestrian 2	Vehicle	Obstacle
Picture				
Criticality	Critical scenario	Very critical scenario	Moderately critical scenario	Moderately critical scenario

Warning types

For the present study two different kinds of traffic signs were used as warning symbols. The warning was presented in a HUD, projected over the roadway, not obscuring the view of drivers. The size of the stop sign warning (SW) in the HUD was 19x17 cm; the caution sign warning (CW) was 21x19 cm. The presentation of the different warning symbols happened about 2.5 s before the critical event occurred (e.g., crossing pedestrian, braking vehicle). The duration of the warning was individual for every subject, since the onset and the offset of the warning was triggered at particular flow points. In the present study there were two experimental conditions, the SW and the CW, as well as a control condition receiving no warning at all. It was expected that these two different warning signs would lead to different driver reactions. Table 2 gives an overview of the three warning conditions.

Table 14: Warning types and warning symbols

Warning types	*Presentation in HUD*
Control (C)	No warning
Stop sign warning (SW)	
Caution sign warnings (CW)	  

Procedure

After reading and signing a consent form, the participants were instructed in written form about the procedure of the experiment. Drivers were told to drive as they normally would in their own vehicles and to obey all traffic rules (e.g., speed limits).

Next, drivers completed a training drive (lasting about 25 min) in order to get familiar with the simulator and prevent simulator sickness. The test drive began after the training drive. During the experiment the researcher was seated in a separate room, having the opportunity to communicate with the participant via a microphone. The test drive lasted about 15 minutes. Afterwards, the subjects were asked to answer two questions concerning the criticality and surprise of each scenario. Next they were debriefed about the purpose of the study, reimbursed and thanked for their participation. The overall duration of the trial was about 1.5 hours. The order of the scenarios in the test drive was not counterbalanced, since the effects of the warning types were of interest (order of scenarios: *Pedestrian 1*, *Pedestrian 2*, *Vehicle*, *Obstacle*).

Data analysis

Driving data was logged by the simulation software SILAB (Krüger et al., 2005). Participant's subjective data was logged by using a two-stage rating scale (see Table 3, 15-point rating scale, Heller, 1982). First, one of the five labelled categories (*low* to *high*) was chosen and then refined by choosing one out of three subcategories (-, 0, +), which were later transformed into numbers from 1 to 15. Table 4 gives a summary of the dependent variables (Driving and Subjective Data) measured.

Table 3. Subjective rating scales to measure the dependent variables criticality and surprise of the four scenarios (Heller, 1982).

low			rather low			moderate			rather high			high		
-	0	+	-	0	+	-	0	+	-	0	+	-	0	+

Table 4: Dependent variables regarded in this study

Variable		*Unit*	*Description*
Driving Data	Brake reaction time	s	Reaction time to pressing the brake pedal after warning was triggered
	Maximum braking value	%	The maximum braking value reached by participants
	Mean velocity at maximum braking value	km/h	Mean velocity participants had at their maximum braking value in the different groups and scenarios
Subjective Data	Criticality	1…15	"How critical was the experienced scenario?"
	Surprise	1…15	"How surprising was the experienced scenario?

For data analysis IBM SPSS Statistics 22 was used. A Kruskal-Wallis test was conducted to evaluate the subjective data, as well as an ANOVA with one within-subject factor (Type of Warning) for the driving data. A significance level of alpha $p = .05$ was adopted for all statistical tests. The error bars in the figures represent the standard deviation.

Results

Different criticalities of scenarios

In order to find out if the implemented scenarios were indeed of different criticalities (highly critical, critical, and moderately critical) a Kruskal-Wallis test was conducted to evaluate the differences among the four scenarios on median change in the subjective rating of participants surprise in each scenario. The subjective data in Figure 1 show significant differences between the four scenarios when looking at the criticality (χ^2 (3, N = 124) = 22.86, $p < 0.001$) and surprise rating (χ^2 (3, N = 122) = 16.66, $p < 0.001$). The *Pedestrian 2* scenario was rated as being a highly critical and highly surprising scenario, followed by the critical and surprising *Pedestrian 1* scenario. The obstacle and vehicle scenario were rated as being moderately critical and moderately surprising.

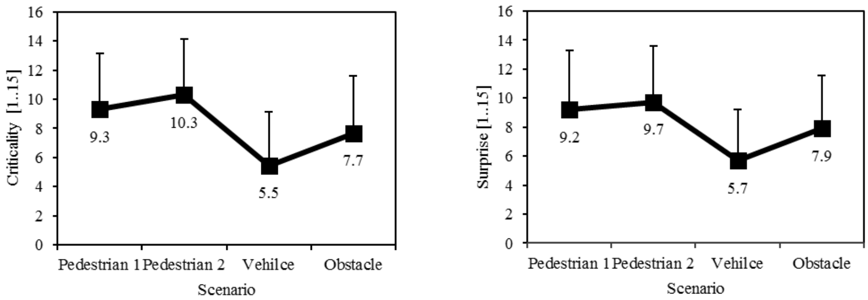


Figure 1. Subjective ratings (mean) of criticality (left) and surprise (right) in each scenario, including all participants.

Do warnings have a positive effect on brake reaction time?

The brake reaction time following the warning was recorded. It was expected that the experimental conditions (SW, CW) would have shorter brake reaction times compared to the control condition (C). For this analysis not all participants were considered. If subjects had only left the gas pedal without pressing the brake pedal, they were excluded from the examination, leading to dissimilar numbers of participants in the groups.

During the *Pedestrian 1* scenario, drivers had a very low velocity when making a left turn (M = 4.7; SD = 2.0). This influenced the brake reaction, in that drivers did not always have to press the brake pedal to avoid an accident with the pedestrian, this possibly being a reason why the brake reaction time in this scenario did not differ significantly between the three conditions (see Fig. 2 left). In the *Pedestrian 2* scenario (see Fig. 2 right) there were significant differences when considering the brake reaction time ($F_{(2,35)} = 4.06$, $p = 0.017$). Furthermore post-hoc tests showed that the control condition (C) had significantly slower brake reactions compared to the SW condition (p = 0.005). Both experimental conditions did not differ significantly in their brake reaction (p = 0.316).

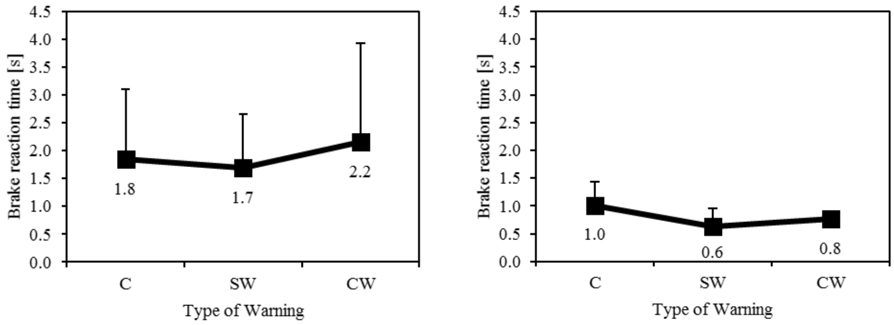


Figure 2. Mean brake reaction times in Pedestrian 1 (left) and Pedestrian 2 (right) scenario.

No overall significant differences concerning the brake reaction time were found in the *Vehicle* scenario. However, when looking at post hoc tests, there was a significant difference between the control and SW condition (p = 0.040), where the SW condition had significantly faster brake reactions compared to the control condition (see Fig. 3 left).

