Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference

de Waard, Dick; Brookhuis, Karel; Wiczorek, Rebecca; Di Nocera, Francesco; Barham, Philip; Kluge, A.; Gerbino, Walter; Toffetti, Antonella

Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference

Human Factors: sustainable life and mobility

Edited by
Dick de Waard, Karel Brookhuis, Rebecca Wiczorek, Francesco di Nocera, Rino Brouwer, Philip Barham, Clemens Weikert, Annette Kluge, Walter Gerbino, and Antonella Toffetti

ISSN 2333-4959 (online)

Please refer to contributions as follows:
Authors (2014), Title. In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, and A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. Downloaded from http://hfes-europe.org
(ISSN 2333-4959)

Available as open source download

Contents

PIE (PSYCHOPHYSIOLOGY IN ERGONOMICS)

Use of a multidimensional approach in the automotive product development: Quality of turn indicator sounds
Wagner & Kallus
Keep calm and don´t move a muscle: Motor restlessness as an indicator of mental workload
Di Nocera, Proietti Colonna, Capobianco, Dessì, Mastrangelo, & Steinhage

SURFACE TRANSPORTATION

Timing of in-vehicle advisory warnings based on cooperative perception
Naujoks & Neukum
Drivers´ hand positions on the steering wheel while using Adaptive Cruise Control (ACC) and driving without the system
Bianchi Piccinini, Prati, Pietrantoni, Manzini, Rodrigues, & Leitão
Ship voyage plan coordination in the MONALISA project: user tests of a prototype ship traffic management system
Porathe, de Vries, & Prison

WORKPLACE

Prevention of physiological and psychological stress at computer-equipped workplaces
Tint, Meigas, Tuulik, Pille, Oha, Reinhold, Karai, Tuulik, & Lauri
Stress and strain in mobile IT-supported work: an empirical study in the area of mobile services involving smartphones, notebooks and similar devices
Bretschneider-Hagemes
Can Immersive Virtual Environments make the difference in training industrial operators?
Nazir, Kluge, & Manca
Strain caused by Head Mounted Displays
Wille, Grauel, & Adolph

AVIATION

Wearable crew support technology on the International Space Station: the mobile Procedure Viewer (mobiPV)
Chintamani, Van Lierde, Maloney, Kiernan, Oliveira, & Wolff
Exceptional performing screener project: methods and findings from the field study
Gibb, Gibb, Barrientos, Flint, Hale, & Wilson
VOD-CA - Assisting Human Flight Performance and Situation Awareness in Lateral Deconflicting
Manske, Meysel, & Boos
Teamwork: collaborative activities of French fighter pilots in joint operations
Marchand

The touch appeal: Why touching things is so popular in Human-Computer Interaction

Stefan Brandenburg & Nils Backhaus
Technische Universität Berlin
Germany

Abstract

This paper utilizes a theoretical framework (the model of user experience) to describe the appeal of touch interfaces. Therefore it starts with a brief discussion of the framework. In the second part it presents empirical evidence undermining the theoretical claims. Study results show that theoretical claims are partially supported by the data. The findings are in line with previous research. The paper concludes that the touch appeal derives from the structure of the (task) environment.

Introduction

Touch-sensitive graphic user interfaces are implemented in a wide range of technical artefacts. However, Human-Computer Interaction researchers still try to explain why this type of interaction is so popular.

For example, Drewitz and Brandenburg (2010) proposed a framework (model of user experience) that relates characteristics of interfaces to feelings of ease of use and joy of use (Davis, 1989). Based on this framework, touch interfaces structure the environment in which an operator solves his tasks. They address mechanisms like affordances, constraints, etc. In turn, these mechanisms foster the elicitation of immediate behaviour and, finally, both feeling of ease of use and joy of use (Drewitz & Brandenburg, 2010; Brandenburg et al., 2013). Therefore, based on Drewitz and Brandenburg (2010) touch interfaces are appealing because of their structured environment.

The structured environment of touch interfaces

Based on the model of user experience, all environments are structured with respect to four characteristics: affordances, constraints, attraction of attention, and mapping (Drewitz & Brandenburg, 2010).

Affordances provide action opportunities. In 1979 Gibson suggested that people do not just perceive an object. Instead, they additionally assess what an object affords to them. Based on peoples' background knowledge and motivations, the same object has different affordances. Therefore, affordances are neither a property of an object, nor a property of the artefact. They emerge depending on both, the properties of the artefact (i. e. weight, structure) and the observer (i. e. knowledge and motivation).

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Hence, learning changes affordances as soon as it changes the background knowledge of learners. In turn, to specify affordances one needs to consider specifications of the environment and the observer. An affordance relates properties of the environment to perceived action possibilities for the observer (Gibson, 2000). In line with Gibson (1979), Norman (1988) proposed that affordances provide knowledge to the operation of things. He stated that affordances are on hand if users of technical artefacts are enabled to apply their previously gained knowledge to the current interaction. In line with Norman (1988), we think of affordances as pointed out by Greeno (1994, p. 338): “In any interaction involving an agent with some other system, conditions that enable that interaction include some properties of the agent along with some properties of the other system. [...] The term affordance refers to whatever is about the environment that contributes to the kind of interaction that occurs.” In line with the Gibsonian idea of affordances, we assume that affordance-congruent behaviour is fast and unconscious. Till now, affordances have been intensively studied and partially successfully implemented in interface design (e. g. Vogel et al., 2011).

Having a lot of competing affordances, the user can only follow one of them at a time. Therefore the design of an artefact has to be constrained. In human computer interaction constraints are limitations in the interface or the operational concept of the artefact. For example, limiting the number of actions that can be executed at a time increases the chance that a user accomplishes his interaction goal. Norman (1988) pointed out that the users’ mental model should map to the designers model of an artefact and both communicate over the system image. Therefore it is important that the environment that is created by the artefact maps to the users knowledge in terms of appearance and functionality (cf. Norman, 1988).

Finally, attraction of attention means that system designers apply (visual, auditory, etc.) cues to attract the user’s attention. Moreover, designers use these cues to guide the user’s attention towards subsequent interaction steps. For example, Vogel et al. (2011) found that using colour to guide the user’s attention through the interaction fostered interaction and increased positive user experience. They concluded that highlighting subsequent interaction steps relieves users working memory and thus decreases cognitive strain.

Touch interfaces met most of these characteristics. Since the operating principle of these interfaces mainly relies on (simulated) 3D Buttons, the affordance of pressing them is very prominent (cf. Vogel et al., 2011). Moreover, as Vogel et al. (2011) also pointed out, semantically meaningful icons additional foster affordance congruent behaviour. However, touch interfaces are also heavily constrained in terms of hardware. They mostly provide a large screen almost without any hardware buttons. Thus the number of competing affordances (pressing hard- or software buttons) is reduced. Regarding their software, touch interfaces are also very constrained. Due to the minimum size of software buttons and limited screen size only a limited amount of action properties can be presented (cf. Backhaus & Brandenburg, 2013). This reduces the number of competing affordances as well. If designers attract the attention of users and map the interface structure to the user’s mental model, the effects of the structured environment on behaviour and his or her experience might be positive.

Effects of the structured environment on behaviour and experience

According to the framework of Drewitz and Brandenburg (2010) the structure of the environment determines the degree of users Immediate Interactive Behaviour (Neth et al., 2007); knowledge acquisition (Ease of Induction Brandenburg et al., 2009) and feelings of ease of use and joy of use (Davis, 1989). The concept of immediate interactive behaviour "[...] entails all adaptive activities of agents that routinely and dynamically use their embodied and environmentally embedded nature to support and augment cognitive processes." (Neth et al. 2007, p. 33). That means the users' interaction with the environment utilizes simple interaction routines (i.e. tapping or pressing) which are fast (1/3 to 3 seconds), interaction-intensive and without cognitive effort (Neth et al., 2007). There is no deliberate behaviour that can occur on a time scale smaller than that of immediate interactive behaviour (cf. Newell, 1990). The application of simple interaction routines on the one hand alters the environment the agents are acting in and they alter the state of the operators' cognitive system on the other hand (Neth et al., 2007). Its application impacts subjects' motivation due to the experience of their ability to make progress in their interaction. Furthermore, people have feelings of competence based on the successful application of their knowledge (Drewitz & Brandenburg, 2010). If the skill acquisition is facilitated due to the appearance of IIB, operating knowledge (i.e. rule based knowledge) is acquired easily by induction (Brandenburg et al., 2009). Brandenburg et al. (2009) point out that the subsequent execution of simple interaction routines increases the probability that these actions will be linked up with each other and merged into new (procedural) knowledge. In consequence people will experience feelings of ease of use. Examples for such mechanisms of skill acquisition are provided by recent theories of cognition, for instance the production compilation mechanism in ACT-R (Anderson et al. 2000). Closely linked with the ease of use is the joy of use (Davis, 1989). When people extensively show immediate interactive behaviour, ease of induction as well as positive emotional and motivational reactions occur (Drewitz & Brandenburg, 2010; Brandenburg et al., 2013). Again, while proceeding in a task, people become aware of the progress they make. Not getting stuck in an impasse but approaching the interaction goal is accompanied by feelings of success and self-efficacy. As touch interfaces meet most of the characteristics of the structured environment, users should show high levels of immediate interactive behaviour, ease of induction and positive feelings of ease of use and joy of use if being confronted with a touch interface. As pointed out earlier, even the software running on a touch interface can further increase the touch appeal of these artefacts. As in previous works (e. g. Brandenburg et al., 2013) the focus on this paper lies on the software design of touch interfaces.

Research objectives

So far, empirical investigations did not yet deliver clear evidence for the aspect of software affordances (e. g. Brandenburg et al., 2013). Therefore the present study investigates the role of affordances for the emergence of user experience on a multi-touch table. Based on the theoretical input it can be assumed that the presence of affordances does facilitate subjects' performance and leads to more positive user experience compared to their absence. To demonstrate that affordances do affect immediate interactive behaviour and user experience differently than standard signals, we included the presence or absence of arrows as standard signals in the experiment as well (see also Norman, 1999). In contrast to affordances, standard signals are conveyed culturally and their symbolic meanings have to be learned explicitly. Artificial visual indicators are often described as signs or symbols (Petocz et al., 2008). Following Petocz et al. (2008), signs are arbitrary cues that convey a pre-defined message. Thus, signs and symbols are not affordances (Norman, 1999). "They are examples of the use of a shared and visible conceptual model, appropriate feedback, and shared, cultural conventions." (Norman, 1999, p. 41). Arrows are some kind of symbols, defined as a line with one end marked, inducing an asymmetry (cf. Kurata et al., 2005). Arrows have a diversity of semantic roles, e.g. moving direction, physical change, labelling, focusing attention, which have to be learned and distinguished in a given situation (Kurata et al., 2005). Since arrows are a somewhat well learned signal, their presence should increase immediate interactive behaviour and thus elicit a more positive user experience compared to their absence. However, the positive effect of affordances should be larger than the effect of signals.

Including both factors in a 2x2 between subjects design, it was possible to test the single and joint effects of both types of information on subjects' performance and experience.

Method

Subjects and material

A total of $N = 48$ multi-touch table novices (age: $M = 25.9$, $SD = 4.5$, 17 female/31 male) voluntarily participated in the experiment. The multi-touch interface consisted of a text box presenting the actual task, a working environment and three blue squares on the right hand side (see Fig.1). These three objects had to be manipulated using different gestures. The subjects' task was to execute the right gestures for rotating, scaling and cutting a blue square (see Fig.2). In the experiment, all subjects saw (light) touch areas indicating where they had to place their fingers. The arrows were presented to the subjects in the corresponding group, only. Affordances were defined as an initial movement of the object into the direction of gesture execution. For example, for scaling a blue object, subjects put their fingers in the lighted corners of the blue square. As soon as both fingers were placed in opposite corners, the object started to gradually enlarge itself to 130 percent of its original size. Then it shrunk back to the original size. This movement was repeated as long as participants started to execute the gesture. Additionally, participants were asked to fill in the NASA-TLX (Hart & Staveland, 1989) to assess their subjective experience.

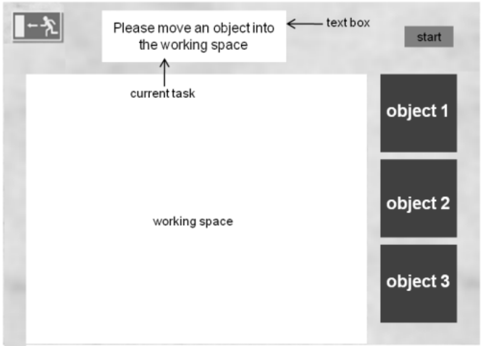


Figure 1. Experimental environment presented on the multi-touch table

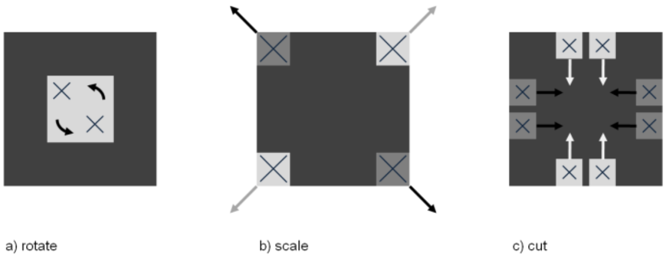


Figure 2. Visualization of the three gestures. In the experiment, all subjects did not see the (light) touch areas. The arrows were presented to the subjects in the corresponding group, only. Crosses are shown for visualization purposes.

Procedure and experimental design

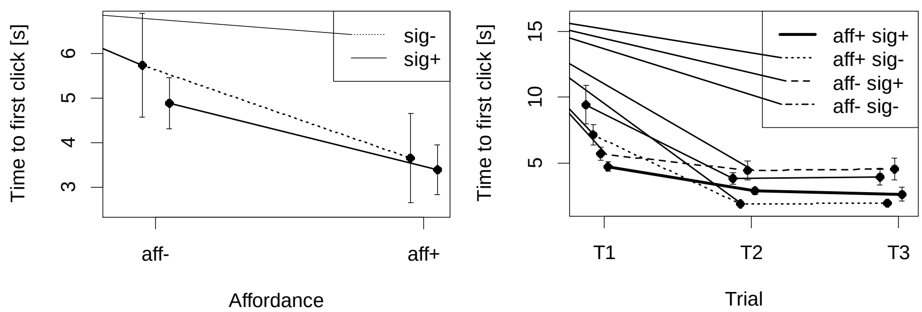
All subjects were randomly assigned to one of the four experimental groups. First, subjects familiarized themselves with the multi-touch table in an exercise trial. In this first trial, participants tested the movement of objects and the multi-touch surface sensitivity. Therefore, subjects were asked to move one of the three squares from the right hand side (see Fig.1) into the working space. Now participants received the instruction that their task was to execute the three different gestures: rotate, scale and cut (see Fig. 2a-c) three times. Hence each subject accomplished three trials, each of them containing all three gestures. The mapping of gestures to objects was randomized over subjects and trials. Hence, it was impossible to associate an object with a special gesture. For each manipulation of a square, participants dragged one of them into the middle of the working space. Then, they read the current task in the interfaces text box. If participants felt that they understood the task, they pressed the start button and initiated the gesture execution. At the end of each trial all participants filled in the NASA-TLX. The entire experiment took about 30 minutes.

Two independent variables were manipulated, both as between-subjects variables: affordances that we defined as object movements indicating the gesture specific direction of finger movements and signals that were set out as arrows showing the gesture specific direction of finger movements. Furthermore the three manipulations (rotate, scale, cut) as well as the three trials were included as within-subjects factors. Dependent variables were defined as: *time to first click* that was assessed from the pressing of the start button to the initial manipulation of the object in the working space (it served as one measure of immediate interactive behaviour, see also Drewitz & Brandenburg, 2010), *total task time* that was measured from pressing the start button to the end of task, i.e. the time to correct task accomplishment. It assessed the degree of ease of induction. *Immediate interactive behaviour* was defined as a (reaction) time to the first click that was faster than three seconds. Hence the time to first click was transformed in a dichotomous variable with the categories *immediate interactive behaviour* (time to first click < 3 sec) and *no immediate interactive behaviour* (time to first click ≥ 3 sec) for each of the nine trials. The trials with people showing immediate interactive behaviour have been summed up to get an overall score.

Results

A 2 (Affordances, (aff)) x 2 (Signals (sig)) Analysis of Variance (ANOVA) was calculated for the time to first click, the total task time, immediate interactive behaviour and subjective measures (NASA-TLX). A reparameterized model (Fox, 2008) has been built and contrasts for each hypothesis have been *t*-tested in form of parametric functions (*Ψ*c'*). In addition, effect sizes (Cohen's *d*) are reported for each test following Cohen's (1988) interpretation (small effect $d \geq 0.3$; medium effect $d \geq 0.5$; large effect $d \geq 0.8$).