In the *Obstacle* scenario, there was an overall significant difference in the brake reaction time ($F_{(2,35)}$ = 6.36, p = 0.005). Post-hoc tests also showed that the control condition led to slower brake reactions compared to the SW condition (p = 0.001). Figure 3 (right) gives an overview of the brake reaction times in the *Obstacle* scenario.

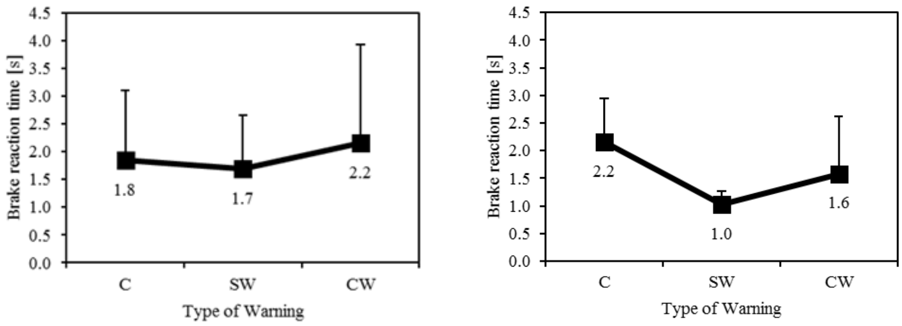


Figure 3. Mean brake reaction times in Vehicle (left) and Obstacle (right) scenario.

Do warnings have a positive effect on the maximum braking value?

After participants pressed the brake pedal, the maximum braking value reached was recorded. This variable gives information about the sturdiness (in %) of pressing the pedal.

For scenarios, *Pedestrian 1* and *Pedestrian 2* there were no significant differences between the three types of warnings. Figure 4 gives an overview of the maximum braking value in the *Pedestrian 1* (left) and *Pedestrian 2* (right) scenarios.

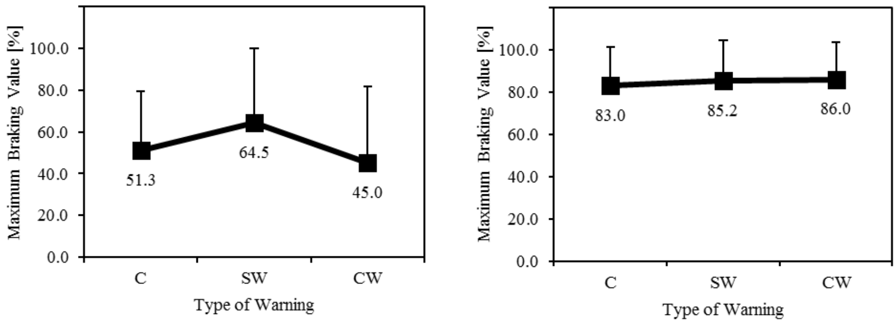


Figure 4. Mean maximum braking value in the Pedestrian 1 (left) and Pedestrian 2 (right) scenarios.

While there were no significant differences in the *Vehicle* scenario, there were significant differences in the *Obstacle* scenario ($F_{(2,35)} = 15.67$, $p < 0.001$). Post-hoc test in the *Obstacle* scenario revealed significant differences between the control and the SW condition ($p < 0.001$) as well as between the SW and the CW condition ($p < 0.001$). Figure 5 gives an overview of the maximum braking value in the *Vehicle* (left) and *Obstacle* (right) scenario. Overall, the SW condition had usually the highest maximum braking value in every scenario (except *Pedestrian 2*).

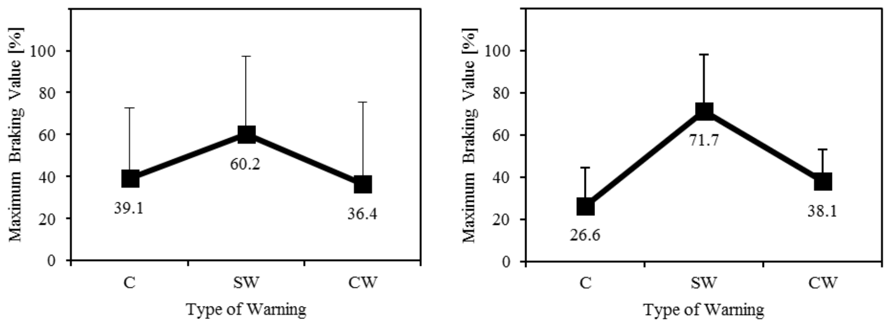


Figure 5. Mean maximum braking value in the Vehicle (left) and Obstacle (right) scenario.

Do warnings have a positive effect on the mean velocity at maximum braking value?

When participants reached their maximum braking value, also their velocity was recorded. This was done to see if participants had different velocities and if one of the experimental conditions rather led to slowing down more.

In the *Pedestrian 1*, *Pedestrian 2* as well as in the *Vehicle* scenario there were no significant differences in the velocity at the maximum braking value in the three conditions. When looking at the *Obstacle* scenario there was a significant difference in the velocity at the maximum braking value ($F_{(2,35)} = 12.94$, $p < 0.000$). As seen in Figure 6, the SW condition had a lower mean velocity at the maximum braking value compared to the control condition ($p < 0.001$), as well as compared to the CW condition ($p < 0.001$).

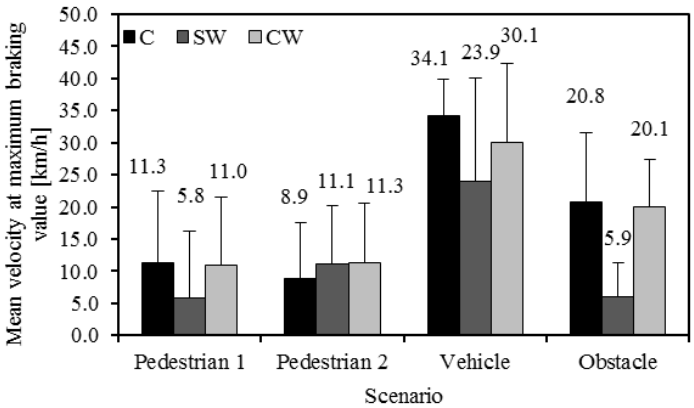


Figure 6. Mean velocity at maximum braking value in all scenarios.

Conclusions

The aim of this work was to create scenarios of different criticalities and to determine if warning systems have a positive effect on the driving performance of older drivers in urban areas.

The subjective data demonstrated that participants rated the *Pedestrian 2* and *Pedestrian 1* scenario as being (highly) critical and (highly) surprising, revealing that the four implemented scenarios were indeed of different criticalities. As of the results following the driving data, the brake reaction time in the SW condition was the fastest in every scenario, followed by the CW condition. In this study the SW condition had a positive effect on the shortening of the brake reaction time. Similar results are found when considering the maximum braking value. This variable was descriptively the highest in the SW condition, meaning that participants hit the brake pedal the strongest here compared to the CW and the control condition.

Furthermore, the SW condition led to almost always the lowest mean velocity when reaching the maximum braking value.

There are some limitations to this study that need to be mentioned. To test warning systems and their effect on accident reduction, it is important that critical situation occur. In this study though, older drivers had a very low amount of accidents. These results are in contrast to the findings from literature of an overrepresentation of older drivers in accidents. One possible explanation might be that the scenarios here were still not critical enough and need to be adjusted. Another reason might be the use of a driving simulator itself that contributes to these findings. Older drivers might have been especially cautious, as the driving simulator was new to them, leading to a compensatory behaviour, which they may not show in real traffic. In order to examine these factors closely, field studies are needed. Moreover, a selection bias might have occurred, since the old drivers volunteering for this study might be those that are in good shape and feel competent to drive. Besides, the warning in the *Pedestrian 2* scenario seems to have been triggered to late, since the results of all variables are almost the same in the three warning conditions. This scenario set up might need to be fine-tuned.

In summary, the results of the present study showed that it is important to know what behaviour a warning might trigger. In scenarios where the driver has to react immediately, the caution sign warning (CW) does not have a significant effect on the brake reaction time and maximum braking value. In such situations, the stop sign warning (SW) would be more suited. Examples of these kinds of situations are the *Pedestrian 1* and *Pedestrian 2* scenarios, in which a fast brake reaction is needed to avoid an accident. Yet, if a sudden and firm brake reaction is not needed, as in the *Vehicle* and *Obstacle* scenario, the SW warning could have negative effects (e.g., high maximum braking value). When a strong brake reaction is triggered other road users might perceive this behaviour as traffic blocking, leading to hazardous behaviours of other drivers (Dotzauer, 2013). In rather moderately critical scenarios leaving the gas pedal or slightly pressing the brake pedal might be sufficient, which might be by the CW.

The two different warning types in this study are to be used in different critical situations. Moreover, it is possible to create a warning cascade that combines these two warning types. For example, if a situation only requires the attention of the driver, the CW warning is triggered. If the driver reacts for example with a slight braking and the situation is not about to get more critical, the SW warning will not follow. Yet, if the CW warning did not raise the attention of the driver, it is possible to trigger an SW warning, so that the driver reacts immediately and a collision can be prevented. This warning cascade will be examined in future studies.

References

Ablassmeier, M., Poitschke, T., Wallhoff, F., Bengler, K., & Rigoll, G. (2007) Eye gaze studies comparing head-up and head-down displays in vehicles. In *2007*

IEEE International Conference on Multimedia and Expo (pp. 2250-2252). Beijing.

Alm, L., & Nilsson, H. (1995) The effects of a mobile telephone task on driver behaviour in a car following situation. *Accident Analysis and Prevention, 27,* 707-715.

Alves, P.R.J.A., Goncalves, J., Rossetti, R.J.F., Oliveira, E.C., & Olaverri-Monreal, C. (2013). Forward collision warning systems using heads-up displays: Testing usability of two new metaphors. Intelligent Vehicles Symposium Workshops (IV Workshops), IEEE International Conference on Communications (pp. 23–23), Budapest, Hungary.

Dotzauer, M. (2013). Researching safety issues with intersection assistance systems for the older driver. In A. Stevens, J. Krems, and C. Brusque (Eds.). *Driver adaptation to information and assistance systems* (pp. 123-143). Stevenage, UK: IET Institution of Engineering and Technology.

Evans, L. (2004). *Traffic Safety*. Science Serving Society, Bloomington, Hills.

George, P., Thouvenin, I., Frémont, V. & Cherfaoui, V. (2012) DAARIA: Driver assistance by augmented reality for intelligent automobile. In *2012 IEEE Intelligent Vehicles Symposium,* (pp. 1043–1048) Michigan, USA.

Hancock, P.A., & Verwey, W.B. (1997). Fatigue, workload and adaptive driver systems. *Accident Analysis & Prevention, 4,* 495-506

Heller, O. (1982). Theorie und Praxis des Verfahrens der Kategorienunterteilung (KU) [Theory and Practice of the subdivision process of categories]. In O. Heller (Ed.), *Forschungsbericht 1981* (pp. 1-15). Würzburg, Germany: Psychologisches Institut der Universität Würzburg.

Kantowitz, B.H. (2000). Attention and mental workload. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Human Factors and Ergonomics Society* (pp. 456-460), Santa Monica, USA

Krüger, H.-P., Grein M., Kaussner, A., & Mark, C. (2005). SILAB-A task-oriented driving simulation. In *2005 DSC Driving Simulator Conference,* (pp. 232-331). Orlando, USA

Manstetten, D., Bengler, K., Busch, F., Färber, B., Lehsing, C., Neukum, A., Petermann-Stock, I., & Schendzielorz, T. (2013). “UR:BAN MV Human Factors in Traffic” – a German Research Project to Increase Safety in Urban Traffic. In *Proceedings of the 20th ITS World Congress* (pp. 596-605). Tokyo, Japan.

Marshall, D.C., Wallace, R.B., Torner, J.C. & Leeds, M.B. (2010). Project Report: *Enhancing the Effectiveness of Safety Warning Systems for Older Drivers* DOT HS 811 417

McGwin Jr., G., & Brown, D.B. (1999). Characteristics of traffic crashes among young, middle-aged, and older drivers. *Accident Analysis and Prevention 31,* 181-198.

Morgenroth, J., Wolf, L., Macht, S., Sasse, A., Hecker, P., Schack, M., Vollrath, M., Seiffert, U., Proskawetz, K., & Otte, D., 2009. Improving the Safety of Pedestrians by Using a Cooperative System. In: Proceedings of the 9th International Conference on Intelligent Transportation Systems (ITS) Telecommunications, Lille, France

Plavsic, M., Bubb, H., Duschl, M., Tönnis, M., & Klinker, G. (2009). Ergonomic Design and Evaluation of Augmented Reality Based Cautionary Warnings for Driving Assistance in Urban Environments. In *Proceedings of Intl. Ergonomics Assoc*.

Statistisches Bundesamt, 2012. Verkehr. Verkehrsunfälle 2011. Wiesbaden: Statistisches Bundesamt.

Staubach, M. (2009). Factors correlated with traffic accidents as a basis for evaluating advanced driver assistance systems. *Accident Analysis and Prevention, 41,* 1025-1033.

Weller, G., & Schlag., B. (2004). Verhaltensadaptationen nach Einführung von Fahrerassistenzsystemen: Vorstellung eines Modells und Ergebnisse einer Expertenbefragung [Behavioral adaptations after the introduction of driver assistance systems: presentation of a model and results of an expert survey]. In *Schlag, B. (Hrsg.): Verkehrspsychologie. Mobilität – Sicherheit – Fahrerassistenz (pp. 351 - 370). Lengerich: Pabst Science Publ.*

The effect of external and environmental factors on perceived comfort: the car-seat experience

Alessandro Naddeo, Rosaria Califano, Nicola Cappetti, & Mariarosaria Vallone
Dept. of Industrial Engineering, University of Salerno
Italy

Abstract

Today, comfort and discomfort related to the automobile seat are widely studied. A previous work, published by this study's authors (Naddeo et al., 2014a), affirmed that the evaluation of the perceived comfort associated with the driving experience could not be performed by considering only the driver's seat. The authors offered a theoretical matrix to evaluate the comfort of car seats through identification of all involved aspects and the interactions with external factors inside the vehicle (personal, task and environment characteristics). To verify this hypothesis, a sample of subjects evaluated a car seat by interacting with it in five ways: interaction with the real prototype, presentation of a real prototype, photographic presentation of a real prototype, presentation of a 3D virtual digitized prototype, and interaction with the real prototype placed inside the car. The Kansei technique has been used as an evaluation tool to assess the individual and subjective emotional impressions of the car seat where all the senses of the consumer are involved. The results of the study show how the same object looks different if evaluated in different ways and the effects of external and environmental factors on the perceived comfort.