Overall mean time to first click showed a significant, medium effect as to affordances ($t(44) = -2.059$, $p = 0.023$, $d = -0.651$) but neither significant effect as to signals nor as to the interaction of both factors (see Fig.3a). The influence of affordance was not significantly larger than the signal influence, although a small effect could be found ($t(44) = 1.003$, $p = 0.161$, $d = 0.317$) showing that the impact of affordances (natural indicators) on the time to first click is larger than the impact of signals (artificial indicators). Large immediate *learning effects* (see Fig. 3b) were obtained over all gestures ($t(44)= -3.48$, $p = 0.001$, $d = -1.1$). These learning effects did not differ between the conditions of affordances or signals or the interaction of both. No significant learning effect ($d \geq 0.3$) occurred from trial 2 to trial 3.



a) b)

Figure 3. The Effect of a) affodances and signals on TFC and b) learning on TFC

With respect to the total task time (time needed to manipulate a square) large *learning effect* was observed from trial 1 to trial 2 over all groups ($t(44)= -4.157$, $p < 0.001$, $d = -1.314$). The learning effect from trial 2 to trial 3 was significant but of smaller effect size ($t(44)= -1.868$, $p = 0.034$, $d = -0.591$). Analysis revealed a tendency for immediate interactive behaviour occurring more often in the group with affordances (Fig. 4) than in the group without affordances ($t(44) = 1.568$, $p = 0.062$, $d = 0.496$). For signal conditions, the interaction of signals and affordances as well as for the comparison of signals and affordances, neither significant differences nor notable effects were found.

For the subjective data (NASA-TLX) no significant main effects were found, neither for affordance nor signal. However, large effects of the repeated measures factor *time* were obtained for all scales of the NASA-TLX and the overall score. *F*-values were larger than $F = 5.49$, *p*-values smaller than $p = 0.01$ for all scales and effect sizes larger than eta = 0.35. Bonferroni corrected post-hoc tests showed that for almost all scales subjective strain decreased from first to second trial (*p*-values were smaller than $p = 0.01$) and from first to third trial (*p*-values were smaller than $p = 0.01$). Only for cognitive and temporal strain, subjects indicated no change in strain between first and third measurement. However, no scales differed significantly between second and third point in time (*p*-values were larger than $p = 0.20$).

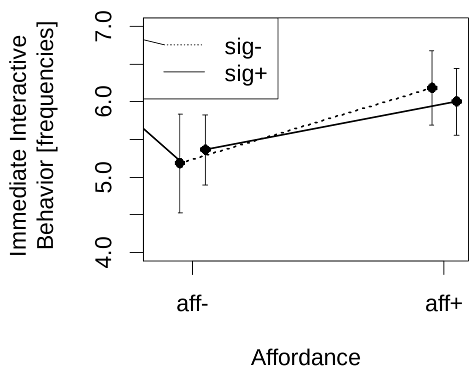


Figure 4. The effect of affordances (aff) and signals (sig) on IIB

Post-hoc power analysis

Post-hoc power analysis should estimate the minimum effect size of tests used in prior analyses of experimental data. Power analysis was calculated for a sample of $N = 48$, $\alpha = 0.05$, $\beta = 0.2$, and one-sided t-tests in the parameterized *2 x 2* ANOVA model. For a power of $1-\beta = 0.8$ effects of $|d| \geq 0.73$ should be found significant. Given the obtained power, it was only likely to detect large effects. Hence, small to medium effects might have been missed in the present study.

Discussion

The present work claims that the appeal of touch interfaces dates back to the way they structure the interaction environment. A multi-touch environment was set up testing the effects of one aspect of the structured environment on subjects' behaviour and experience. Therefore the effects of two forms of support (i.e. signals and affordances) on users' performance and experience were tested based on the model of user experience framework (Drewitz & Brandenburg, 2010). First, it was hypothesized, that the presence of affordances facilitates the users' performance and experience. Overall, empirical data revealed partial support for this hypothesis. More specifically, subjects initiated an interaction more quickly (i.e. time to first click was lower) if affordances were present. In addition, affordances fostered the occurrence of immediate interactive behaviour. The results are partially in line with previous findings (e.g. Brandenburg et al., 2013) and theoretical considerations (Scarantino, 2003; McGenere & Ho, 2000). For example Brandenburg et al. (2013) showed that affordances did foster the subjects' initial reaction to the artefact and the task completion times. However, here the effect of affordances depended on the presence of signals (Brandenburg et al., 2013). In the current study, affordances clearly outperformed signals regarding the time to first click and immediate interactive behaviour. Theoretical considerations about the inner workings of affordances deliver additional support for the presented results. Following Gibson (1979) and McGenere and Ho (2000), direct interaction fosters direct perception. In turn, direct perception leads to the fast execution of actions based on perceived affordances (see also Albrechtsen et al., 2001; Costall, 1984; Scarantino, 2003). Nevertheless, subjective strain ratings did not support the interpretation that users benefit from affordances in multi-touch interaction. Further investigations should use other methods to assess user experience in these contexts, since NASA-TLX is no user experience measure in the proper sense.

In addition to the hypothesized effects, large learning effects were obtained. Interestingly, participants mostly learned within the first trial. After having performed the gestures once, they were significantly faster in the second trial. After that no further significant learning occurred. This result is in line with previous work of Drewitz and Brandenburg (2010) and Brandenburg et al. (2013) who obtained similar learning curves for time to first click. However, this investigation also revealed some shortcomings. With respect to performance data, large error variances occurred which might have hindered smaller effects to become significant. One reason for this methodical issue is assumed to be due to missing information of location for the index fingers. In other words, if subjects did not know where to put the fingers, they were likely to search for the points of location at the beginning of the gesture execution process. Thus, this search operation might influence empirical as well as subjective data. Regarding subjective data, the present experiment focused on ease of use. Therefore, further studies should focus on the assessment of positive user experience as well. Hence emotions play an important role in user experience processes and influence subjective experience in touch interaction.

References

Albrechtsen, H., Andersen, H.H.K., Bødker, S., & Pejtersen, A.M. (2001). *Affordances in activity theory and cognitive systems engineering* (Report Risø-R-128). Roskilde, Denmark: Risø National Laboratory.

Anderson, J.R. (2000). *Kognitive Psychologie.* Heidelberg: Spektrum Verlag.

Backhaus, N., & Brandenburg, S. (2013). Wie viel ist zu viel? Empfehlungen für die Anzahl von Bedienelementen auf kleinen Displays. In E. Brandenburg, and H. Smiezek (Eds.), *Beiträge zur 10. Berliner Werkstatt Mensch-Maschine-Systeme* (pp.181–186). Düsseldorf: VDI-Verlag.

Brandenburg, S., Drewitz U., Thüring, M., Minge, M., & Brune, T. (2009). Touch me if you can - Immediate Interactive Behavior als Bedienkonzept für Multimediageräte. In A. Liechtenstein, C. Stößel, and C. Clemens (Eds.), *Der Mensch im Mittelpunkt technischer Systeme: 8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 248–252). Düsseldorf: VDI Verlag.

Brandenburg, S., Vogel, M., & Drewitz, U. (2013). User experience starts at the keystroke level: The Model of User Experience (MUX). In A. Marcus (Ed.), *DUXU/HCII 2013, Part I, LNCS 8012* (pp. 449–458). Berlin, Heidelberg: Springer.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Costall, A.P. (1984). Are there theories of perception necessary? A review of Gibsons' the ecological approach to visual perception. *Journal of the experimental analysis of behavior, 41*, 109–115.

Davis, E.D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *Management Information Systems Quarterly, 13*, 319–340.

Drewitz, U., & Brandenburg, S. (2010). From design to experience: Towards a process model of user experience. In J.C. Lin, D.M. Lin, and H. Chen (Eds.), *Ergonomics for all: Celebrating Ppcoe's 20 years of excellence: Selected papers of the Pan-Pacific Conference on Ergonomics* (pp. 1-6). Oxford: CRC Press.

Fox, J. (2008). *Applied regression analysis and generalized linear models* (2nd edition). Los Angeles, CA: Sage.

Gibson, E.J., (2000). Where is the Information for Affordances? *Ecological Psychology, 12*, 53–56.

Gibson, J. (1979). *The Ecological Approach to Visual Perception.* Boston, MA: Houghton Mifflin.

Greeno, J.G. (1994). Gibson's Affordances. *Psychological Review, 101*, 336–342.

Hart, S.G., & Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. Hancock, and N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). Amsterdam: Elsevier.

Kurata, Y., & Egenhofer, M. (2005). Semantics of Simple Arrow Diagrams. In T. Barkowsky, C. Freksa, M. Hegarty, and R. Lowe (Eds.), *AAAI Spring Symposium on Reasoning with Mental and External Diagram: Computational Modeling and Spatial Assistance* (pp. 101–104). Pala Alto, CA: AAAI Press

McGenere, J., & Ho, W. (2000). Affordances: Clarifying and evolving a concept. In Canadian Human-Computer Communications Society (Ed.) *Proceedings of Graphics Interface 2000* (pp. 179–186). New York, NY: Lawrence Erlbaum Associates.

Neth, H., Carlson, R.A., Gray, W.D., Kirlik, A., Kirsh, D., & Payne, S.J. (2007). Immediate Interactive Behavior: How Embodied and Embedded Cognition Uses and Changes the World to Achieve its Goals. In D.S. McNamara, and J.G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 33–34). New York, NY: Lawrence Erlbaum Associates.

Newell, S. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Norman, D.A. (1999). Affordance, conventions, and design. *interactions, 6*, 38–43.

Norman, D.A. (1988). *The Psychology of Everyday Things.* New York, NY: Basic Books.

Petocz, A., Keller, P.E., & Stevens, C.J. (2008). Auditory warnings, signal-referent relations, and natural indicators: Re-thinking theory and application. *Journal of experimental psychology: Applied*, *14*, 165–178.

Scarantino, A. (2003). Affordances explained. *Philosophy of science*, *70*, 949–961.

Vogel, M., Brandenburg, S., & Drewitz, U. (2011). Zur Gestaltung grafischer Benutzerschnittstellen: Einflussfaktoren für das Nutzererleben. In M. Eibl (Ed.), *Mensch & Computer 2011: überMEDIEN|ÜBERmorgen* (pp. 111–120). München: Oldenburg.

Transitional Journey Maps: Capturing the dynamics of operational policing

Reinier J. Jansen, René van Egmond, Huib de Ridder, & Sacha Silvester
Delft University of Technology, The Netherlands

Abstract

Operational police work can be characterized by the continuous switching between surveillance, responding to incidents, and office activities. Transitions between these activities are initiated by radio contact, messages on a mobile data terminal, or personal observations. The "information environment" emerging from these channels may cause cognitive overload during demanding activities. Although the notion of fragmented work is acknowledged in police literature, detailed descriptions are lacking. The goal of this study is to better understand cognitive load in police officers by capturing the dynamics of operational policing. Ten officers of the Dutch police force were accompanied while on patrol with their car. The method of contextual inquiry was used to collect 28 hours of data. Activities were mapped on a pre-defined set of categories. Attention was paid to how officers experienced their information environment while performing these activities. All was captured in the Transitional Journey Map, a new method to visualize workflow. The Transitional Journey Map augments a sequence of activities with experiential and contextual information. This method was used to identify cognitive overload situations and differences between solo and dual patrol work. These insights are relevant for improving the information system that assists officers in their patrol vehicle.

Introduction

"After a short break at the police station, C. and S. return to their surveillance duty. The dispatcher calls: 'A missing young girl possibly showed up at relatives and should be picked up.' While S. tries to write down the address in his notebook, they realize they missed the girl's full name and the house number. S. feels stupid for having to ask again. Directly afterwards an alarm goes off. S. glances at the mobile data terminal: 'It's a white vehicle with an unpaid fine.' C. looks around, locates the car, and immediately makes a turn. Just as S. tries to request information on the driver, the dispatcher interrupts him: 'We are detaching you from the previous call. Someone has been spotted in a building that burned down last week.' C. recognizes the address, turns the car again, and accelerates. On their way, S. declines another alarm with a lower priority. They arrive at the scene only minutes later, to find a man in ragged clothes carrying a bag full of copper." (field notes from the present study)

This example illustrates how police officers are faced with a perpetual switching between activities, often resulting from the technology that surrounds them. The

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Dutch police force is currently looking for ways to improve the information system of their police vehicles, including pushing information (e.g., neighbourhood updates, on-board training) to the vehicle. A central question is how much information officers can process in various work situations. Streefkerk et al. (2006) argued that a mobile police information system should be context-aware (i.e., involving time, location, environmental, and social factors) to prevent cognitive overload. This implies that the dynamics of police work should be taken into account for the development of such a system. For example, indications on average time spent on an activity and the corresponding cognitive load may assist in determining the length and appropriateness of an information event (i.e., a moment during which information is presented). As will turn out, a detailed description of work dynamics is lacking in police literature. Therefore, the goal of this study is to better understand cognitive load in police officers by capturing the dynamics of operational policing. First, a background on operational policing, equipment, and activities is discussed. Next, a new approach to visualize workflow is applied to the data of a field study.

Operational policing in The Netherlands

The Dutch police organisation consists of one national unit, and ten regional units. The national unit deals with, e.g., highway patrol, organized crime, and terrorism. In addition, two types of police work can be found in a regional unit (Stol et al, 2004): community policing (Dutch: 'gebiedswerk') and operational policing (Dutch: 'noodhulp'). Community policing is pro-active and preventive, and involves considerable time on networking with civilians (Stol et al., 2004; Smith et al., 2001). Operational policing on the other hand is mostly reactive: it is time and safety critical work based upon officers attending incident sites by car (Sørensen & Pica, 2005). The scope of this paper is operational policing. For this type of work, each police station employs a number of concurrently operating police vehicles. Two officers usually occupy one vehicle, although some regions are experimenting with additional solo patrol vehicles. When an operational police team in The Netherlands is not assigned to a current call, officers typically spend their time on criminal investigation (i.e., based on assignments handed out during briefing), or law enforcement (e.g., surveillance, traffic control).

A police officer's information environment

The information environment of a patrolling officer consists of numerous concurrent visual and auditory channels. The patrol car is equipped with a specialized in-vehicle information and communication system. The Mobile Data Terminal (MDT) is a touch screen device positioned on the vehicle console in-between the driver and co-driver, providing the officer with a number of functionalities (see Figure 1). First, the vehicle is equipped with an Automatic Number Plate Recognition (ANPR) system. This system compares license plates scanned by on-board cameras with a database of delicts linked to specific number plates. In case of a 'hit', an alarm can be heard through the car's speakers, and information on the vehicle is displayed on the MDT. Additionally, officers use the MDT to acquire information on a person, to control the lights on top of the car, and for navigation.

Figure 1. View on the cockpit of a police vehicle. The Mobile Data Terminal is located between the driver and co-driver.

There are two main modes of communication between the control room and patrolling officers: direct contact using a mobile phone, and two-way broadcasting. Regarding the latter, officers are equipped with a portophone for radio contact, which consists of an earpiece, a microphone, and a channel selector. Additionally, the vehicle's interior loudspeakers may be used. Broadcast radio messages typically start with a numerical code consisting of the region and the team it is intended for. Consequentially, officers continuously monitor incoming codes to detect if a call is meant for them. Pen and paper are used to memorize details of a call, as well as observations made when dealing with a call.

Monitoring this complex environment may have consequences on performance. Multiple Resource Theory (MRT) predicts that time-sharing between two tasks is best when they require the use of different processing stages (e.g., cognitive vs. response), processing codes (e.g., spatial vs. verbal), and modalities (e.g., visual vs. auditory) (Wickens, 2008). However, the independence of modalities claimed by MRT has been criticized. For example, Spence and Read (2003) showed that dual-task performance decreases when the spatial location of an auditory speech shadowing task does not coincide with the spatial location of a visual driving simulator task. Since police officers typically monitor incoming messages through their earpieces (i.e., from one side), one can expect lower dual-task performance than would be predicted by MRT. These decrements may be enlarged when the traffic conditions become more demanding (Patten et al., 2006), for example during pursuits and high priority calls. Additionally, Anderson et al. (2005) found that police officers frequently perform more than two tasks at a given time, which may also result in performance decrements (e.g., Recarte & Nunes, 2003). Therefore, designing the cockpit of a police vehicle requires an understanding of the situations in which in-vehicle technologies are used.