Introduction

Thousands of people experience comfort or discomfort on their automobile seat daily, especially drivers on long trips. Comfort during the task of driving is a multifaceted phenomenon that is not exclusively related to the seat but to all elements that interact with the driver. All these elements can influence the driving comfort experience and, in particular, the user's sensation of comfort about the seat. To test this hypothesis, the same seat was represented in different ways and in different environments, and evaluated by a sample of subjects.

Naddeo and Cappetti model of comfort perception

The authors' previous work represented the comfort experience using a big matrix in which the comfort-related elements were classified and studied. According to the Naddeo and Cappetti (NC-model) model of perception (Naddeo et al., 2014a), the experience of comfort or discomfort in a generic environment was represented by

In D. de Waard, K.A. Brookhuis, A. Toffetti, A. Stuiver, C. Weikert, D. Coelho, D. Manzey, A.B. Ünal, S. Röttger, and N. Merat (Eds.) (2016). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2015 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

the logic sum of four main aspects that contribute to HMI description and classification: Person (Pe), Product Characteristics (Pr), Task and Usage (T), and Working Environment (WE) as in Figure 1.

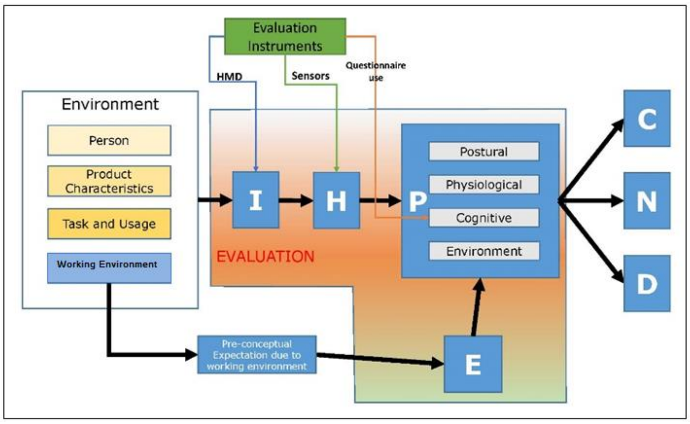


Figure 1. Naddeo and Cappetti model of comfort perception.

The interaction (I) with an environment is caused by the contact (also nonphysical contact) with a consequent internal human body effects (H). The perceived effects (P) are influenced by the human body effects but also by expectations (E). These are interpreted as "comfortable" (C) or "you feel nothing" (N), or they can lead to "feelings of discomfort" (D) (Vink & Hallbeck, 2012).

In Naddeo et al. (2015c), it's demonstrated that if the product has a particular design and therefore a higher cost than other products, the users have higher expectations, thereby influencing the comfort perception. Another aspect that cannot be underestimated, also integrates this model: the perception modification due to experimental devices needed to evaluate comfort. Each element of the NC-model has been deployed and divided in primary or modifier elements. The primary elements (P) are the ones that weigh on the real interaction ability of a person while the modifier elements (M) weigh on the perception ability and are related to person and environment characteristics. For each kind of interaction (I), one or more human body effects (H) have been found. All data have been organized in a large comfort-matrix (see Table 1).

Driving model: Literature review helps build the NC-matrix model

Naddeo at al., (2014a) have identified a suitable application of the matrix: the automobile seat, a sub-matrix in which all elements that influence a driver's perception of comfort are considered (see schematization in Figure 2). The class labeled "person" comprises personal characteristics; the physical characteristics that influence comfort include anthropometric measures (Reed et al., 1994; Kolich, 2003; Fazlollahtabar, 2010) and consequent postures (Naddeo et al., 2015b) that are the starting point for dimensions, physique (BMI) and physical problems. Mental status, personality and psychological conditions influence the level of tiredness, the level of attention, biomechanical loading (Nimbarte et al., 2012) and fatigue, shown as increments of the drivers' physiological and emotional discomfort levels (Gerald, 2002).

Table 1. Seat-comfort evaluation sub-matrix.

SEAT-COMFORT			
PHYSIOLOGICAL	*EMOTIONAL-COGNITIVE*	*ORGANIZATIONAL-ENVIROMENTAL*	*POSTURAL*
PERSONAL CHARACTERISTICS			
PHYSICAL CHARACTERISTICS			
Anthropometric measures			
(M)	(M)		(P)posture overload, muscle complaint
Physique (BMI)			
(P)localized blood-pressure, body temperature, heartrate, metabolism	(M)tiredness		(P)muscle effort, posture overload, muscle complaint
Physical problems			
(P)tactile sensation, localized blood-pressure, body temperature, heartrate	(M)work overload, tiredness		(P)muscle effort, posture overload, muscle complaint
PERSONAL DATA			
Gender			
(P)localized blood-pressure, body tenmperature, heartrate, metabolism	(P)		(P)
Age			
(P)tactile sensation, localized blood-pressure, body temperature, heartrate, metabolism	(P)lack of attention		(P)muscle effort
EXPECTATIONS			
Expectations			
(M)	(P)perceived safety , aggressiveness and irritability, stress		(M)
WORK/TASK CHARACTERISTICS			
WORKSTATION			
Posture: angles and joints			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle complaint
CHARACTERISTICS OF TOOLS/OBJECTS WITH WHICH A PERSON INTERACTS			

Shape			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle complaint
Customization of the workstation (sitting)			
(M)tactile sensation	(M)tiredness	(M)perceived safety	(P)muscle effort, posture overload, muscle complaint
WORKING ENVIRONMENTS' CHARACTERISTICS			
VISUAL WELL-BEING			
Colors			
(M)	(M)aggressiveness and irritability, tiredness , lack of attention	(P)	(M)
Odors			
(M)	(M)aggressiveness and irritability, lack of attention	(P)	(M)
AUDITIVE WELL-BEING			
Vibrations			
(M)	(M)work overload, safety, aggressiveness and irritability, tiredness , lack of attention	(P)	(P)muscle effort, posture overload, muscle complaint
THERMAL WELL-BEING			
Interface temperature			
(P)tactile sensation, localized blood-pressure, body temperature	(M)lack of attention	(M)	(M)
Thermal resistance of clothing			
(P)tactile sensation, localized blood-pressure, body temperature	(M)	(M)	(M)
Persistence in a thermal condition			
(P)tactile sensation, localized blood-pressure, body temperature, aggressiveness, nervousness	(M)aggressiveness and irritability, lack of attention	(M)level of perceived safety	(M)
Contact pressure			
(P)tactile sensation, localized blood-pressure	(M)	(M)	(M)
TOOLS AND INSTRUMENTS FOR COMFORT MEASURING			
Invasivity			
(M)	(M)aggressiveness and irritability , lack of attention	(M)	(M)

Tactile interference			
(M)tactile sensation, localized blood-pressure	(M)lack of attention	(M)	(M)
Restriction of movements			
(M)	(M)work overload, aggressiveness and irritability, tiredness , lack of attention	(M)	(M)muscle effort, posture overload, muscle complaint
Override of action/ position			
(M)	(M)perceived safety, aggressiveness and irritability, tiredness , lack of attention	(M)	(M)muscle effort, posture overload, muscle complaint

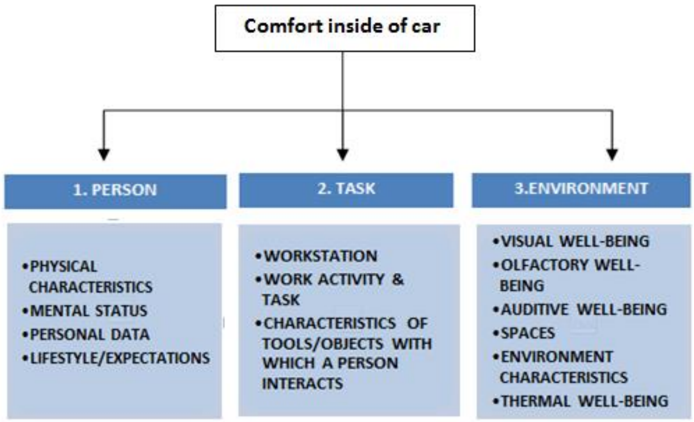


Figure 2. Schematization of elements that influence comfort inside of a car.

Identifying the correlation between age and gender in regard to perceived level of comfort is difficult because people of the same gender and age can have different perceptions. However, it can be stated that with increasing age, the level of perceived safety differs and the level of attention decreases, as well as gender influences on muscle strain and sensitivity to noise and vibration (Dauris et al., 2008) (see Table 2).

Table 2. Deployment of the comfort framework for Personal Characteristics related to comfort while driving.

PERSONAL CHARACTERISTICS			
PHYSIOLOGICAL	*EMOTIONAL-COGNITIVE*	*ORGANIZATIONAL-ENVIROMENTAL*	*POSTURAL*
PHYSICAL CHARACTERISTICS			
Anthropometric measures			
(M)	(M)		(P)posture overload, muscle complaint
Physique (BMI)			

(P)localized blood-pressure, body temperature, heartrate, metabolism	(M)tiredness		(P)muscle effort, posture overload, muscle complaint
Physical problems (chronic illness, trauma, and previous fractures)			
(P)tactile sensation, localized blood-pressure, body temperature, heartrate	(M)work overload, level of perceived tiredness		(P)muscle effort, posture overload, muscle complaint
MENTAL STATUS			
Personality			
(M)	(P)work overload, perceived safety, aggressiveness and irritability, level of perceived tiredness, stress, lack of attention		(M)
Psychological diseases (anxiety, stress)			
(M)body temperature, heartrate	(P)perceived safety, aggressiveness and irritability, tiredness, stress, lack of attention		(M)
PERSONAL DATA			
Gender			
(P)localized blood-pressure, body tenmperature, heartrate, metabolism	(P)		(P)
Age			
(P)tactile sensation, localized blood-pressure, body temperature, heartrate, metabolism	(P)lack of attention		(P)muscle effort
LIFESTYLE/EXPECTATIONS			
Lifestyle (diet, smoking, sports, sedentary lifestyle,. ..)			
(P)body tenmperature, heartrate, metabolism	(M)aggressiveness and irritability, lack of attention		(M)muscle effort
Expectations			
(M)	(P)level of perceived safety, aggressiveness and irritability, stress		(M)

During the driving task, the user needs to perform (with a certain frequency) a series of actions such as pressing buttons, turning the steering wheel and using pedals. The work activity aspects influencing comfort are evaluated by different indicators: type of loads and actuation, operating speed (Apostolico et al., 2013; Naddeo et al. 2014b), an action's frequency, rest–pause duration and frequency. The perception of comfort in every situation can be affected by duration of the task (Moore & Garg, 1995; Kee & Karwowski, 2001), and particularly in car seat's design (Vergara & Page, 2002).

In our paper, defining the driver's comfort level refers only to the seat. In literature, in fact, the seat's geometry (Reed et al., 1994; Kolich, 2003; Apostolico et al., 2013), breathability and rigidity are considered the most important indexes of driver comfort. During the driving experience, the driver interfaces not only with the seat but also with a high number of other elements (steering wheel, pedals, knobs, etc.). Each element's shape (Kuijt-Evers et al., 2004), position (Ellegast et al., 2012; Naddeo et al., 2014b; Naddeo et al., 2015a; Patrick et al., 2004; Lars et al., 2003) and orientation can make the vehicle cockpit more or less comfortable.

Today, people spend much more time in the car, especially those whose work involves driving for many hours (taxi driver, couriers, truck drivers, etc.); as such, the comfort of the car cockpit can be studied in ways similar to any work environment. The colours of the interior, artificial lighting conditions, air quality and space are factors that influence the driver's perceived comfort.

Vibrations, for example, are one of the most important environmental aspects (Smith et al., 2006; Falou et al., 2003). The consequences of exposure to vibration can be a decrease in cognitive and postural comfort. Vibrations, in fact, result in an increase in the level of irritability, lack of attention and postural overload.

One issue involving improved functioning of the automobile are unpleasant smells due to materials used in the interior. Realizing this, Yamada et al. (2000) created advanced seat fabrics with high performance deodorant function that effectively controls the smell in the passenger compartment; in the matrix in Table 3, olfactory well-being is related to air quality and odors.

The thermal aspect is the consequence of influence factors such as humidity, temperature and thermal-resistance of clothes (D'ambrosio Alfano, & Liotti, 2004). Also ambient lighting is important inside the car, for improving the sense of spaciousness, as well as the impression of safety, value and comfort (Caberletti et al., 2009). In our study, aspects such as colours, luminance level, the combination of lighting materials and their reflection have been included in the matrix.

One of the most common evaluations on automobile seats is based on the interface pressure of the driver seat. Previous studies have shown that preferred pressure levels are different for different body parts as well as different anthropometric groups (Kolich, 2004; Dunk & Callaghan, 2005; Oudenhuijzen et al., 2003; Kyung et al., 2008). There are definite associations between interface pressure and seated discomfort. Based on the results, specific approaches are recommended to improve the driver's sitting experience: (1) lower pressure ratios at the buttocks and the higher-pressure ratios at the upper and lower back, and (2) balanced pressure between the bilateral buttocks and between the lower and upper body (Table 4).

Table 3. Deployment of the comfort framework for Work/Task Characteristics related to comfort while driving.

WORK/TASK CHARACTERISTICS			
PHYSIOLOGICAL	*EMOTIONAL-COGNITIVE*	*ORGANIZATIONAL-ENVIROMENTAL*	*POSTURAL*
WORKSTATION			
Posture: angles and joints			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle complaint
Individual safety equipment: overall dimensions and heaviness			
(M)tactile sensation, localized blood-pressure, body temperature	(M)perceived safety, lack of attention	(M)	(M)muscle effort, posture overload, muscle complaint
WORK ACTIVITY & TASK			
Type of loads and actuation (lifting, pulling, pushing)			
(M)localized blood-pressure, body temperature, heartrate	(M)level of perceived tiredness	(M)	(P)muscle effort, posture overload, muscle complaint
Operating speed			
(M)body temperature	(M)work overload, tiredness, stress	(P)	(M)muscle effort, posture overload, muscle complaint
Actions' frequency			
(M)body temperature	(M)work overload, tiredness, stress	(P)	(M)muscle effort, posture overload, muscle complaint
Rest-pause duration and frequency			
(M)	(M)work overload, perceived safety, aggressiveness and irritability, tiredness, stress, lack of attention	(P)	(M)muscle effort, posture overload, muscle complaint
Revel of precision			
(M)	(M) aggressiveness and irritability, tiredness, stress, lack of attention	(P)	(M)muscle effort, posture overload, muscle complaint
Time maintaining of the posture with and/or without loads			
(M)localized blood-pressure, body temperature, heartrate	(M)aggressiveness and irritability, tiredness	(M)	(P)muscle effort, posture overload, muscle complaint
Time and duration of work activity/tasks			
(M)	(M) aggressiveness, work overload, irritability, tiredness, stress, lack of attention	(P)	(M)muscle effort, posture overload, muscle complaint
Work-shifts			
(M)muskular exertion, aggressiveness, nervousness, tiredness	(M)work overload, perceived safety, aggressiveness and	(P)perceived safety	(M)

	irritability, tiredness, lack of attention		
CHARACTERISTICS OF TOOLS/OBJECTS WITH WHICH A PERSON INTERACTS			
Shape			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle complaint
Weight			
(M)	(M)	(M)perceived safety	(P)muscle effort, posture overload, muscle complaint
Relative position between person and object/tool			
(M)	(M)	(M)perceived safety	(P)muscle effort, posture overload, muscle complaint
Frequency of lifting / pulling / pushing			
(M)heartrate, localized blood-pressure, body temperature	(P)tiredness	(M)	(M)muscle effort, posture overload, muscle complaint
Handling characteristics (grip, grasp, pinch, ….)			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle complaint
Customization of the workstation (sitting)			
(M)tactile sensation	(M)tiredness	(M)perceived safety	(P)muscle effort, posture overload, muscle complaint
Commands' layout			
(M)	(M)	(M)	(P)muscle effort, posture overload, muscle compl.