Fragmentation in police work

Lundin and Nuldén (2007) identified five ways in which Swedish officers used their patrol car: *'on their way to an incident'*, *'on their way from an incident'*, *'at the site of an incident'*, *'general surveillance when driving around or parked at a specific location'*, and *'parked at the station handling detained people or paperwork'*. A comparable categorization was found in a study on British police officers interacting with mobile technology (Sørensen & Pica, 2005). Here, the researchers distinguish five primary activity types: *'waiting in the car before an incident'*, *'driving to an incident'*, *'taking action at the incident'*, *'driving from the incident'*, and *'waiting in the car after an incident'*. Furthermore, they emphasize that this so-called *'generic cycle of operational policing'* can be interrupted and rearranged due to intermediate events (e.g., incoming calls with a higher priority). Borglund and Nuldén (2012) share this statement, identifying work rhythm as problem area in the Swedish police force: *"Much of police work is characterized by interruptions. Planned and ongoing activity can be discontinued at any time. Current routines and access to computer-based systems create a somewhat fragmented work situation for the officers."* Similar accounts have been reported for the U.S. (Straus et al., 2010) and Dutch (Bouwman et al., 2008) police forces. Thus, the notion of fragmented work seems acknowledged in literature on operational policing.

Given the continuous switching between activities, it is important to not only focus on stationary cognitive load during an activity, but also to consider the effects of transitions between activities on cognitive load. Yet, detailed investigations into police routines are typically represented through activity statistics using a full work shift as time window (e.g., Anderson et al., 2005; Frank et al., 1997; Smith et al., 2001). These statistics do not provide information on whether an activity is executed without interruptions, or about patterns of fragmentation. Moreover, these investigations do not reflect police officers' subjective experiences related to these activities. While attempts to characterize police work fragmentation using scenarios (Borglund & Nuldén, 2012) or narratives (Sørensen & Pica, 2005) do include subjective experiences, they too fail to quantify fragmentation. Therefore, the present study aims to unite a quantitative description of work dynamics with subjective experiences related to cognitive load.

Method

A series of ride-alongs with Dutch police officers were arranged. Based on the method of contextual inquiry (Beyer & Holtzblatt, 1997), officers were interviewed and observed in their natural work environment, where they provided explanations as their work unfolded.

Participants

Ten officers (8 males, 2 females) volunteered to be accompanied in their patrol cars. Each officer had at least two years of experience with operational policing. Four ride-alongs were arranged, including three full eight-hour shifts and two shift changes in total. Hence, the vehicle was chosen as central focus during ride-alongs, while personnel configurations changed from shift to shift. The ride-alongs included solo (2 cases) and dual patrol (4 cases). With durations varying between 4,5 and 11 hours, in total 28 hours of data were collected. Colleagues of the officers often asked the researcher about his presence during stops at the police station. Their comments on work dynamics and organization are treated as part of the study results.

Apparatus

Data were collected with pen and paper, featuring timestamps, descriptions of the current activity, events in the officer's information environment that caused a transition to another activity (e.g., incoming calls, comments following an officer's observation), and utterances that expressed influences on cognitive load. All data were logged on a template with six rows of pre-defined activity categories. These categories were based on the framework by Lundin & Nuldén (2007), consisting of 'driving to the incident', 'engaging at the incident', 'driving to the station', 'driving surveillance', 'parked surveillance', and 'parked at the station'. As opposed to the original framework, surveillance activities were logged separately for the driving and non-driving category, as the corresponding cognitive load is likely to be different.

Procedure

Before the ride-along began, the researcher explicitly stated that the study was not intended to judge the officers' performance. Agreements were made on safety and privacy. During the ride-alongs, the researcher tried to minimize hindrance by discretely observing what was going on. This non-participatory research approach was at times violated, for example, when an officer asked for details about a recent call. Existing studies recommend that the relationship with the officer should not be sacrificed for the sake of minimizing reactivity (Stol et al., 2004; Spano & Reisig, 2006). Interestingly, such a question can be regarded as a verbalization related to high cognitive load. Officers were occasionally asked to explain what happened during transitions, but only if the work demands allowed for such concurrent reports. Otherwise, they were asked to give a retrospective report shortly after the event.

Results

A new method to visualize workflow will be introduced. The method is used to report findings on cognitive overload, and differences between solo and dual patrol.

Transitional Journey Maps

Connecting consecutive activities and experiences logged in the data template gives a sense of order and time. The graphical representation of these objective and subjective data is referred to as *Transitional Journey Map*. Four Transitional Journey Maps were constructed, one for each ride-along. An example can be found in the lower part of Figure 2. The excerpt shown on top corresponds with the anecdote introducing this paper. The vertical axis displays six activity categories, whereas time is found on the horizontal axis. The main actors are represented through three thick lines: the police vehicle (red violet), the driver (dark blue), and in case of dual patrol, the co-driver (light blue). A journey through activity

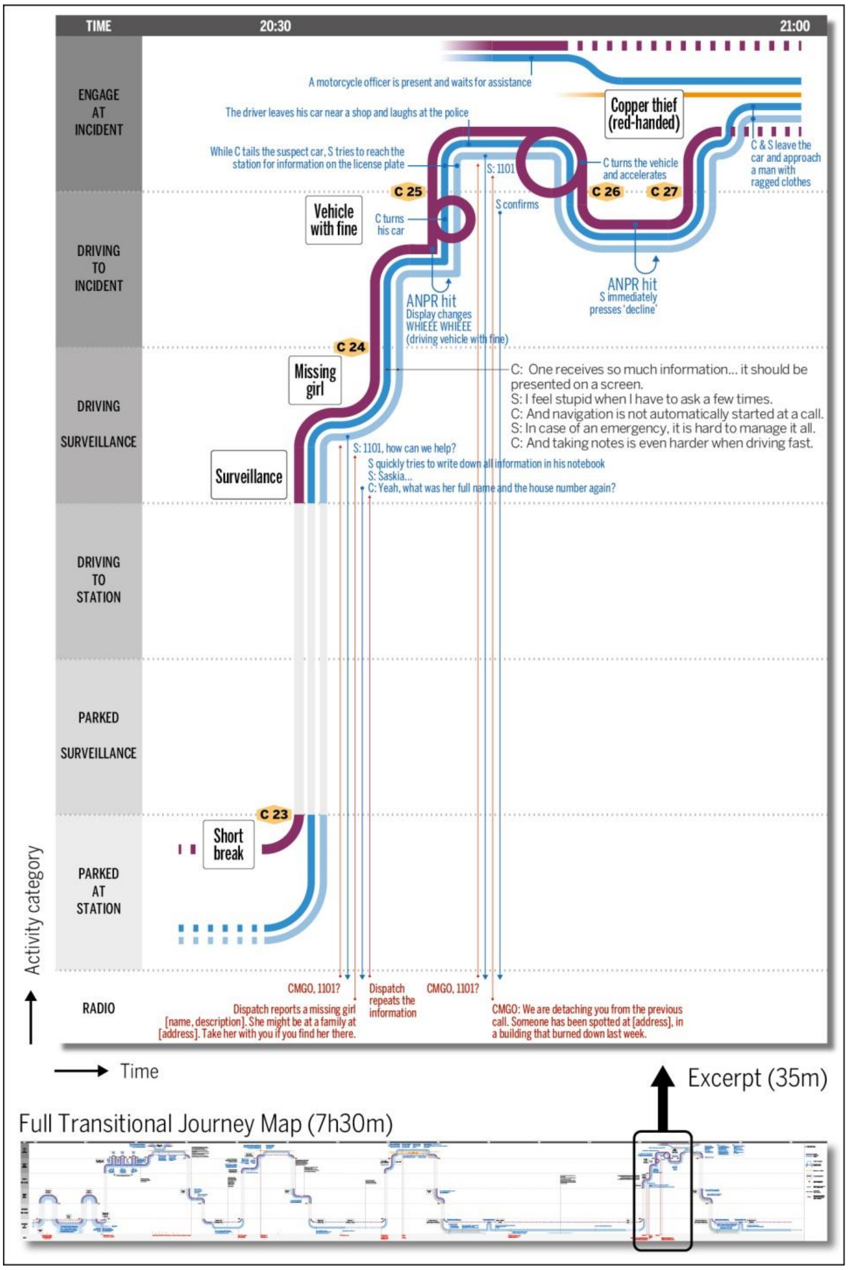


Figure 2. Excerpt of a Transitional Journey Map. The horizontal and vertical axes display time and six activity categories, respectively. Coloured lines refer to the vehicle (red violet), driver (dark blue), and co-driver (light blue). Additional details are described in the text.

categories is created as the actors cross the underlying framework. Additional lines may be used in case other actors come into play (e.g., the case of the copper thief, here represented in orange). Stationary vehicles are depicted with a dashed line.

Similarly, dashed lines are used when officers are having a break. Upon entering their car, officers' corresponding lines are joined with the vehicle's line. Segments of activities are demarcated by the time between adjacent transitions.

A transition is defined as a change from one activity category to another one. In Figure 2, transitions are labelled with yellow boxes, a character for the corresponding ride-along, and a number for the order of occurrence. For example, 'C23' refers to a segment of previous activity at the police station, and marks the transition from 'parked at station' to 'driving surveillance'. Descriptions for ongoing activities are depicted in white boxes for quick reference. Because of its dominant role in police work, instances of radio communication can be found in a separate row. The thin alternating red and blue lines in Figure 2 show how messages are going back and forth between the officers and the dispatcher (e.g., the call of the missing girl).

Applying Transitional Journey Maps to operational policing

Visual inspection of a full Transitional Journey Map confirms the notion that police work is fragmented. The lower part of Figure 2 shows periods of many short activity segments followed by relatively long stretches of paperwork at the police station. This is reflected in the boxplots of Figure 3, which show the durations of activity segments per activity category, including all ride-alongs.

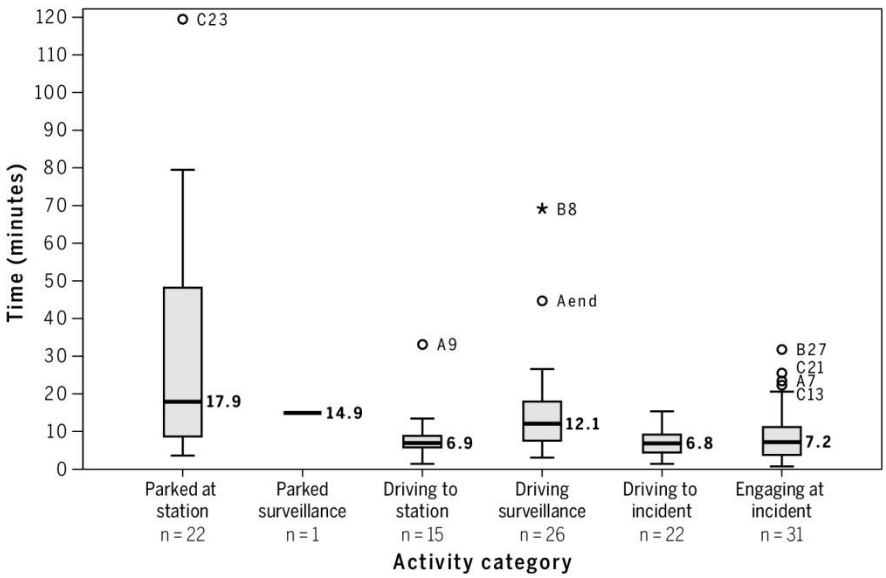


Figure 3. Boxplots of time spent in each activity category, summarized over all ride-alongs. Median values are shown next to each box. Whiskers depict the lowest and highest datum within a 1.5 interquartile range of the lower and upper quartile, respectively. Outliers labelled with an asterisk or a circle concern solo and dual patrol, respectively.

Outliers in the 'engaging at incident' category were cases where victims and/or suspects were questioned, namely theft (A7, C21) and domestic violence (B27, C13). All of these cases required more than half an hour of paperwork, with an outlier at two hours (C23). However, officers were often interrupted by incoming calls before finishing their office work, as reflected by the median duration of 17.9 minutes. Other outliers refer to picking up remote colleagues (A9), surveillance while bringing the researcher to the train station (Aend), and surveillance across a deserted national park (B8).

The categories ‘parked at station’ and ‘driving surveillance’ seem to take longer than ‘driving to station’, ‘driving to incident’, and ‘engaging at incident’, who seem to have similar segment durations (see Figure 2). Given the skewed distributions, non-parametric tests (SPSS v20) were used to compare between activity categories. As only one instance of ‘parked surveillance’ occurred, this category was excluded from further analysis. Segment duration is significantly affected by activity category ($H(4) = 23.71$, $p < .001$). Seven Mann-Whitney tests were used to follow up this finding. Therefore, a Bonferroni correction was applied, and all effects are reported at a .007 level of significance. The duration of activities in the ‘parked at station’ category was generally significantly longer than ‘driving to station’ ($U = 70$, $r = -.48$), ‘driving to incident’ ($U = 90$, $r = -.54$), and ‘engaging at incident’ ($U = 148$, $r = -.48$), but not longer than ‘driving surveillance’ ($U = 208$, $r = -.23$). Furthermore, activities performed in the ‘driving surveillance’ category took significantly longer than ‘driving to incident’ ($U = 139$, $r = -.44$) and ‘engaging at incident’ ($U = 235$, $r = -.36$), but not longer than ‘driving to station’ ($U = 108$, $r = -.37$). It can be concluded that most time for an informing event can be found when officers are working at the police station, or during surveillance while driving. Based on this dataset, an informing event should take less than 6.8 minutes if at least half of these events are to be fully processed in any activity category before a next transition takes place. However, these statistics do not address whether officers have spare capacity to successfully process the information.

Reports of cognitive overload

Comments by police officers regularly contained descriptions of situations witnessed during other ride-alongs, which were indicators for cognitive overload. For example, compare the following anecdote with Figure 2:

“An incoming call instructs the officers to advance to a car that was broken into. L. takes a notebook from her pocket to record the address: ‘This way you don’t have to ask again.’ A. responds: ‘On the group radio one often hears colleagues asking for a repetition of the suspect description. At the time they receive a call and they have to move as fast as possible, their mind set is already preoccupied.’” (field notes ride-along A)

Because of the activity descriptions and their characteristic visual pattern, the layout of a Transitional Journey Map facilitates remembering and retrieving events with related comments. Furthermore, the content of a comment dictates in which activity category it should be placed (e.g., a colleague at the station talking about an arrest belongs to ‘engaging at the incident’). Thus, an overview of information processing issues within an activity category can be obtained by scanning along the corresponding row in the Transitional Journey Maps. This approach resulted in the identification of an information processing paradox.

On the one hand, police officers not only monitor the radio for messages addressed to themselves, but they also want to stay informed on the whereabouts and tasks of their colleagues. One reason is safety: *"If there is a call with violence, it's good to know if colleagues are nearby... then you know if and how long you should wait before stepping in."* Vice versa, officers may offer assistance. Second, there are functional implications: *"Those officers are busy over there, so I'll compensate by patrolling more centrally in this area."* Finally, it is part of a social system: when returning at the station after an emotionally demanding call, officers find support from colleagues that listened in. One officer commented that he was missing too much information, even though three channels were concurrently monitored (i.e., car radio and two earpieces).

On the other hand, police officers have trouble in processing all information. As described above, incoming calls regularly contain too much information to remember. This is further inhibited by situational and state related factors: *"If a situation is dangerous, you feel the adrenaline, stress, fatigue and tension, and this affects your ability to concentrate. In those situations it is hard to hear something amidst other voices."* Messages are often hard to comprehend due to auditory masking by the police vehicle (e.g., when driving at high speed, often accompanied by a siren) and signal degradation in the communication system. In the meanwhile, the continuous monitoring and filtering of radio messages takes its toll. Up to 26 messages were counted in a time span of five minutes. Officers complained about high volumes, occasional feedback beeps, and fatigue: *"My left ear is deaf for other sounds because of the earpiece. After a busy shift I still hear the voices at home."*

Comments on the necessity of monitoring radio communication were found in all activity categories, except for 'parked surveillance'. However, the representativeness of this exception is doubtful, since action in this category was observed only once. Comments on auditory masking were found in all activity categories that involved driving. Comments on overload were found in all activity categories, except for 'driving to station.' Overall, the observations and comments suggest that police officers want more information than they can handle with the current system.