Table 4. Deployment of the comfort framework for Environment Characteristics related to comfort while driving.

WORKING ENVIRONMENTS' CHARACTERISTICS			
PHYSIOLOGICAL	*EMOTIONAL-COGNITIVE*	*ORGANIZATIONAL-ENVIROMENTAL*	*POSTURAL*
VISUAL WELL-BEING			
Colors			
(M)	(M)aggressiveness and irritability, tiredness, lack of attention	(P)	(M)
Artificial lighting conditions			
(M)	(M)perceived safety, aggressiveness and irritability, tiredness	(P)perceived safety	(M)muscle complaint
Natural lighting conditions			
(M)	(M)aggressiveness and irritability, tiredness	(P)	(M)muscle complaint
Lights' reflection and refraction on walls and objects			
(M)	(M)aggressiveness and irritability, tiredness, lack of attention	(P)	(M)muscle complaint

OLFACTORY WELL-BEING			
Air quality			
(M)aggressiveness, nervousness	(M)aggressiveness and irritability	(P)	(M)
Odors			
(M)	(M)aggressiveness and irritability, lack of attention	(P)	(M)
AUDITIVE WELL-BEING			
Noises			
(M)	(M)perceived safety, aggressiveness and irritability, tiredness, lack of attention	(P)perceived safety	(M)
Vibrations			
(M)	(M)work overload, perceived safety, aggressiveness and irritability, tiredness, lack of attention	(P)	(P)muscle effort, posture overload, muscle complaint
SPACES			
Workspace			
(M)muskular exertion, aggressiveness, nervousness	(M)perceived safety, aggressiveness and irritability	(P)perceived safety	(M)muscle effort, posture overload, muscle complaint
Layout			
(M)	(M)	(P)perceived safety	(M)
ENVIROMENT CHARACTERISTICS			
Cleanliness			
(M)	(M)aggressiveness and irritability	(P)	(M)
Tidiness			
(M)	(M)work overload, aggressiveness and irritability	(P)	(M)
THERMAL WELL-BEING			
Air-temperature			
(M)body temperature, aggressiveness and nervousness	(M)aggressiveness and irritability, lack of attention	(P)perceived safety	(M)
Interface temperature			
(P)tactile sensation, localized blood-pressure, body temperature	(M)lack of attention	(M)	(M)
Humidity			
(M)localized blood-pressure	(M)	(P)	(M)
Thermal resistance of clothing			
(P)tactile sensation, localized blood-pressure, body temperature	(M)	(M)	(M)

Persistence in a thermal condition			
(P)tactile sensation, localized blood-pressure, body temperature, aggressiveness, nervousness	(M)aggressiveness and irritability, lack of attention	(M)perceived safety	(M)
Contact pressure			
(P)tactile sensation, localized blood-pressure	(M)	(M)	(M)
Air speed			
(M)body temperature	(M)	(P)	(M)

Materials and Methods

Participants

Twenty volunteers took part in this study and were recruited from among students taking Engineering courses at the University of Salerno. In this way, the sample was quite homogenous. This aspect was necessary to ensure the investigation was valid and consistent. Students' ages were between 20 and 29 years and had a valid driving license for at least three years.

Apparatus

A 2005 Ford Fiesta MY seat (see Figure 3) was used for the test. The seat is characterized by the ability to adjust seat height and adjust the headrest. The seat has a fabric cover. Five ways to represent the car seat vis a vis human interaction have been chosen: physic interaction with the real prototype, presentation and observation only of real prototype, photographic presentation of real prototype, presentation of 3D virtual digitized prototype (the 3D virtual digitized prototype was realized by the reverse engineering method), and interaction with the real prototype placed inside the car.

Figure 3. 2005 Ford Fiesta MY seat.

Methods to acquire data

The Kansei Engineering methodology specializes in the translation of affective values into concrete product design parameters (Nakada, 1997; Nagamachi, 2002). To achieve this, Kansei Engineering uses semantic differential scales (SD-scales) as a central pillar. The "5-point scale" questionnaire utilized during the test is shown in Figure 4. Within the questionnaire, there are explicit questions on the perceived level of comfort as well as other questions relating to various indicators and facets of comfort on automobile seats considered in the literature (Zhang et al., 1996; Da Silva et al., 2012).

Comfortable	☐ ☐ ☐ ☐ ☐	Uncomfortable
Compact	☐ ☐ ☐ ☐ ☐	Not Compact
Elegant	☐ ☐ ☐ ☐ ☐	Not Elegant
Sober	☐ ☐ ☐ ☐ ☐	Excessive
Sport	☐ ☐ ☐ ☐ ☐	Tourist
It'll continue to like	☐ ☐ ☐ ☐ ☐	It will go to tire me
Pleasant	☐ ☐ ☐ ☐ ☐	Unpleasant
Gaudy	☐ ☐ ☐ ☐ ☐	Dull
Luxurious	☐ ☐ ☐ ☐ ☐	Cheap
Exciting	☐ ☐ ☐ ☐ ☐	Boring
Curved	☐ ☐ ☐ ☐ ☐	Linear
Design	☐ ☐ ☐ ☐ ☐	Rough
Usual	☐ ☐ ☐ ☐ ☐	Unusual
Refined	☐ ☐ ☐ ☐ ☐	Unrefined
Soft	☐ ☐ ☐ ☐ ☐	Hard
Relaxing	☐ ☐ ☐ ☐ ☐	Not Relaxing
Innovative	☐ ☐ ☐ ☐ ☐	Ordinary
Functional	☐ ☐ ☐ ☐ ☐	Decorative
Robust	☐ ☐ ☐ ☐ ☐	Easily broken
Pleasant	☐ ☐ ☐ ☐ ☐	Unpleasant
Good	☐ ☐ ☐ ☐ ☐	Not Good

Figure 4. Questionnaire with Kansei terms.

Procedure

During the test, participants were invited to interact with four kinds of prototypes. During the physical interaction with the real prototype, the participants placed themselves on the seat, allowing them to acquire information about the seat. In this case, the participants could evaluate the finish of the surfaces, the softness and the compactness of the coatings, as well as the level of comfort. The second test was evaluation of real prototype without any interaction with it and without giving any extra information. The participants could evaluate only the colors and the geometry of the seat. The third test was prototyping through photography. In this case, a short introduction was shown on the state-of-the-art automobile seat. The fourth type of interaction was a 3D presentation as a virtual digitized prototype (Figure 5).