Comparison solo and dual patrol

All outliers in Figure 3 were cases of dual patrol, except for B8. This suggests a considerable difference in time spendings between solo and dual patrol, and as a result, more time for information events during dual patrol. Non-parametric tests were performed per category. Using a Bonferroni correction, the effects were compared with an alpha level of .008. None of the tests on time spending reached statistical significance. Nevertheless, police officers did mention differences between solo and dual patrol modes. The biggest impact is the opportunity to distribute tasks among officers in case of dual patrol. Generally, the driver only concentrates on driving, whereas the co-driver is responsible for communication and surveillance tasks. Many officers commented that it is hard to operate the mobile

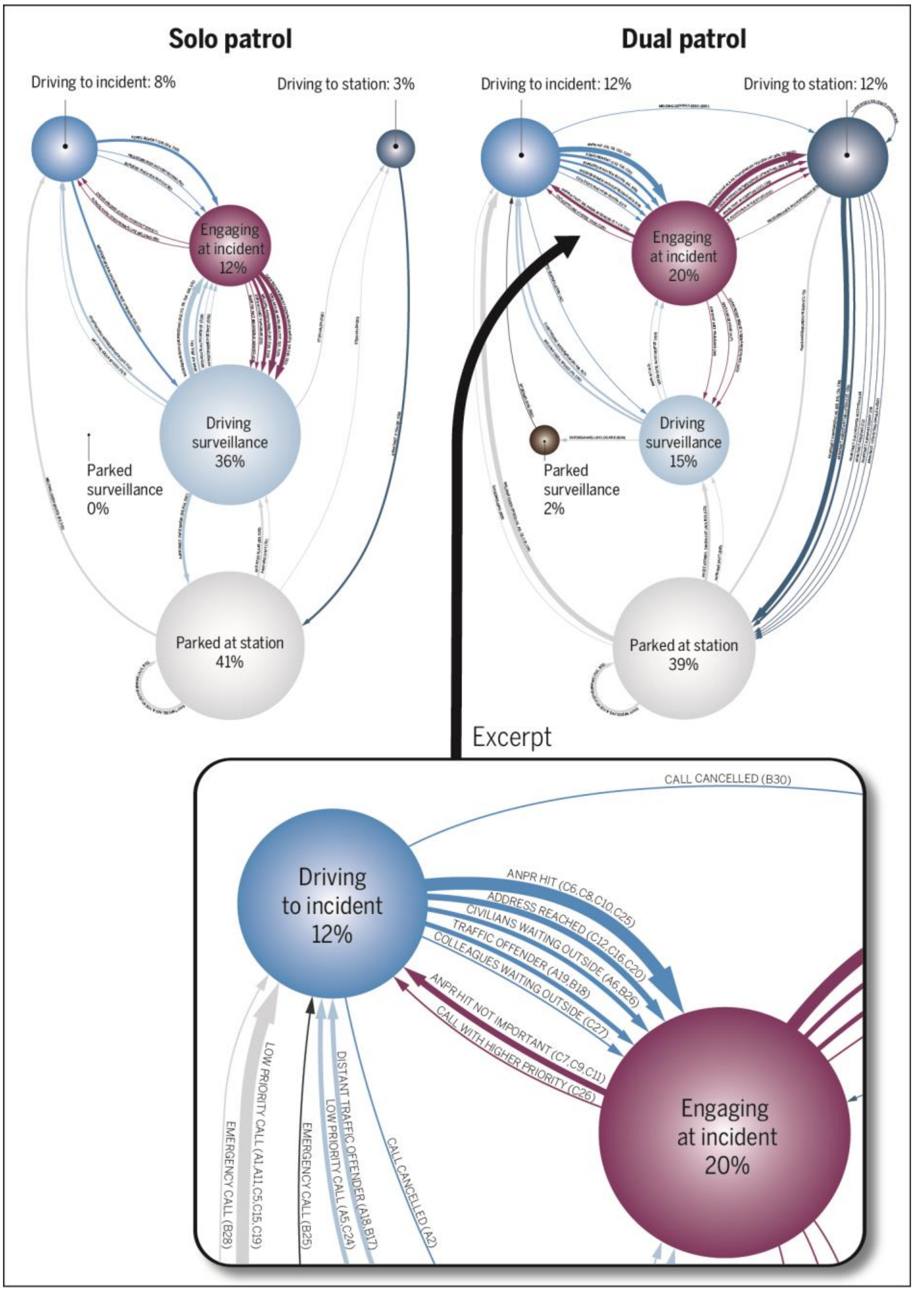


Figure 4. *State diagrams of transitions during solo patrol (top left, 12.6 hours observed) and dual patrol (top right, 15.1 hours observed). The excerpt shows observed causes for transitions between activity categories. Codes in parentheses refer to transitions in the Transitional Journey Maps.*

data terminal while driving solo. Additionally, there are organizational differences between the patrol modes: *"If you're patrolling solo, you only get a call when the others cannot handle it. In cases of violence we always operate with couples."* This suggests that differences may be found between the distributions of transitions.

Figure 4 depicts state diagrams for solo and dual patrol. An arrow line represents each cause for a transition between two activity categories. Thicker lines are used if the same cause was observed more than once. The total time spent observing solo and dual patrol were 12.6 hours and 15.1 hours, respectively. The relative time spent in each activity category is represented by the size of the corresponding circles. The two figures reveal that solo patrol involves relatively more 'driving surveillance' activity than dual patrol (36% vs. 15%). Solo patrol involved more transitions from 'driving surveillance' to 'engaging at incident' (10 vs. 2), but less transitions from 'driving to incident' to 'engaging at incident' (5 vs. 12). Interestingly, in both patrol modes 15 transitions were counted towards 'engaging at incident'. However, relatively more time on 'engaging at incident' was spent in dual patrol (20% vs. 12%). This was caused by the longer times spent on investigating incidents with violence (see outliers in Figure 2). Additionally, dual patrol involved more time spent on 'driving to station' (12% vs. 3%), which may be due to the large amount of paperwork after serious incidents, and a higher likelihood of transporting victims or suspects afterwards. In sum, the state diagrams on solo and dual patrol reflect the organizational differences uttered by the police officers.

Discussion

The Transitional Journey Map is a model to describe workflow, by connecting objective and subjective data along a timeline. Similar models have been developed, but there are a number of structural differences. Although the customer journey map used in service design (e.g., Zomerdijk & Voss, 2010) served as inspiration, it is based on a fixed sequence of consumption activities. In contrast, the Transitional Journey Map is aimed at capturing unpredictability, which is one of the essential characteristics of operational police work. The Cognitive Pathway visualization, which was developed to map the workflow of nurses (Wolf et al., 2006), does consider unpredictability. However, it does not include the subjective experiences of participants, or the interaction with other professionals in the work domain. Therefore, the Transitional Journey Map, with its emphasis on time and transitions, may be a valuable addition to the existing palette of work models (e.g., Beyer & Holtzblatt, 1997). In case the time dimension is less relevant for analysis, the derivative state diagrams could offer insights from a different perspective. Finally, Transitional Journey Maps may also be employed as scenarios to compare alternative solutions in a design context.

It has been argued that knowledge on work dynamics and related cognitive load is beneficial for the development of an information and communication system. Regarding work dynamics, the current data suggest that in most cases an information event that takes place while driving may not take longer than 6.8 minutes if it is to be fully processed before a next transition takes place. However, this finding does not address cognitive load. The comments made by officers suggest that the demands of concurrent driving, surveillance, and monitoring force them into a permanent state of task-related effort (De Waard, 1996). Therefore, although no complete performance breakdowns were observed, the continuously experienced high workload was exhausting in the long run. ‘Driving to station’ is the only activity category that may be used for additional information events, given the absence of comments on cognitive overload. However, the applicability would be limited during solo patrol. Overall, the data suggest that the means with which police officers currently obtain information from the system should be improved, before pushing of additional information can be considered.

The present study was an initial exploration into the dynamics of operational policing. Therefore, more data on segment durations are required before a final statement can be made on opportunities for information events. More specifically, exposure to activities in the ‘parked surveillance’ category is needed, as well as observations during early mornings. Police officers may have been tempted to show parts of their work that they considered most interesting to show. An automated logging system may help in gathering a more detailed work pattern. However, this would be at the cost of subjective experiences.

Conclusion

The Transitional Journey Map, a new visualization method, describes workflow by connecting observed events with subjective experiences on a timeline with a set of activity categories. This method was applied to the data of a field study on operational policing. Embedded in an unpredictable workflow, the information environment of a police officer includes numerous visual and auditory information channels. Analysis of workflow fragmentation and officers’ comments on cognitive overload suggests that the police vehicle’s cockpit should be improved, before pushing additional information can be considered. Particularly, alleviation of the auditory channel is needed. The current results warrant further application of the Transitional Journey Map method in other contexts, such as information processing by other emergency services, and tracking group dynamics for crowd management.

Acknowledgments

This study was carried out at the request of, and partially funded by Staf Korpsleiding / Directie Facility Management, Nationale Politie. The authors wish to thank all police officers who volunteered for their cooperation.

References

Anderson, G.S., Courtney, A., Plecas, D. & Chamberlin, C. (2005). Multi-tasking behaviors of general duty police officers. *Police Practice and Research, 6,* 39-48.

Beyer, H. & Holtzblatt, K. (Eds.) (1997). *Contextual design: defining customer-centered systems*, San Francisco: Morgan Kaufmann.

Borglund, E. & Nuldén, U. (2012). Personas in uniform: police officers as users of information technology. *Transactions on Human-Computer Interaction, 4,* 92-106.

Bouwman, H., Haaker, T. & de Vos, H. (2008). Mobile applications for police officers. In *BLED 2008 proceedings*, *Paper 39*.

De Waard, D. (1996). *The measurement of drivers' mental workload*. PhD thesis, University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre.

Frank, J., Brandl, S.G., & Watkins, R.C. (1997). The content of community policing: a comparison of the daily activities of community and 'beat' officers. *Policing: An International Journal of Police Strategies & Management, 20*, 716-728.

Lundin, J. & Nuldén, U. (2007). Talking about tools – investigating learning at work in police practice. *Journal of Workplace Learning, 19*, 222-239.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L., & Svenson, O. (2004). Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis & Prevention, 36*, 341-350.

Recarte, M.A. & Nunes, L.M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Applied Experimental Psychology, 9*, 119-137.

Smith, B.W., Novak, K.J. & Frank, J. (2001). Community policing and the work routines of street-level officers. *Criminal Justice Review, 26*, 17-37.

Sørensen, C. & Pica, D. (2005). Tales from the police: rhythms of interaction with mobile technologies. *Information and Organization, 15*, 125-149.

Spano, R. & Reisig, M.D. (2006). "Drop the clipboard and help me!": the determinants of observer behavior in police encounters with suspects. *Journal of Criminal Justice, 34*, 619-629.

Spence, C. & Read, L. (2003). Speech shadowing while driving: on the difficulty of splitting attention between eye and ear. *Psychological Science, 14*, 251-256.

Stol, W.P., van Wijk, P., Vogel, G., Foederer, B., & van Heel, L. (Eds.) (2004). *Politiestraatwerk in Nederland: noodhulp en gebiedswerk* [Police streetwork in The Netherlands]. Zeist, The Netherlands: Kerckebosch.

Straus, S., Bikson, T., Balkovich, E., & Pane, J. (2010). Mobile technology and action teams: assessing BlackBerry use in law enforcement units. *Computer Supported Cooperative Work (CSCW), 19*, 45-71.

Streefkerk, J.W., van Esch-Bussemakers, M.P., & Neerincx, M.A. (2006). Designing personal attentive user interfaces in the mobile public safety domain. *Computers in Human Behavior*, *22*, 749-770.

Wickens, C.D. (2008). Multiple resources and mental workload. *Human Factors, 50*, 449-455.

Wolf, L.D., Potter, P., Sledge, J.A., Boxerman, S.B., Grayson, D., & Evanoff, B. (2006). Describing nurses' work: combining quantitative and qualitative analysis. *Human Factors, 48*, 5-14.

Zomerdijk, L.G. & Voss, C.A. (2010). Service design for experience-centric services. *Journal of Service Research, 13*, 67-82.

Usage of a Driving Simulator in the Design Process of new HMI concepts for eco-driving Applications (eCoMove project)

Eva García, Rosa Blanco, Abel Nogueira , Moisés Rial
CTAG (Centro Tecnológico de Automoción de Galicia)
Porriño (Pontevedra)
Spain

Abstract

In the framework of the eCoMove project, CTAG has carried out a study in its Dynamic Driving Simulator focused in analyses the most suitable way to present speed and gear recommendations to the driver. The aim of this study was to provide useful recommendations based on the results obtained during the development of the new concepts of ecoHMI carried out within the project. In order to analyse the best way to display the recommendations three different options to display the information were compared. The experimental design defined was a within-subject experiment where 24 participants drove through all the experimental conditions. With the purpose to evaluate the perceptions and opinions that the participants had about the system, different subjective measures were collected and in order to evaluate workload the DALI subjective tool was also used. After a general introduction of the eCoMove project and the purpose of the study, this paper will describe the methodology defined to carry out the experiment and the main results and conclusions obtained in terms of speed and gear recommendations for an eco-driving support system.

Introduction

eCoMove project aims to reduce fuel consumption (and therefore CO2 emissions) by applying the latest vehicle-to-infrastructure and vehicle-to-vehicle communication technologies. The project will create an integrated solution comprising eco-driving support and eco-traffic management to tackle the main sources of energy waste by passenger and goods vehicles having the target of the 20% reduction of fuel consumption (CO2 emmissions).

Within the project, one important aspect is the HMI development as it is a key factor for the acceptation and usage of the system by the drivers, especially for the on-trip advices. For this reason, the design process of the HMI was done interactively and focussing on a user centred design approach, including a previous study based in surveys and empirical studies in driving simulators, previously conducted to develop the final concepts.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Main objective of this study was the analysis, during the development phase, of different aspects related with the HMI for the systems developed in eCoMove project, aiming to clarify and solve some design concerns:

- Disposition of information on the screen.
- How to present the gear and speed recommendations (position of icons, clarity, colours…).
- Analysing distraction and driver workload in terms of graphical design.
- Estimation of time for the information in the display.

The study was performed using a general HMI designed in the previous phases of the project and allowing the testing in the driving simulator. This concept was designed taking into account the main results of one pre-study (survey) carried out in the previous steps.

Method

The tests were carried out in the CTAG's Dynamic Driving Simulator. This simulator is composed of the following subsystems: the movement platform with 6 DOF and 4000 Kg payload, the visual system (180º cylindrical screen and 3 rear view displays), the acquisition and control systems and the SCANeR© II (release 2.24) software, which builds a realistic virtual environment.

The instrumented vehicle inside the cabin is a commercial vehicle, in which the only changes done in the car were the replacement of the steering-wheel by the Active Steering Wheel System and the sensors mounted in pedals, gear stick, etc. The vehicle has an automatic gear change, but, for this experiment, the manual mode was used. A 7" screen was placed on the up side of the IPC in order to display the graphical information to the driver.

Figure 1. CTAG Dynamic Driving Simulator

Scenarios and test routes

The road network selected for the study was placed in a rural area, without many buildings. The road was composed of straight sections, curves and intersections (including roundabouts). All the roads used for testing had one lane in each direction. There were different values for the speed limit. These values varied between 50 and 90 Km/h, and the traffic density could be light or medium.

Figure 2. Image from the scenario

Three different routes were selected to test each option for HMI. The routes have very similar characteristics and they are located in the same areas of the scenario and the length was the same for all three routes, around 4 kilometres and 500 meters.

Description of HMI

The HMI proposed for the study was based in a screen located on the up side of the central screen of the vehicle (right side of the driver), three different graphical proposals for displaying the information were developed, changing the position and the order of the information presentation (i.e. recommended gear and speed).

Figure 3. Position of the display in the vehicle

In general, the layout proposed for the screen was divided into five different areas (1, 2, 3, MAPs and Reserved areas):

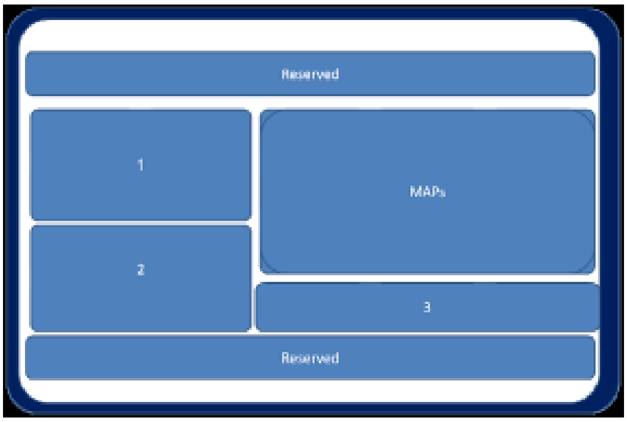


Figure 4. Layout of the Screen

- The upper and lower areas are reserved.
- The left side is the main driving information area (numbers 1 and 2 in the image), in which the recommended and current speed and gear information will be located.
- The right side is separated into two areas. The upper area shows the route information (map) and the lower area shows the ecoInformation advices (use of energy consumers inside the vehicle, maintenance information and load).

The proposals A and B differed only in the position of the gear and speed information: in proposal A the gear is shown in the position 1 and speed in position 2. In the proposal B, the location of the information is just the opposite (speed is in position 1 and the gear in position 2).

Figure 5. Proposal A (left) and proposal B (right)

In the proposal C, the areas 1 and 2 were treated as only one area where speed or gear information is displayed at a time (no simultaneous recommendation of speed and gear). This information was provided automatically and alternatively every 5 seconds (areas 1+2).

In all proposals (A, B and C) the ways in which the map information (see area ‘MAPs’) and the ecoInformation advices (see area 3 in the Layout) are displayed do not change. The map area and the ecoInformation area (position 3 in the layout) work as a simulated system, without real data.