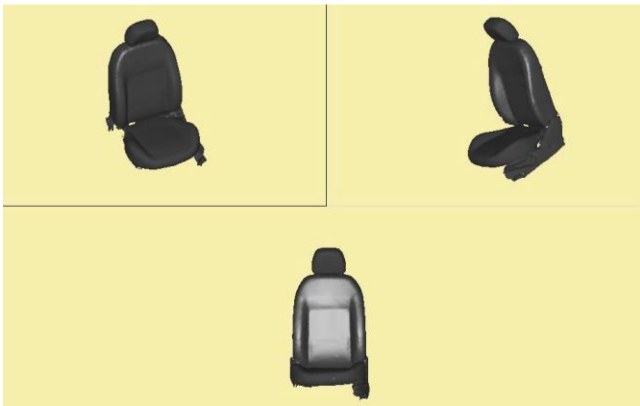


Figure 5. Virtual prototype of the seat

Through this method, the participants experienced 3D vision of the object (similar to recent web-based car presentations). The final test was physical interaction with the automobile seat inside the car. In this case, the participants interacted with the seat; they could position the seat according to their preferences, interact with the other elements inside the car and listen the music (but without turning their attention from the seat). After each test, the subjects were required to complete the same questionnaire as the other prototypes to evaluate their comfort sensations. The aim of this study was to demonstrate that the same object, presented in different ways, results in different comfort sensations.

Results

After data acquisition, all the terms of the questionnaire were analyzed. The results demonstrated that the highest score of comfort was obtained when the subjects evaluated the seat inside the car, simulating the driving experience for a few minutes (Figure 6).

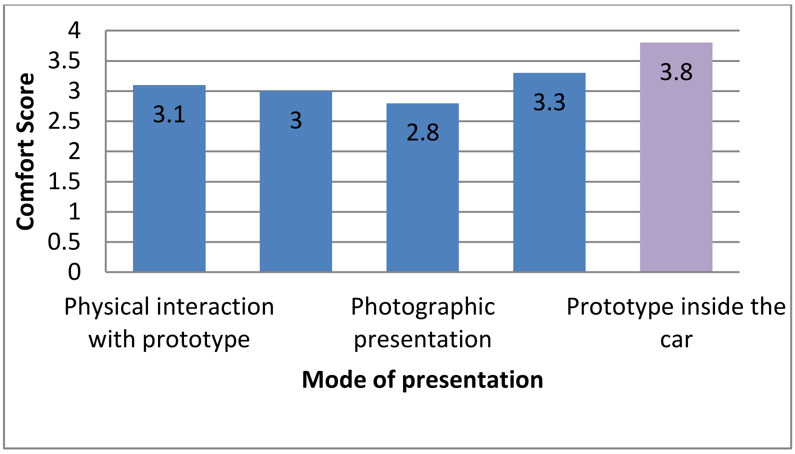


Figure 6. Average comfort score related to the different kinds of presentation of the seat.

The results of the questionnaire for each type of presentation were compared with results obtained by the physical interaction with the automobile seat inside the car. The scores, in most cases, were highest when the subjects were inside the car (see Figures 7–10).

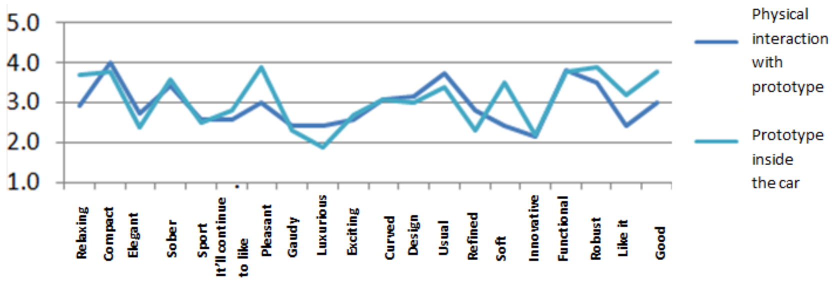


Figure 7. Comparison of the questionnaire average scores between the physical interaction with the prototype and interaction with the seat inside the car.

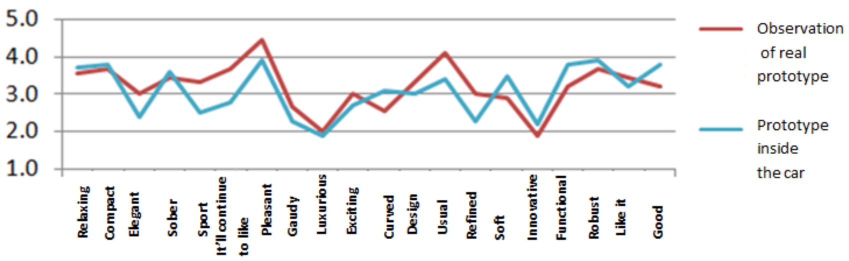


Figure 8. Comparison of the questionnaire scores between observation of real prototype and interaction with the seat inside the car.

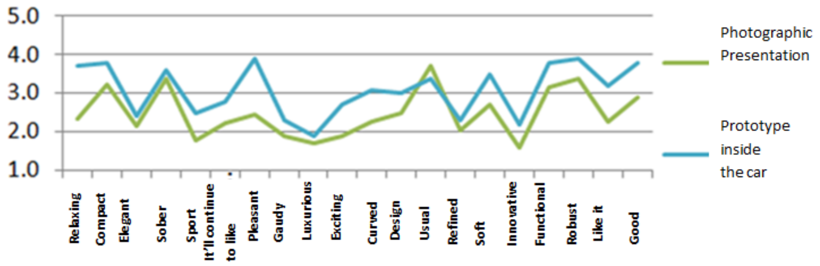


Figure 9. Comparison of the questionnaire scores between the photographic presentation of the seat and interaction with the seat inside the automobile.

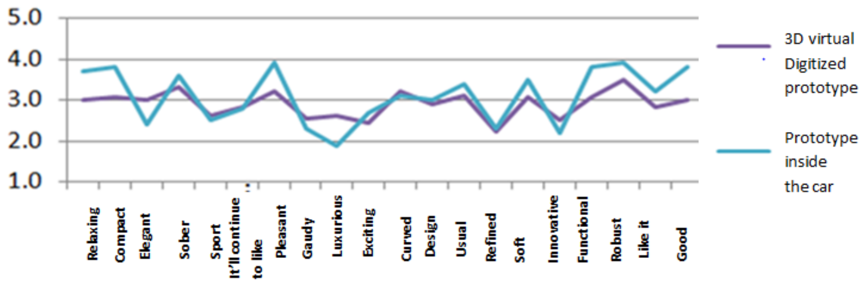


Figure10. Comparison of the questionnaire scores between the 3D virtual digitized prototype and interaction with the seat inside the automobile.

Discussion

The aim of this study was to demonstrate that the feeling of comfort/discomfort related to an object could not be determined if only the object itself is considered. The study has tested the hypothesis of the Naddeo and Cappetti model: according to the Naddeo and Cappetti model, to evaluate levels of satisfaction of a condition of service, we must consider all the elements of environment and how they interact with the object itself. To demonstrate this hypothesis, the authors applied the matrix to the case of the automobile seat. The automobile seat was evaluated by a sample of subjects in different ways, adding some information each time: from a simple photo to a journey inside the automobile. From the analysis of the questionnaire results, the seat when evaluated inside the car scored as most comfortable. In this case, in fact, the comfort index is 3.8, higher than all other ways of representation.

This demonstrates that it is insufficient to evaluate only the finishes of the surfaces, the softness and the compactness of the coatings (construction interaction), only the colours and the geometry (observation of real prototype without interaction and photographic interaction) or only the 3D vision of the object. Instead, it is necessary to evaluate the object inside its natural environment and during a simulation of the activity. In this way, all the elements of the matrix are considered. The results deeply sustain the Kansei technique and prove the formulated hypothesis: to analyze an emotional design, it is necessary to apply a methodology in which all the consumers' senses are involved. Therefore, it was not unexpected that the lowest index was correlated to the photographic presentation in which interaction between the user and the object is next to nothing.

References

Apostolico, A., Cappetti, N., D'Oria, C., Naddeo, A. & Sestri, M. (2008). Postural comfort evaluation: experimental identification of Range of Rest Posture for human articular joints. *International Journal on Interactive Design and Manufacturing , 8*, 109-120.

Caberletti, L., Kümmel, D. M., Elfmann, K. & Schierz, C. (2009). Influence of Ambient Lighting in Vehicle Interior on the Drivers Perception. In *International Conference on the Effects of Light on Wellbeing*.

Da Silva, L., Bortolotti, S. L. V., Campos, I. C. M. & Merino, E. A. D. (2012). Comfort model for automobile seat. *Work: A Journal of Prevention, Assessment and Rehabilitation, 41*, 295-302.

D'ambrosio Alfano, F.R. & Liotti, F. (2004). La qualità degli ambienti confinati non industriali: il discomfort termo igrometrico [Quality of confined non industrial environments: thermal-hygrometric discomfort]. *Giornale Italiano di Medicina del Lavoro ed Ergonomia, 26*, 401-415.

Dauris, D.D.I., Deros, B.M.D. & Noor, M.J.M. (2008). Understanding the effect of noise, vibration and seat discomfort towards a passenger´s vehicle driver through self-reported survey. *Journal Teknologic, 49*, 39-53.

Dunk, N.M. & Callaghan, J.P. (2005). Gender-based differences in postural responses to seated exposures. *Clinical Biomechanics, 20*, 1101–1110.

Ellegast, R.P., Kraft K., Groenesteijn, L., Krause, F., Berger, H. & Vink, P. (2012). Comparison of four specific dynamic office chairs with a conventional office chair: impact upon muscle activation, physical activity and posture. *Applied Ergonomics, 43*, 296-307.

Falou, W.E., Duchene, J., Grabisch, M., Hewson, D., Langeron, Y. & Lino, F. (2003). Evaluation of driver discomfort during long-duration car driving. *Applied Ergonomics, 34*, 249–255.

Fazlollahtabar, H. (2010). A subjective framework for seat comfort based on a heuristic multi criteria decision making technique and anthropometry. *Applied Ergonomics, 42*, 16-18.

Gerald, M. (2002). Towards a transactional ergonomics for driver stress and fatigue. *Theoretical Issues in Ergonomics Science, 3*,195-211.

Kee, D. & Karwowski, W. (2001). LUBA: an assessment technique for postural loading on the upper body based on joint motion discomfort and maximum holding time. *Applied Ergonomics, 32*, 357-66.

Kolich, M. (2008). A conceptual framework proposed to formalize the scientific investigation of automobile seat comfort. *Applied Ergonomics, 39*, 15–27.

Kuijt-Evers, L.F.M., Groenesteijn, L., De Looze, M.P & Vink, P. (2004). Identifying factors of comfort in using hand tools. *Applied Ergonomics, 35*, 453-458.

Kyung, G., Nussbaum, M.A. & Babski-Reeves, K. (2008). Driver sitting comfort and discomfort (Part I): use of subjective ratings in discriminating car seats and correspondence among ratings. *International Journal of Industrial Ergonomics, 38*, 516-525.

Lars, H., Willfried, W. & Lena, S. (2003). A control handling comfort model based on fuzzylogics, *International Journal of Industrial Ergonomics, 31*, 87–100.

Moore, J.S., & Garg, A. (1995). The Strain Index: a proposed method to analyze jobs for risk of distal upper extremity disorders. *American Industrial Hygienists Association Journal, 56*, 443-458.

Naddeo, A., Cappetti, N., Vallone, M. & Califano, R. (2014a). New trend line of research about comfort evaluation: proposal of a framework for weighing and evaluating contributes coming from cognitive, postural and physiologic comfort perceptions. In T. Ahram, W. Karwowski and T. Marek (Eds.).Proceedings of the *5th International Conference on Applied Human Factors and Ergonomics*

Naddeo, A., Cappetti, N. & Ippolito, O. (2014b). *Dashboard reachability and usability tests: a cheap and effective method for drivers' comfort rating*. SAE Technical Paper 2014-01-0455.

Naddeo, A., Apicella, M. & Galluzzi, D. (2015a). *Comfort-Driven Design of Car Interiors: A Method to Trace Iso-Comfort Surfaces for Positioning the Dashboard Commands*. SAE Technical Paper 2015-01-1394.

Naddeo, A., Cappetti, N. & D'Oria, C. (2015b). Proposal of a new quantitative method for postural comfort evaluation. *International Journal of Industrial Ergonomics, 48*, 25-35.

Naddeo, A., Cappetti, N., Vallone, M. & Califano, R. (2015c). The role of expectation in comfort perception: The mattresses' evaluation experience. In Proceedings of the *6th International Conference on Applied Human Factors and Ergonomics,* 26-30 of July 2015, Las Vegas (USA).

Nagamachi, M. (2002). Kansei engineering as an ergonomic consumer-oriented technology for product development. *Applied ergonomics, 33,* 289-294.

Nakada, K. (1997). Kansei engineering research on the design of construction machinery. *International Journal of Industrial Ergonomics, 19*,129-146.

Nimbarte, A.D., Al Hassan, M.J., Guffey, S.E. & Myers, W.R. (2012), Influence of psychosocial stress and personality type on the biomechanical loading of neck and shoulder muscles, *International Journal of Industrial Ergonomics, 42*, 397-405.

Oudenhuijzen, A., Tan, K. & Morsch, F. (2003). *The relationship between seat pressure and comfort*. SAE Technical Paper 2003-01-2213.

Patrick, G.D., Raymond, W. & Niall, V.O. (2004). The effects of work height, workpiec e orientation, gender and screwdriver type on productivity and wrist deviation. *International Journal of Industrial Ergonomics, 33,* 339–346.

Reed, M.P., Schneider, L.W. & Ricci, L.L. (1994). *Survey of auto seat design recommendations for improved comfort* (Report UMTRI-94-6).University of Michigan, Transportation Research Institute.

Smith, D.R., Andrews, D.M. & Wawrow, P.T. (2006). Development and evaluation of the Automotive Seating Discomfort Questionnaire. *International Journal of Industrial Ergonomics, 36*, 141-149.

Vergara, M. & Page, A. (2002). Relationship between comfort and back posture and mobility in sitting comfort. *Applied Ergonomics*, *33*, 1-8.

Vink, P. & Hallbeck, S. (2012). Editorial: comfort and discomfort studies demonstrate the need for a new model. *Applied Ergonomics, 43,* 271–276.

Yamada, Y., Owaki, M., Mori, T., Fukumoto, K., Miura, S., Furuta, M. & Matsuyama, M. (2000). Advanced seat fabrics with high performance deodorant function. *Society of Automotive Engineers of Japan*, *21*, 543-547.

Zhang, L., Helander, M. G. & Drury, C.G. (1996). Identifying factors of comfort and discomfort in sitting. *Human Factors, 38*, 377-389.