System description and use

Below, it is explained how the speed and gear recommendations were given to the driver.

Gear information presentation

While the participants were driving, the current gear that the driver has engaged from 1 to 6 is displayed in the screen (simultaneously with speed information in proposals A and B and intermittent with the speed information in proposal C).

When the revolutions are below 1800 rpm’s the system suggests to the driver a gear change to a lower gear by means of a face down green triangle blinking. Until the adequate gear position is reached, the triangle blinks. The system suggests a change in the gear where it should be engaged.

Figure 6. Gear down suggested

In the opposite side, driving in a high level of revolutions, above 3200 rpm’s, the system suggests to engage a higher position of gear. In this case, the green triangle blinking in the upper side of the current gear is displayed.

As in the previous case, until the adequate position of the gear is reached (taking into account the rpm) the triangle continues blinking.

Driving in the range of 1800 – 3200 rpms, the system does not show any recommendation of change, only the current gear.

Speed information presentation

The current speed is displayed in the centre of an oval, where different colours showed if the speed is the ‘eco’ speed (green section) or not.

The oval around the speed will illuminate a number of sections depending of the percentage value of the current speed with respect to the recommended value (always in the centre of green zone). Three different colours indicated if the speed is the recommended one (green section of the oval), not recommended at all (red section of the oval) and a transition zone between the not recommended speed and the recommended one (orange section of the oval).

Figure 7. Speed zones

System implementation in the Driving Simulator

For the implementation of the system in the driving simulator, a screen was placed on the up side of the central screen of the vehicle (right side of the driver), where the information was displayed. A dedicated PC controls the information displayed in this screen. This PC was connected to the driving simulator network and it received information from a specific software module (called 'HMI server'), developed specifically for the communication between the driving simulator software (SCANeR© II) and the HMI.

Subjective measures

For analyzing the participant's feedback (perceptions and opinions) about the HMI concept, five different questionnaires were used: general questionnaire, DALI (Pauzié & Pachiaudi, 1997), HMI General Evaluation, Pre-Trip and Post-Trip information, perceived HMI distractiction evaluation (Andreone et al., 2011).

Experimental Design

A within-subject experimental design was defined: each person drove three times and each time tested a different HMI proposal (A, B or C). Furthermore, it was necessary to control the order effect of the displays presentation using complete counterbalance. 6 groups of participants (4 persons in each group) were defined, depending on the order of presentation of the three displays. All of them had to drive through three different routes (one for each chosen display) into the same scenario.

Table 1. Configuration of groups

	Groups Screen Presentation Order			Nº participants
Group A	Proposal A	Proposal B	Proposal C	4 persons
Group B	Proposal A	Proposal C	Proposal B	4 persons
Group C	Proposal B	Proposal A	Proposal C	4 persons
Group D	Proposal B	Proposal C	Proposal A	4 persons
Group E	Proposal C	Proposal A	Proposal B	4 persons
Group F	Proposal C	Proposal B	Proposal A	4 persons
			Total:	24 persons

Experimental Procedure

Each participant drove a single session lasting approximately 1 hour, following the indications given to them. Before starting the session, general information was provided to them. The sessions were recorded (audio and video) in order to be used during the data analyses. The collection of information was done maintaining and protecting the privacy of participants.

Participants

The total sample was composed by 24 persons that were assigned randomly to the 6 experimental groups.The recruitment of them was done using the CTAG's database.The total sample is composed by 24 participants (M=27.29; SD= 3.483) with a range of age between 22 and 37 years old. Two thirds of the sample were male, meanwhile one third were female (Women = 8; Men = 16).

Results

In this section results from DALI questionnaire, evaluation of the graphical interfaces and distraction are presented. Furthermore, results from general questions about pre-trip and posttrip items are analyzed from a qualitative perspective.

DALI results

Five factors from the DALI questionnaire (Driving Activity Load Index), were analyzed for the three options presented to the participants: Global attention demand, visual demand, temporal demand, interference and stress. As it can be appreciated in Figure 13 the highest values are for the dimension 'Visual demand'. The median value for this variable is in a range between 2.46 for Option A and 3.00 for Option C.

Moreover, the option C presented the most negative results; in fact, the highest media values in all the factors evaluated are assigned to this option as it is shown in Table 2. The results of the Friedman Test indicated that there were a statistically significant different in all the factors except for the temporal demand.

A Wilcoxon Signed Rank Test indicated that there was a statistically significant difference between Option A and C regarding Global attention (Option A presents a lower value than Option C), $z=-1.992$, $p<.05$, with a medium effect size ($r=.30$) (Cohen, 1988). Regarding the Visual Demand, there was also a statistically significant difference between Option A and C ($z=-2.056$, $p<.05$), and Option B and C ($z=-2.044$, $p<.05$) in both the effect size was medium ($r=.30$). Finally, regarding the stress, a Wilcoxon Signed Rank Test revealed a statistically significant increase concerning option A (Md=1.00) and option B (Md=2.00), with $z=-2.285$, $p<.05$, with a medium effect size ($r=.33$).

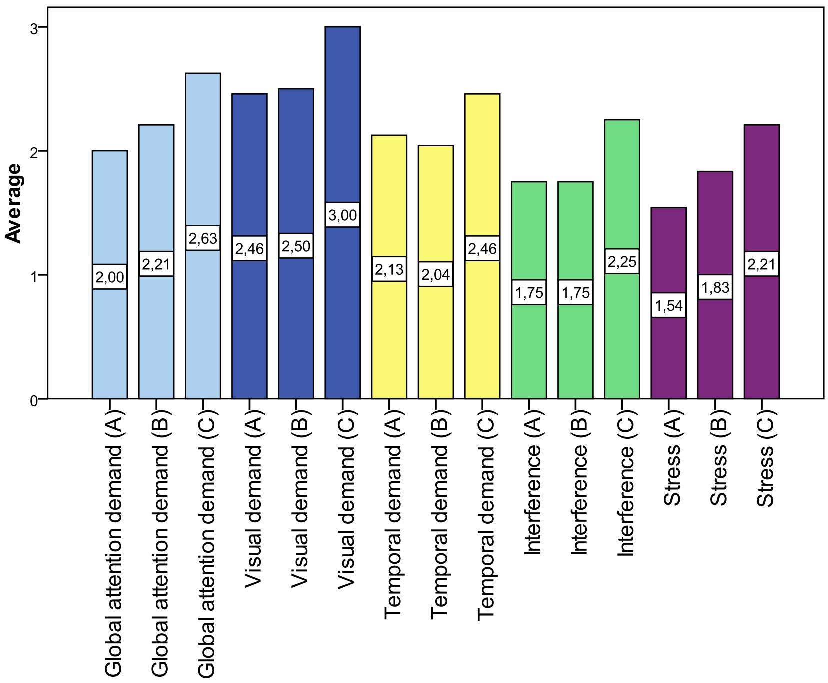


Figure 8. Average scores for the DALI factors

Table 2. Friedman Test Results (DALI questionnaire) (p < 0.05)*

Friedman Test	*Option A*		*Option B*		*Option C*	
Global attention	Median	*2.00*	Median	*2.00*	Median	*2.50*
			*χ(2,24)=7.015**			
Visual demand	Median	*2.00*	Median	*2.00*	Median	*3.00*
			*χ(2,24)=8.784**			
Temporal demand	Median	*1.00*	Median	*1.25*	Median	*2.00*
			χ(2,24)=3.205			
Interference	Median	*1.00*	Median	*1.50*	Median	*2.00*
			*χ(2,24)=6.127**			
Stress	Median	*1.00*	Median	*2.00*	Median	*2.00*
			*χ(2,24)=7.508**			

HMI evaluation results

For the different items included in this questionnaire the scores obtained are in general positive. The mean for all of them is over 5 points (except for degree of distraction and degree of driver annoyance, but in this case, values lower than 5 points are positive). The different values are better in option A and B than in option C. It seems that option C is the less pleasant option for the participants.

Friedman test results show statistically significant differences only for degree of utility, ease of learning and global value (see Table 3). The results of the Friedman Test indicated that there was a statistically significant difference in degree of utility scores for the three options (option A, B and C), $\chi(2,24)=8.955$, $p<.05$.

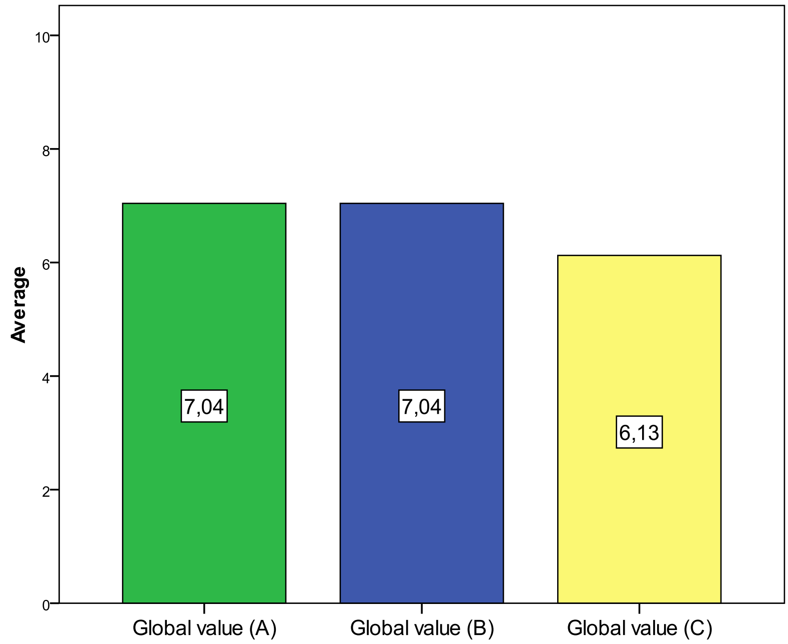


Figure 9. Average scores for global HMI evaluation

The analyses of the median values shown a decrease in degree of utility scores from option A and B (Md=7.00) to option C (Md=6.00). A Wilcoxon Signed Rank Test shown a statistically significant reduction in degree of utility scores concerning option A and C ($z=-2.275, p<0.05$. with a medium effect size ($r=.30$)), and concerning option B and C ($z=-2.265, p<0.05$. again with a medium effect size ($r=.30$)).

A trend similar to the one found in degree of utility can be observed in the ease of learning analyses: options A and B (Md=8.00) present the highest value (option C: Md=7.00) and the differences are statistically significant: $\chi(2,24)=7.524$, $p<.025$. A Wilcoxon Signed Rank test revealed a statistically significant reduction in easy of learning values concerning option A and option C, $z=-1.997$, $p<.05$, with a small effect ($r=.28$).

Table 3. Friedman Test Results (HMI Evaluation) (* p < 0.05)

Friedman Test	*Option A*		*Option B*		*Option C*	
Degree of utility	*Median*	*7.00*	*Median*	*7.00*	*Median*	*6.00*
			*χ(2,24)=8.9**			
Ease of learning	*Median*	*8.00*	*Median*	*8.00*	*Median*	*7.00*
			*χ(2,24)=7.52**			
Global value	*Median*	*7.00*	*Median*	*7.00*	*Median*	*6.00*
			*χ(2,24)=16.44**			

Finally, the results of the Friedman Test for the Global value were significantly different, the median values showed a decrease ($\chi(2,24)=16.444$, $p<.001$) in Global value scores from option A and B (Md=7.00) to option C (Md=6.00). Once again, a Wilcoxon Signed Rank Test indicated a statistically significant reduction in degree of utility scores between option A and C, $z=-2.811, p\leq 0.005$ with a medium effect size ($r=.40$), and between option B and C, $z=-3.581$, <0.000 again with a large effect size ($r=.51$).

Distraction evaluation

Concerning distraction evaluation, option C presented the highest value (M=5.22). Option A and B showed similar mean scores (Mean for A=4.46 and for B=4.58). But statistical differences among the three options were not found. These results are in accordance with the data obtained in HMI results about degree of distraction for the three proposals.

Duration of information in the screen for option C

Regarding the estimation of time the information is in the screen for option C, the average value was 4.04 seconds (SD=2.56). 66.6% of the participants undervalue the time giving a value under 5 seconds (the actual time the information is maintained in the screen). Only 16.7% estimates it correctly.

Focusing on the comments from the participants, part of the sample considered the duration to be adequate: the time is considered enough to see the information without being distracted. In addition, they suggested that if the duration is lower, this situation would generate stress and, on the contrary, if the time were higher, they thought they would be worried about the change (waiting it) and the number of times they'd look to the screen would be increased.

Those who estimated that the duration should increase (25% of the sample), claimed that it should increase with 1, 2, 5, 7 or 10 seconds, and those who thought the duration should decrease (12.5% of the sample) suggested to decrease by 1, 3 or 5 seconds

Some participants commented that the time the information is presented in the screen should depend on the characteristics of the road and it should be presented in the screen while the recommendation is applying to the driving situation.

Preference of screens

The option A was the first choice for 54.17%, while 41.67% of the participants chose option B. For the second choice, the percentages were similar: 45.83% chose option A as second option and 50% chose option B. Option C was the least selected option for first and second choice, option C was the most preferred as third choice.

The participants who choose option A commented that for them the most appreciated characteristic of this option is the position the gear indicator on the top because it is easier to see it, it is easier to understand and it is more intuitive.

Regarding to option B, the participants who selected it justified their selection because they found it easy to understand the information, easy to see the gear information and the position is more intuitive, also the position of gear information is associated with the position of gear shift in the vehicle.

Not all participants valued option B positively, mainly because they think that it is complicated to see the gear in this position (below the speed information).

Option C had the most negative values. The participants commented that they did not like the alternation of the information in the screen because this change distracts them. Those who liked this option commented that, as the information is not displayed simultaneously, when they look at the display they need only to understand one advice and not both of them at the same time.

Qualitative results

Most of the participants that made a comment suggested that the indicator (the arrow) that recommended changing the gear and the area that shows the eco-driving should be more visible. Furthermore, an important part of the sample thought that it would be recommendable to show the speed and gear indicators together only in those cases in which that would coincide.

Some subjects thought that the amount of information displayed in the screen is quite high and the presentation mode causes distraction. As an alternative, they suggested to use sound for some cases or to reduce the arc of colour (it involves a lot of workload).

Regarding the oval shape and the arc of colours designed for the speed recommendations, some participants considered it complicated to make a quick association between colours and eco driving. They judged that the use of colour levels in both sides is not needed. Nevertheless, other part of the sample believed that the arc of colours is very intuitive and comfortable.

Finally, many of participants would place the display in the dashboard and some subjects considered the system to be useful, but dispensable.

Pre-trip

The kind of information that drivers would like to provide at the begin of the trip is the following destination, driving efficiently (mode), consumption, load, route, meteorology, selection of efficient road (mode), etc.

Moreover, participants answered (using a 0 to 100 points scale) to the question: in which degree do you like the system will help you? The mean value of the answers was of 24.17 points with a standard deviation of 31.31. The scores given by 30% of the sample were below 50 points. The other part of the participants provided higher values (25% of them gave 100 points).

Post-trip

Participants were also asked about what information they would like the system to provide at the end of the trip. The information that participants demanded by to end the trip was the following: Fuel Consumption and Average consumption, time eco-driving, percentage of fuel consumption, difference between eco-driving and real driving, percentage of time the gear position is correct, time eco-driving, driving cost, time eco-driving, a graph with the sections in which the fuel consumption was higher, etc.

Second option

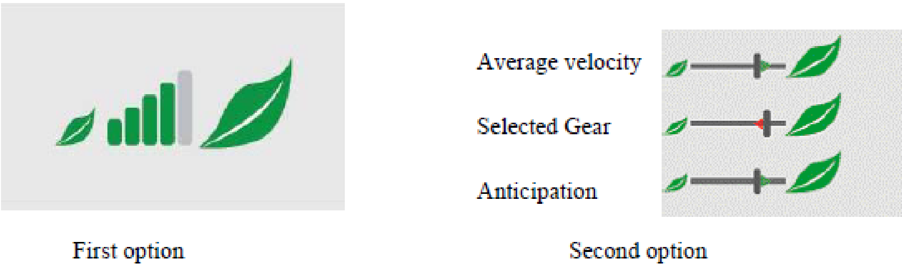


Figure 10. Proposal of presentation of information

Summarizing, the general result is that the participants preferred the first option as a default mode and the second option as an advanced mode in which more information is displayed.

Discussion

Added to the general evaluation of the graphical interface, the study has focused on analyzing the best way to give the gear and speed recommendations. Taking as a reference other studies in which the design and evaluation of different interfaces was done using simulated driving situations (Chalmé et al. 1999), an experiment in a dynamic driving simulator was carried out in order to do the subjective evaluation of three different graphic modes designed for presentation of the on-trip information.

The workload was evaluated using the DALI questionnaire (excluding only the auditory demand item not needed for this evaluation). This questionnaire is broadly used for evaluating workload in drivers (Pauzié & Pachiaudi, 1997). In this study, the visual demand plays an important role (presentation of information only in visual mode) so it has sense the highest values of the DALI were assigned to this factor. In any case, the presentation of the information does not annoy to drive. In this sense, it was important to take into account the use of visual information to maintain an ecological driving. Fricke (2007) found that "in general, visual information presentation is best where the presented information need to be persistent because one cannot immediately respond to it".

When the results of the DALI questionnaire were compared for the three options, option A presents a lower value for Global attention demand than option C (significantly different with medium effect size). Focusing in the Visual demand, there was also a statistically significant difference between Option A and C and between Option B and C (lowest values for A and B). Regarding the stress, a statistically significant increase was appreciated concerning option A and B. Looking at these results, it seems that option A presents better results in terms of Global attention, Visual Demand and Stress and option C present the worst ones.

The differences between options A, B and C are also reflected in the answers to the questionnaires. The differences between option A and B are minimum, but there is a clear difference with option C: drivers preferred the combination of the information instead of the presentation in an alternative way. A change in the presentation mode was irritating for them in general (Fricke, 2007). Then, when information is presented alternatively (Option C) the drivers do not feel comfortable. In fact, the participants agreed that they did not like changeable information and this fact caused distraction.

Moreover, participants expressed that the HMI evaluation was positive for all screens. Once more, the alternative presentation of information has the worst values (statistical significant differences were found in 'degree of usefulness', 'easy of use' and 'global score'). In general, the results are supporting the conclusion about the preference of the drivers for the presentation of gear and speed information combined at the same time.

On the other side, no significant results were found regarding the distraction, but, once again, option C presents the highest values.

Regarding the duration of information, after this study it is not clear that the subjects are good in estimating the duration of information. These analyses should be taken into account in further studies.

The analyses of the qualitative results show that, in general, the drivers appreciated the help offered by an eco-driving support system and they were willing to provide clues to improve it. Focusing on the post-trip application, when only a general evaluation of the driving is given, the drivers considered this as a default mode meanwhile when more descriptive information is provided, they perceive that this mode should be an advanced option. This is related to the fact that "several new approaches to assisting the driver in dealing with the complex task of the driving situation are being develop with the aim to give support in the secondary task and at the same time facilitate the primary driving task" (Fricke, 2007). From this study it could be understood that the fact of providing aid to green driving is a secondary task that would help drivers in their primary driving task.

References

Andreone, L., Bernatallada, G., Bersiner, L., Blanco, R.,Brockmann, M., Damiani, S., García, E., Gröning, A., Kassner, A., Kuck, D., Lübbert, K., Rommerskirchen, C., Schebitz, N., Schießl, C., Staubach, M., Themann, P., & Uhrner, T. (2011). ecoSmartDriving ecoHMI. Project ecoMove Deliverable D3.6.

Fricke, N. (2007). Effects of adaptive information presentation. In Proceedings of the Fourth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design (pp. 292-298). Stevenson, Washington, USA.

Pauzié, A., & Pachiaudi, G. (1997): Subjective Evaluation of the Mental Workload. In T. Rothengatter & E. Carbonell Vaya (Eds.), *Driving Context. Traffic and Transport Psycology: Theory and Application*. (pp. 173-182). Bingley, UK: Pergamon.

Pauzié, A. & Manzano, J. (2007). Evaluation of driver mental workload facing new in-vehicle information and communication technology. 20th Enhanced Safety of Vehicles Conference (ESV), Lyon, France, 17-21 June 2007, on CD-ROM, Paper number 07-0057, 10 p.

Pauzié, A., Manzano, J., & Dapzol, N. (2007). Driver's Behavior and Workload Assessment for New In-Vehicle Technologies Design. In Proceedings of the 4th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design, July 9-12, 2007. (pp.572-580) Stevenson, Washington, USA.

Pauzié, A. & Gelau, C. (2010). Methodology for safety and usability assessment of ITS for riders. *In Proceedings of the European Conference on Human Centred Design for Intelligent Transport Systems* (pp.457-465), Berlin, Germany: HUMANIST.

Van der Laan, J.D., Heino, A., & de Waard, D. (1997). A simple procedure fort he assessment of acceptance of advanced transport telematics. *Transportation Research Part C, 5*.,1-10.

Shifting from manual to automatic gear when growing old: good advice? Results from a driving simulator study

Dafne Piersma & Dick de Waard
University of Groningen, Traffic Psychology Group, The Netherlands

Abstract

Older people may be advised to switch from manual to automatic gear shifting, because they may have difficulties with dividing their attention between gear shifting and other driving tasks such as perceiving other traffic participants. The question is whether older drivers show a better driving performance when using automatic gear shifting instead of manual gear shifting. Twenty young and twenty older drivers participated in a driving simulator study. Each participant drove both in the manual and the automatic gear conditions. Young drivers reported significantly higher risk taking behaviour in traffic than the older drivers. The older drivers experienced more collisions in the driving simulator than the young participants, but this was only true in the manual gear condition. Older participants swerved more than young participants in both gear conditions. Altogether, the driving performance of the older drivers was better in the automatic gear condition compared with the manual gear condition. This research supports the advice for older drivers to use automatic gear shifting.

Introduction

For many people, driving is very important and their preferred mode of transport (Gruber et al., 2013). Car drivers feel they are independent and have optimal mobility options. As a result, most aged drivers continue to drive (Unsworth et al., 2007).

There is no consensus about who drive safer, young or older drivers. Actually, both young and old drivers are overrepresented in crashes (Li et al., 2003). Young drivers may show a bad driving performance due to inexperience (Isler et al., 2011) and/or increased risk taking (O'Brien & Gormley, 2013; McGwin & Brown, 1999) while older drivers may suffer from age-related changes which could compromise driving performance (Brouwer & Ponds, 1994). In older drivers, both cognitive resources (Verwey, 2000; Withaar et al., 2000) as well as their motor skills may decline (Wheatley & Di Stefano, 2008). This may explain the overrepresentation of older drivers in crashes in complex situations (McGwin & Brown, 1999). Complex situations include crossing intersections, merging and overtaking. As driving is a multifaceted task different kinds of errors can be made. Any driving behaviours that

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

increase the risk of collision may be considered driving errors. Drivers show good driving performance when the number of driving errors is low, which corresponds to relatively safe driving.

Michon's model defines three levels of driving behaviour; strategic, tactical, and operational (Michon, 1985). On the strategic level planning takes place such as determining the route. On the tactical level perceiving and reacting to the environment is crucial such as keeping enough distance to others drivers. The operational level includes steering, operating the accelerator, brake, and clutch and other vehicle controls. The three levels may be active simultaneously and might influence one another. This implies, for example, that an operational level error could negatively affect driving performance on the strategic and tactical levels.

Research comparing driving performance of young and older drivers showed contradictory results (McGwin & Brown, 1999). Some researchers report that young drivers make more driving errors than older drivers (Carr et al., 1992; Shinar et al., 1998). Others demonstrate that older drivers have a worse driving performance compared with young drivers (Chaparro et al., 2005; Selander et al., 2011). Perryman and Fitten (1996) have shown that ageing may decrease operational driving skills such as maintaining a proper speed and lane position. Gear changing is another operational driving skill. Novice drivers may well find this hard and practice is necessary to learn to automatically change gears (Shinar et al., 1998). While middle-aged drivers may change gears automatically, in older drivers execution of movements may become less automatized again (Brouwer & Ponds, 1994). In the latter case, gear shifting requires attention again, thus demanding cognitive resources which are less available in older drivers. Consequently, older drivers might have difficulties with dividing their attention between gear shifting and other driving tasks such as perceiving other traffic participants (McGwin & Brown, 1999). Selander et al. (2011) reported incorrect gear shifting as fourth most common driving error in older drivers.

Older drivers may be advised to switch from manual to automatic gear shifting, but so far limited research has been performed on the role of transmission type on driving safety. Warshawsky-Livne and Shinar (2002) recorded brake reaction and movement times in a driving simulator during manual and automatic gear conditions. Reaction times increased with age, but also movement times were longer during manual compared with automatic gear shifting. The study implicates that automatic gear shifting might improve driving performance. Shinar et al. (1998) investigated sign detection by young novice drivers and experienced drivers during manual and automatic gear shifted driving. Novice drivers missed more signs in the manual gear condition than in the automatic gear condition. This difference was not found in the experienced driver group indicating that young drivers could benefit from automatic transmission. Selander et al. (2012) compared the number of driving errors made by young and older drivers during manual and automatic gear shifted driving. Here, participants in the older group showed more driving errors in both transmission conditions than their younger counterparts, but automatic transmission reduced the number of errors of the older drivers. This study supports automatic transmission for older drivers. So far, this study has not been repeated, nor has it been performed in a driving simulator. The aim of the current study is to investigate the driving performance of young and older drivers during manual and automatic gear shifting. Swerving will be measured and the number of collisions will be counted. The hypothesis is that older drivers show more driving errors than young drivers. Also, the expectation is that participants generally will make fewer driving errors in the automatic transmission condition than in the manual transmission condition and that the performance of older drivers improves most.

Materials and methods

Participants

Overall, forty participants were recruited through distribution of flyers, advertisements and the word-of-mouth. All participants were native Dutch speakers in possession of a valid driver's license for at least one year. Twenty young and twenty older participants entered the study. One older participant was excluded, because the person did not possess a valid driver's licence. Another three older participants were excluded due to simulator sickness. Twenty young (M=22.7 ± 1.6) drivers completed the study, ten males and ten females. Sixteen older (M=73.4 ± 4.8) drivers completed the study, eight males and eight females.

Apparatus

A fixed-based driving simulator located at the University of Groningen was used for the study. The simulator consisted of an open cabin mock-up with a steering wheel, gear box, gas pedal, brake pedal, clutch and simulated driving sound. Four LED screens, three in front and one on the left of the mock-up provided the participant with a view on the road. Each screen provided a 70° view, leading to a view of 280° in total. The car windows, side mirrors and rear view mirror were realized on the screen. During driving the participants wore the safety belt. Three computers were used, one for graphical rendering, one for the traffic simulation and one for system control showing a user interface for the simulator operator. The graphical interface was designed with StRoadDesign (StSoftware) and the scenario was programmed with scripting language StScenario (StSoftware).

Procedure

Every participant drove three rides in both the manual and the automatic gear condition. Half of the young and half of the older participants started with the manual gear condition and the other half started with the automatic gear condition. In both gear conditions the same three rides were performed. The first ride was in a rural environment on a slightly winding road, the participants had to maintain their lane position on the right lane while oncoming traffic approached on the left lane. During the first part of the ride the participant chose a comfortable speed, in the second part the participant was asked to hurry and drive as fast as was just within safety margins. Participants were told there was no speed limit. The second ride was in a rural environment as well, but now the participant encountered six intersections with different priority regulations, including one with traffic lights, and also a car that suddenly pulled out of a parking lot in front of the participant. Speed limits differed between 60, 70 and 80 km/h. The participant was told to obey the traffic rules. In the third ride the participant merged into a crowded motorway, was then asked to pass one vehicle and subsequently to leave the motorway. After the participant had driven all three rides in both gear conditions a short interview took place. The participants were asked to report their age and to reflect on their own driving competence and risk taking behaviour in traffic. Driving competence was indicated on a 10 point Likert scale running from 1 to 10 (10 being extremely competent), and risk taking behaviour was indicated on a 5 point Likert scale running from 1 to 5 (5 being very high risk taking).

Analyses

The subjective scores on driving competence and risk taking in traffic were compared between young and older drivers using t-tests. Swerving was measured as the standard deviation of the lateral position (SDLP) in cm during the first ride in both gear conditions. A repeated measures ANOVA was used to investigate differences in swerving between the manual and automatic gear conditions and between young and older participants. During the second ride in both gear conditions the number of collisions was counted. Chi-square tests were used to analyse differences in the number of collisions between both gear conditions and between the two age groups.

Results

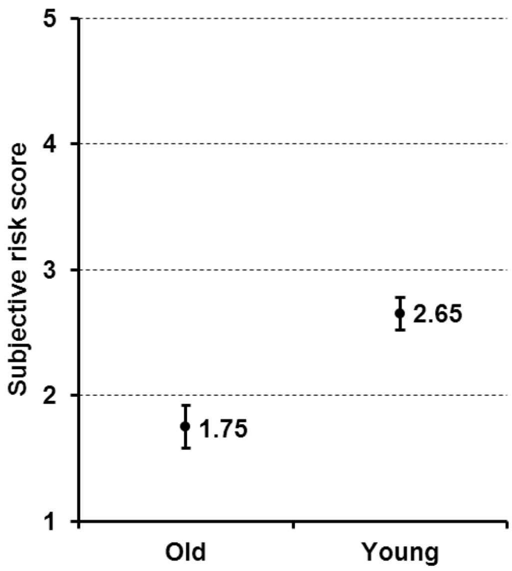


Figure 1. The subjective risk taking scores of the old and young subjects. Error bars represent the standard errors of the means.

The subjective score on driving competence is comparable in the young (M=7.30 ± 1.08) and elderly (M=7.17 ± 0.72) groups, t(34)=-0.417, NS. However, the subjective risk taking scores of older drivers (M=1.75 ± 0.68) were significantly lower compared with younger drivers (M=2.65 ± 0.59), t(34)=-4.250, p<.001 (Figure 1).

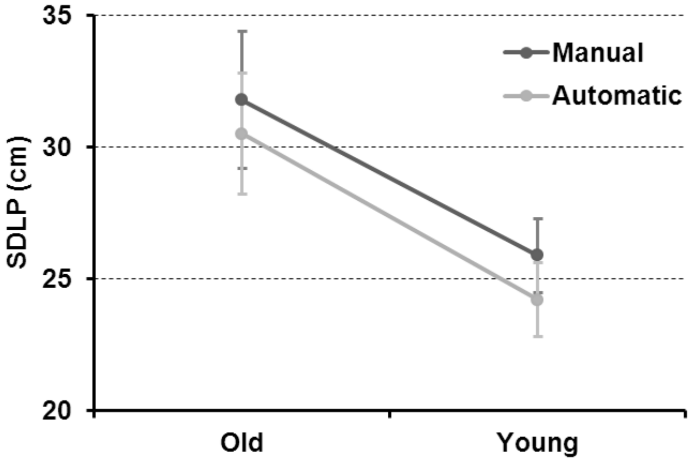


Figure 2. Swerving by old and young subjects in the manual and automatic gear conditions. SDLP = standard deviation of the lateral position. Error bars represent the standard errors of the means.

There was no significant difference in swerving between the manual and automatic condition, F(1,35)=1.08, NS. Yet, in both conditions the older participants swerved more than the young participants, F(1,35)=9.64, p=.004 (Figure 2).

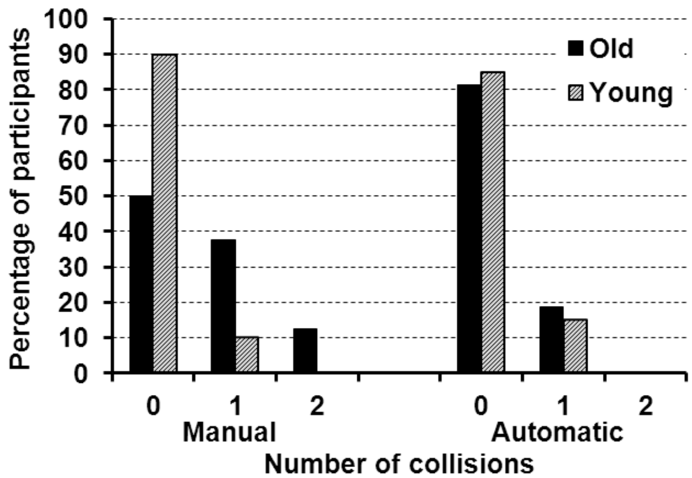


Figure 3. The percentage of older and young participants involved in 0, 1 or 2 collisions in the manual and automatic gear conditions.

Overall, there was no significant difference in the number of collisions between the manual and the automatic gear condition, χ^2(df=2, N=36)=0.62, NS. Moreover, in the automatic gear condition the young and older drivers did not differ in the number of collisions, χ^2(df=1, N=36)=0.09, NS. However, in the manual gear condition the older drivers were involved in significantly more collisions than the young drivers, χ^2(df=2, N=36)=7.49, p=.024 (Figure 3).

Discussion

Older drivers reported to take less risk in traffic than young drivers. Yet, older drivers swerved more than younger drivers in both transmission conditions. In addition, older drivers were involved in more collisions than young drivers in the manual gear condition. In the automatic gear condition this difference was not found.

The finding that older drivers take less risks in traffic compared with young drivers is consistent with previous studies (O'Brien & Gormley, 2013; McGwin & Brown, 1999). Nevertheless, the driving performance of the older drivers was worse compared with the young drivers. This may not be caused by deliberate risk taking, but by age-related changes. Older drivers swerved more than young drivers in the driving simulator. Although, the mean SDLPs were higher in the manual compared with the automatic conditions in both age groups, the variation was large and the difference between the transmission conditions not significant. Selander et al. (2012) also investigated lane-keeping control with automatic gear cars compared to manual gear cars. However, they reported older drivers to drive significantly less frequently too far to the left in an automatic car compared with a manual car. When driving too far to the left, the middle line of the road could be crossed which gives a higher risk of collision with approaching traffic. In the current study the number of middle line crossings was counted in the first and second rides in both gear conditions. In the first ride, the number of middle line crossings was significantly higher in the manual compared with the automatic gear condition, χ^2(df=16, N=36)=32.00, p=.010. This finding was, however, caused by a minority of all participants, since 22 of the 36 participants did not cross the middle line at all in the first rides. The second ride was more difficult and required more gear shifts in the manual gear condition due to the intersections. In the second rides only six out of all participants did not cross the middle line at all. Again the number of middle line crossings was higher in the manual compared with the automatic gear condition, χ^2(df=63, N=36)=99.733, p=.002. In the first ride, the difference in number of middle line crossings between the older and younger drivers approached significance, but was neither significant in the automatic gear condition, χ^2(df=4, N=36)=9.00, p=.061, nor in the manual gear condition, χ^2(df=4, N=36)=8.49, p=.075. In the second ride, the same was found in the automatic condition, χ^2(df=7, N=36)=12.11, p=.097. On the contrary, in the second ride of the manual condition older drivers did show significantly more middle line crossings than their younger counterparts, χ^2(df=9, N=36)=22.28, p=.008. Automatic transmission may thus improve lane-keeping control of, especially older, drivers.

Older drivers were involved in more collisions than younger drivers in the manual gear condition. The higher number of collisions in the older driver group may be explained by the type of collisions that could occur. Collisions were either the result of braking too late for a car that pulled out or caused by crashing into other cars at intersections. In the first case, brake reaction and movement times are crucial. Warshawsky-Livne and Shinar (2002) already reported that older drivers have slower reaction times and with manual gear shifting longer movement times arise. Together these results may well lead to an increased risk of collision as described in the first case when an older driver uses manual gear shifting. With regard to the second case, an explanation might be that older drivers have an increased risk of multiple-vehicle intersection crashes compared with young people (IIHS, 2007 & 2011). With an automatic gear older drivers no longer caused more collisions than young drivers. Young drivers were involved in as few collisions in the manual as the automatic condition. This result indicates that the older drivers were benefitting from the automatic gear shifting.

The practical implication of this research is that older drivers should be advised to use automatic gear shifting instead of manual gear shifting. To do this, one needs to know *when* a driver becomes an older driver. In the current study older participants were 65 years old or older. In the study of Selander et al. (2012) older drivers were aged between 70 and 90 years. In the study of Warshawsky-Livne and Shinar (2002) the senior group was between 50 and 82 years of age and with a mean of 62 years clearly younger than previously mentioned groups. Over-involvement in crashes is most prominent for drivers above 75 years of age (IHHS, 2007; Li et al., 2003). Therefore, a switch from manual to automatic gear shifting before the age of 75 may be advised. More research is needed to define at which age it is the best moment to switch to automatic gear. For people who have been driving cars with manual transmission all their lives it might take some time to get used to automatic gear shifting. Especially older people may experience difficulties with learning new skills. For this reason, people could be advised to acquire driving experience in an automatic gear car already at a younger age too.

Over-involvement in crashes is not only reported for older drivers, but also for drivers below 20 years of age. In all mentioned studies no groups of drivers all aged below 20 were included. A study with this group could show whether the very young drivers would benefit from automatic gear shifting too. This might be expected because Shinar et al. (1998) found that novice drivers missed more signs in the manual gear condition than the automatic gear condition.

It is of importance to investigate further for which age groups automatic gear shifting is preferred over manual gear shifting. Moreover, a longitudinal on-road study with young and old subjects may provide more insight in the number and type of collisions drivers cause in both manual and automatic transmission cars.

Conclusion

There appears to be benefit of driving a car with automatic gear for the older driver. Older drivers reported lower risk taking scores than young drivers, but showed a worse driving performance in the driving simulator. The older drivers swerved more in both transmission conditions than young drivers. Only in the manual condition the older drivers were involved in more collisions than the younger drives. Altogether, this study supports the advice for older drivers to switch from manual to automatic gear shifting.

Acknowledgments

Thanks to Femke Scholten, Florentine Roerink, Loes Ligtenberg and Alexander Meijer for data collection.

References

Brouwer, W.H., & Ponds, R.W.H.M. (1994). Driving competence in older persons. *Disability and Rehabilitation, 16,* 149-161.

Carr, D. Jackson, T.W., Madden, D.J., & Cohen, H.J. (1992). The effect of age on driving skills. *Journal of the American Geriatrics Society, 40,* 567-573.

Chaparro, A., Wood, J.M., & Carberry, T. (2005). Effects of age and auditory and visual dual tasks on closed-road driving performance. *Optometry and Vision Science, 82,* 747-754.

Gruber, N., Mosimann, U.P., Müri, R.M., & Nef, T. (2013). Vision and night driving abilities of elderly drivers. *Traffic Injury Prevention, 14,* 477-485.

Isler, R.B., Starkey, N.J., & Sheppard, P. (2011). Effects of higher-order driving skill training on young, inexperienced drivers' on-road driving performance. *Accident Analysis & Prevention, 43*, 1818-1827.

Li, G., Braver, E.R., & Chen, L.H. (2003). Fragility versus excessive crash involvement as determinants of high death rates per vehicle-mile of travel among older drivers. *Accident Analysis & Prevention, 35,* 227-235.

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

Michon, J.A. (1985). A critical view of driver behaviour models: what do we know, what should we do. In L. Evans, and R.C. Laoschwing (Eds.), *Human behaviour and traffic safety* (pp. 485–520). New York, Plenum Press.

Insurance Institute for Highway Safety (IIHS). (2013). Fatality facts 2011: Older people. <http://www.iihs.org/research/fatality.aspx?topicName=Older-people>. Accessed on 2013 Aug 8.

Insurance Institute for Highway Safety (IIHS). (2008). Fatality facts 2007: Older people. <http://www.iihs.org/research/fatality.aspx?topicName=Older-people&year=2007>. Accessed August 8, 2013.

O'Brien, F., & Gormley, M. (2013). The contribution of inhibitory deficits to dangerous driving among young people. *Accident Analysis & Prevention, 51,* 238-242.

Perryman, K.M., & Fitten, L.J. (1996). Effects of normal aging on the performance of motor-vehicle operational skills. *Journal of geriatric psychiatry and neurology, 9,* 136-141.

Selander, H., Lee, H.C., Johansson, K., & Falkmer, T. (2011). Older drivers: On-road and off-road test results. *Accident Analysis & Prevention, 43*, 1348-1354.

Selander, H., Bolin, I., & Falkmer, T. (2012). Does automatic transmission improve driving behavior in older drivers? *Gerontology, 58*, 181-187.

Shinar, D., Meir, M., & Ben-Shoham, I. (1998). How automatic is manual gear shifting? *Human Factors, 40,* 647-654.

Unsworth, C.A., Wells, Y., Browning, C., Thomas, S.A., & Kendig, H. (2007). To continua, modify or relinquish driving: findings from a longitudinal study of healthy ageing. *Gerontology, 53,* 423-431.

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

Warshawsky-Livne, L., & Shinar, D. (2002). Effects of uncertainty, transmission type, driver age and gender on brake reaction and movement time. *Journal of safety research, 33,* 117-128.

Wheatley, C.J., & Di Stefano, M. (2008). Individualized assessment of driving fitness for older individuals with health, disability, and age-related concerns. *Traffic Injury Prevention, 9,* 320-327.

Withaar, F.K., Brouwer, W.H., & Van Zomeren, A.H. (2000). Fitness to drive in older drivers with cognitive impairment. *Journal of the International Neuropsychological Society, 6,* 480-490.

Domain independent visual attention assessment in stereoscopic displays

Franziska Michalczik [1], Manfred Thüring [2], Axel Hillebrand [1],
[1]EADS Innovation Works, [2] TU Berlin
Germany

Abstract

Complex operational scenarios often involve three-dimensional (3D) mental models, yet operators are provided with two-dimensional (2D) visual input. Considering that 3D displays and applications are on the rise, it is of special interest whether these developments have negative effects on visual attention deployment (VAD) across various application domains, thus decreasing assumed benefits.
Objectives of the study were to establish an assessment method for VAD in stereoscopic displays. A similar approach from investigations of two-dimensional VAD built the foundation for the current study (Hillebrand et al., 2012). It was required that the methodology for 3D investigations assessed VAD without using expert knowledge, in order to gain domain-independent insight; and to check for time-dependent influences within the task design.
Overt and covert attention was examined using a task combination of the Stroop test and a target detection task. The Stroop test captured visual attention in a confined area, hereby simulating other tasks that usually demand expert knowledge. The target detection task was performed simultaneously, the detection rate served as measure for VAD.
Results show that the task combination is applicable for examining VAD in 3D domain independently. The location of the observer's primary attentional focus influences VAD, especially the direct surroundings of the Stroop test. A bias for upper display areas and support for the attentional gradient theory (Atchley et al., 1997) were found. Time of target appearance influenced target detection, depending on the primary focus.

Introduction

In many technical domains, such as air traffic control, operators must accomplish multiple tasks simultaneously and construct mental models from a variety of information sources. To accomplish this, they have to transfer two-dimensional (2D) visual input from displays into a three-dimensional (3D) mental model of the area they are controlling. At first glance, it appears plausible to facilitate the construction of such a model by simply displaying information on three dimensions in stereoscopic displays. However, it is uncertain how visual attention (VA) is deployed in this case, since findings in natural depth perception cannot be readily

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

transferred to understand how perception works in 3D displays. To clarify this issue, an experiment was performed using a modified version of a methodology called ATTENDO (Hillebrand et al., 2012). ATTENDO is a two-task paradigm which has previously been used to evaluate visual attention allocation for a head-up display scenario in a PC-based flight simulator. For the current purposes, ATTENDO was adapted to 3D settings. This new version enables researchers to investigate VA in stereoscopic displays for any application domain and independently from the knowledge specificities of different experts.

The theoretical basis for this study is derived from research on visual perception and attention. With respect to VA, a distinction between *overt* and *covert VA* is often made. Overt attention is described as attending that requires eye-movements, while covert attention does not (Hunt, 2003; Wright & Ward, 2008). Wolfe and Horowitz (2004), as well as Kollmorgen et al. (2010) have summarized a large part of the available research on the question what guides the deployment of VA. Since they gained insight on general VAD, this insight is important for 3D research, as well. Thus, their notions on low-level and high-level aspects of visual stimuli were accounted for in the design of the visual input for the experiment described in the next section. This input consisted of stimuli with low-level as well as high-level features. While low-level features provide visual salience, high-level features guide VA by the meaning of their content. These features were accounted for in design of the visual stimuli of the experiment described in the next section.

Another aspect of visual attention deployment (VAD) is addressed by the theory of the *attentional gradient* (Atchley et al., 1997). According to this theory, “attention functions like a [funnel-shaped] gradient with maximal processing efficiency at the centre and declining efficiency at more peripheral locations” (p. 524). With respect to depth, there is “a viewer-centred attentional gradient in which attention is allocated from the observer to the point of fixation” (p. 527). This means that visual attention can be directed into space although studies often fail to verify this. Atchley et al. (1997) assume that the reason for this failure is related to perceptual load. If observers are confronted with visual scenery of low perceptual load, depth-awareness is not assessable. Therefore, VAD might not depend on low- and high-level features alone, but might also be influenced by perceptual load. To keep the design of the experiment focused, neither perceptual load nor the aspect of low-level and high-level features mentioned above were treated as independent variables in this study.

Method

To investigate how VA is deployed in stereoscopic displays, a task is needed which requires the monitoring of the complete display space. This is a necessary precondition for finding areas to which attention is paid more (or less) frequently compared to others (Hillebrand et al., 2012). Moreover, the task should account for the multitasking characteristics which are typical for the work of operators. Since they usually must cope with more than one task at a time, operators may focus on an area which they regard as particular important for a current task thus neglecting other areas. In addition to spatial aspects, it is also important to consider at which time task relevant information is presented.

Participants

Twenty four persons (21 male, three female) participated in the experiment. Their age ranged from 20 to 57 years (mean of 33.1 years, SD=9.16) and all of them were native German speakers. They received no payment or any incentives for their participation.

Apparatus and stimuli

A custom-made 3D workbench with a two-projector system and rear projection panels was used for the experiment. Stereo parallax was enabled by using passive polarization glasses; movement parallax was accounted for by head tracking. The measures of the display space were 65x40x70 cm (length x height x depth). The visual angle of the workbench was 22° - 30°. It was not homogeneous for the complete display, since the two projector panels were sloped (see Figure 1).

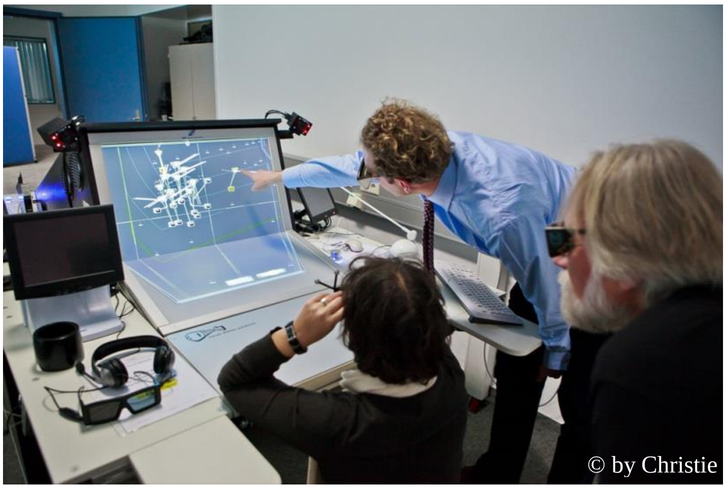


Figure 1. A 3D workbench from the Deutsche Flugsicherung GmbH (source: Christie Digital Systems,Inc.) similar to the one used in this experiment.

The experiment was based on a two-task paradigm consisting of a modified Stroop test (Michalczik et al., 2013) and a target detection task (Hillebrand et al., 2012) adapted to 3D purposes. The Stroop test stimuli were designed to provoke word-colour interference, using the four colours and colour names: red, yellow, blue, and green. The words were displayed in German. The Stroop test presented stimuli (incongruently coloured colour names) separately, unlike the original Stroop experiment, in accordance to presentation structure rules by Dalrymple-Alford and Budayr (1966). Three blocks of the Stroop test were performed. There were 100 incongruent Stroop stimuli per block. Further, eight congruent were introduced per experimental block as control stimuli. The congruent stimuli proofed that participants actually read the Stroop words, instead of naming colours alone. The Stroop test served as primary task and aimed to draw VA to a specific region of the display. The Stroop stimuli were presented serially with an inter stimulus interval of 1 second. Each had a duration of 0.5 seconds and appeared at one of three locations

(S1, S2 and S3, see Figure 2). S1 was located in the upper-right background, S2 in the very centre of every dimension and S3 in the lower-left foreground. The viewing direction of the operator's gaze is slightly from above, as illustrated by the arrow in Figure 2.

The secondary task was a target detection task. It served to assess the distribution of VA over the complete 3D space. There were 27 possible positions for the targets which were coloured dark grey, just above the perception threshold for distinguishing them from the background

The colour of the background was 85% grey with an embedded 75% grey grid. The grid was established for enhancing the impression of depth of the screen content. However, the grid is not shown in Figure 2 since it is a schematic depiction of the scenery that serves illustrational purposes only. To enable a better understanding of the visual content the colours in Figure 2 have been altered, as well.

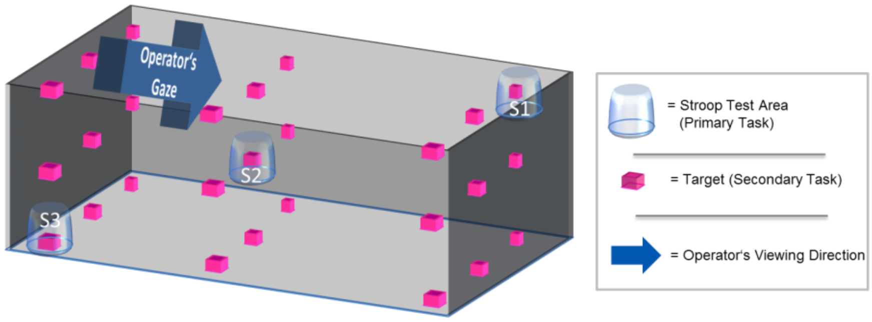


Figure 2. Distribution of primary and secondary task within the 3D display space (Note: targets were of dark-grey colour, use of magenta is only for illustrational purposes.).

Procedure

Before starting the experiment, participants took two vision tests. The first test assessed colour vision impairment, the second one examined binocular visual faculty. Having passed the vision tests, participants were introduced to the workbench. They were placed in front of the horizontal centre of the screen so that their eye level at 50 cm height from the desk.

The experimental phase was preceded by a short training. Since the Stroop test generally shows a relatively stable level of performance after the processing of five blocks (MacLeod, 1991), these were performed first. The training blocks contained 30 Stroop stimuli and 8 targets. The subsequent experimental phase consisted of three blocks of Stroop tasks. Each block was dedicated to one of the three positions of the task (S1, S2, S3). All blocks were presented to each participant in randomized order. Each block consisted of 108 Stroop stimuli, 27 of them together with a target, 81 without a target (note that Figure 2 shows all possible 27 positions, although targets were presented in succession). Within a block, each of the 27 target positions was used once and their order was randomized for every participant. In total, 324 Stroop stimuli and 81 targets were presented in each experimental session. Each target was presented for 500 milliseconds. It appeared either synchronously with the Stroop stimulus (no delay) or asynchronously with a delay of 0.5 seconds or 1.0 seconds. To indicate the detection of a target, participants pressed a button on a nine-key keyboard left of the screen.

Experimental design and hypotheses

The experiment investigated three independent variables: Stroop test position (S1, S2, S3; see Figure 3), target position (27 possibilities; see Figure 3), and delay in target appearance (0.0 s, 0.5 s, and 1.0 s). The performance in the primary task (Stroop test) in terms of correct answer rates as well as the performance in the secondary task (target detection) in terms of correct detections served as dependent variables.

Five hypotheses were tested:

H1: Rates of correct answers in the Stroop task are influenced by the position of the task in the 3D display.

H2: The target detection rate is diminished outside the plane (see Figure 3) where the Stroop task is located.

H3: The target detection rate is diminished at target locations that are more than one position away from the location of the Stroop task.

H4: The target detection rate depends on the delay of target appearance.

H5: There is a trade-off between the performance in the Stroop task and the performance in target detection.

Results

Stroop task performance in relation to its position (H1)

For all positions, the percentage of correct answers to the Stroop task was very high with 94.1% for S1, 93.7% for S2 and 95.6% for S3. Wilcoxon tests showed that the difference between S2 and S3 was the only significant one (α=.02). High percentages of correct responses are common for the Stroop Test; yet, the rates in this experiment indicate that participants had really concentrated on their primary task.

Target detection in relation to the plane of the Stroop task position (H2)

For analytical purposes, the 3D display was segmented into nine target planes which can be defined and classified on the basis of the coordinate system (see Figure 3):

- Vertical planes are spanned by the x and z dimension and are arranged along the y dimension. They are termed *upper [*1], *middle* [2] and *bottom* [3] plane.
- Horizontal planes are spanned by the y and z dimension and are arranged along the x dimension. They are termed *left* [4], *central* [5] and *right* [6]) plane.

- Depth planes are spanned by the x and y dimension and are arranged along the z dimension. They are termed *foreground* [7], *middle ground* [8] and *background* [9] plane.

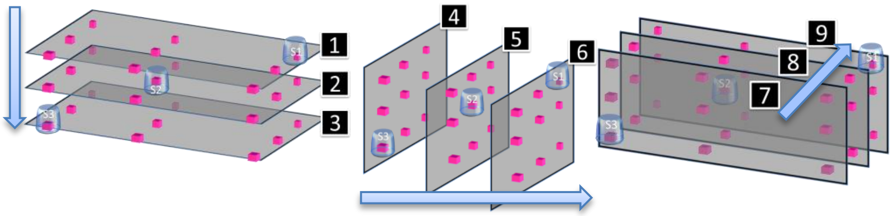


Figure 3. Segementation of the display space in three planes per dimension (Note: colours differ from the experimental design for illustrational purposes).

Target detection rates were analysed by Wilcoxon tests (α=.02). For reasons of simplicity, only significant differences are listed below. The detection rates for each group of planes in relation to the position of the Stroop task are given in brackets.

Vertical planes:

- For S1: upper plane (89%) > middle plane (68%) > bottom plane (22%)
- For S2: upper plane 92%) > middle plane (68%) > bottom plane (49%)
- For S3: upper plane (62%) > middle plane (53%) > bottom plane (32%)

Horizontal planes:

- For S1: right plane (76%) > central plane (63%) > left plane (26%)
- For S3: left plane (58%) > right plane (42%)

Depth planes:

- For S1: Foreground plane (61%) > background plane (35%)
- For S1: Middle ground plane (62%) > background plane (35%)

- For S2: Foreground plane (69%) < middle ground plane (83%)
- For S2: Middle ground plane (83%) > background plane (56%)

- For S3: Foreground plane (72%) > background plane (8%)
- For S3: Middle ground plane (72%) > background plane (8%)

Target detection rate in relation to the distance between target location and Stroop task location (H3)

In addition to the detection rates for the different planes of the display, target detection was investigated in relation to the proximity of the target position to the Stroop test position. For this purpose, the mean detection rates in the direct vicinity of a Stroop task (i.e., only position away) were compared with the mean rates at target positions further off. Figure 4 gives an example for position S3. Targets in the direct vicinity of S3 are magenta-coloured and connected to it. An equivalent area exists for S1. Since for S2 all targets are only one position away, the analysis is confined to S1 and S3.

Comparisons of mean detection rates between target positions close to the Stroop task and those that were further away were accomplished with the *Yates's chi-squared test*. For S1, the empirical chi-square value was significantly above the critical one (χ^2emp =78.68 and χ^2crit (α=.05; df=1; 1-tailed) = 2.71). Targets near the Stroop task position were detected more frequently (81.77%) than targets further off (43.42%), indicating that spatial proximity improved detection. The results for S3 point in the same direction (χ^2emp = 20.41 and χ^2crit (α=.05; df=1; 1-tailed) = 2.71). Again, near targets were detected more frequently (65.62%) than far targets (45.83%).

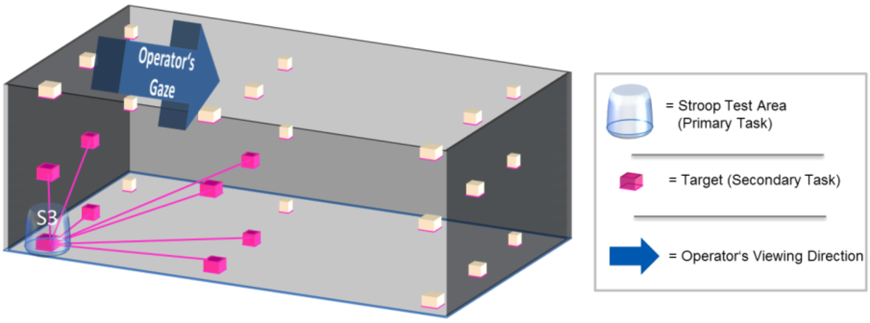


Figure 4. Targets with high spatial target proximity to the primary task (Note: colours differ from the experimental design for illustrational purposes).

Target detection rate in relation to the delay in target appearance (H4)

Again, detection rates were analysed with Wilcoxon Tests (α=.02). Significant differences were found for S1 and S2, but not for S3.

- S1: no delay (63%; SD=.094) > 0.5s delay (57%; SD=.128) > 1.0s delay (47%;SD = .131)
- S2: 1.0s delay (81%; SD=.138) > 0.5s delay (63%; SD=.214)
- S2: 1.0s delay (81%; SD=.138) > no delay (69%; SD=.161)

Figure 5 illustrates the significant differences between target detection rates for the different levels of delays and for the Stroop task positions S1 and S2. Trade-off effects (H5).

Performance in dual tasks is susceptible to trade-offs. Good performance in the primary task may lead to an impairment of performance in the secondary task (or vice versa). To check for trade-off effects in the current experiment, performance in the Stroop task was contrasted against performance in the detection task. As can be seen in Table 1 performance in the Stroop tasks at S2 was descriptively worse than at S1 and S3. As reported above, the difference between S2 and S3 was also significant. The corresponding detection rates are also shown in Table 1. *Related-Samples Wilcoxon Signed Rank Tests* revealed that target detection at S2 was significantly higher than detection at S1 as well as at S3 (α=.02).

The overall correlation between the number of correctly answered Stroop stimuli and detected targets of r=-.019 (Pearson's r) and r^2=.0004 (coefficient of determination). The overall and separate coefficients for S1, S2 and S3 can be found in Table 1.

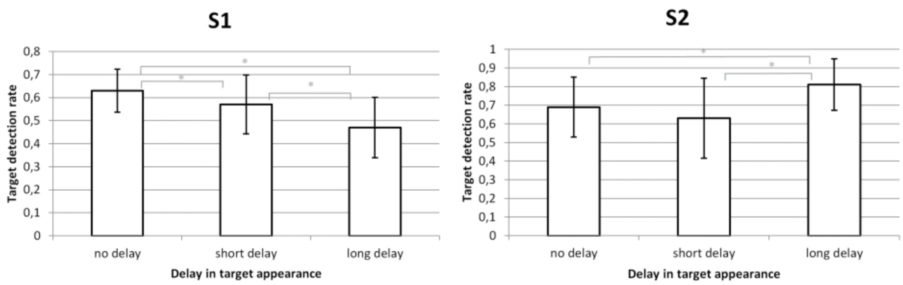


Figure 5. Target detection rates per level of delay in target appearance for Stroop task positions S1 and S2.

Table 1. Rates of correct answers in the Stroop task and rates of correct target detection per position of the Stroop task.

Stroop Task position	**Correct Answers in the Stroop Task**	**Correct target detections**	**Pearson's r**	r^2
S1	94.1%	55.7%	.291	.084
S2	93.7%	70.9%	.096	.009
S3	95.6%	50.1%	-.034	.001
Overall	94.5	58.9	-.019	.0004

Discussion

Five hypotheses guided the investigation of visual attention in the present study. The first one postulated that the rate of correct answers in the Stroop task was influenced by the position of the task in the 3D display. Although the rate of correct answers was generally high, it was worst at S2, i.e., in the centre of the display. This result is rather surprising and requires further investigation. As the following discussion will show, a number of factors may influence performance and the distribution of visual attention in a 3D display.

The second hypothesis was based the assumption that (a) attention would be focused on the location of the primary task and (b) that target detection at positions on the same plane would benefit from the resulting proximity. Therefore, it was predicted that target detection would be better on the plane that contained the Stroop task compared to the other planes. Results for the vertical planes contradicted this hypothesis. For all positions of the Stoop task, there was an advantage in target detection on the upper plane. One explanation might be derived from the participants' posture in front of the 3D workbench. Since they looked at the display from slightly above, the resulting visual angle might have facilitated detection on the upper plane. Future research is needed though to clarify this point. It is possible that the advantage disappears or shifts to another plane when the participants' visual angle is changed.

For the horizontal planes, target detection was significantly better in the periphery when the primary task was located at the respective side. The performance was best on the right plane, when the Stroop task appeared at S1, and best on the left plane when the Stroop task was presented at S3. No similar effect, however, was found for S2. Hence, the results for the lateral planes are in line with H2, while the result for the central one is not.

Differences in detection were also found on the depth planes. When the Stroop test was located at S3, target detection was best in the foreground, second best in the middle ground and very low in the background. This result accords with H2. Results for S2 support the hypothesis as well. The detection rate was significantly better on the middle ground plane than on the foreground and on the background plane. These results support the assumption that a close proximity of targets to the position of the primary task improves detection. Different results, however, are found for S1. Here, the Stroop test was located in the background, yet detection was best in the foreground and second best in the middle ground. A possible explanation can be derived from the theory by Atchley et al. (1997).The funnel-shaped attentional gradient encompasses the fore- and middle ground, but not necessarily the background. Furthermore, the perceptual load might have been too low to enable strong depth-awareness. Since this is a post hoc explanation, it requires further investigation in follow-up studies that treat perceptual load as an independent variable.

In summary, the results for H2 are ambiguous. While a particularity of the experimental (i.e., the participants' position in front of the display) might be responsible for the results on the vertical planes, differences in target detection on the other planes support the assumption that the proximity of targets to the location of the primary task improves the participants' performance. However, proximity might not be the only factor that impact detection. The attentional gradient as well as perceptual load might also play an important role.

In hypothesis three, proximity was operationalized in a different way. The direct vicinity of a Stroop task was defined as the set of targets that were only one position away from it. It was assumed that target detection in the vicinity would be better than at locations further off. This assumption was supported for the Stroop tasks located at S1 and S3. It seems that the primary task attracts VA and creates an attentional halo around its position. Inside the halo, detection seems to be better than outside. However, it must be noted that this kind of proximity could not be investigated for S2 because all targets were just one position away from this location.

The fourth hypothesis stated that target detection would be influenced by the delay of target appearance. In line with this prediction, the delays significantly impacted target detection rates at the Stroop test locations S1 and S2. While it was advantageous when the target was not delayed with the primary task at S1, the opposite is found for S2. Here, the longest delay (1.0 s) provokes the highest detection rate. No influence of delays is found for S3. Hence, the delay between the presentation of the primary task (Stroop tests) and the secondary task seems to affect target detection although no general pattern can be identified.

Hypothesis five addressed the possibility of trade-offs between the performance in the Stroop task and the performance in target detection. Results provide evidence for such trade-off effects. At S1 and S3 correct answers to the Stroop tasks are very frequent (94,1% and 95,6%) while the corresponding detection rates are comparatively small (55.7% and 50.1%). This ratio is changed at S2. Here, performance in the Stroop task is worst (93.7%) while the detection rate is substantially increased (70.9%). However, there are very small correlations between the performance in the Stroop test and the detection rate for S1 ($r=.29$, $r^2=.084$), S2 ($r=.096$, $r^2=.009$) and S3 ($r=-.034$, $r^2=.001$) separately and overall ($r=-.019$, $r^2=.0004$). Therefore, the constellation of performance in the primary and the secondary task does not support the assumption of trade-off effects in the current experiment.

To summarize, this study took a first step towards the development of an ATTENDO-based (Hillebrand et al., 2012) evaluation method for 3D settings. It served as a proof of concept that visual attention in stereoscopic displays can be experimentally investigated with a domain independent dual task paradigm. This paradigm provides a comparatively simple, cost-efficient methodology for future research. Moreover, the results of the study raised a number of new issues that must be investigated to gain deeper insights into the distribution of attention in 3D displays.

References

Atchley, P., Kramer, A.F., Andersen, G.J. ,& Theeuwes, J. (1997). Spatial cuing in a stereoscopic display: Evidence for a “depth-aware” attentional focus. *Psychonomic Bulletin & Review*, *4*, 524–529.

Dalrymple-Alford, E., & Budayr, B. (1966). Examination Of Some Aspects Of The Stroop Color-Word Test. *Perceptual And Motor Skills, 23(3f)*, 1211–1214. Doi:10.2466/Pms.1966.23.3f.1211

Hillebrand, A., Wahrenberg, E., & Manzey, D. (2012), A new method to assess pilots’ allocation of visual attention using a head-up display. In D. De Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier, (Eds.) (2012), *Human Factors: a view from an integrative perspective. Proceedings HFES Europe Chapter Conference Toulouse 2012*, Downloaded from http://hfes-europe.org

Hunt, A. (2003). Covert and overt voluntary attention: linked or independent? *Cognitive Brain Research*, *18*, 102–105.

Kollmorgen, S., Nortmann, N., Schröder, S., König, P., & Friston, K. J. (2010). Influence of Low-Level Stimulus Features, Task Dependent Factors, and Spatial Biases on Overt Visual Attention. *PLoS Comput Biol*, *6* (5), e1000791.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*, 163–203.

Michalczik, F., Hillebrand, A., & Dreßler, A., (2013). Old dog, new tricks - Adapting the Stroop test for visual attention assessment. In: Ulrich Ansorge, Erich Kirchler, Claus Lamm and Helmut Leder (Eds.): Abstracts of the 55th Conference of Experimental Psychologists. March, 24th to 27th, 2013, Vienna, Austria. Lengerich: Pabst Science Publ, p. 187.

Wolfe, J.M. & Horowitz, T.S. (2004). Opinion: What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*, 495–501.

Wright, R.D. & Ward, L.M. (2008). *Orienting of attention*. Oxford , New York: Oxford University Press.

The assessment of a new cockpit colour concept using the Occlusion Method

Martin Götze[1], Antonia S.Conti[1], Andreas Keinath[2], Tarek Said[2], & Klaus Bengler[1]
[1]Institute of Ergonomics, Technische Universität München
[2]BMW Group, München
Germany

Abstract

The first vehicle cockpit colour concepts were developed based on the light used in submarine vehicles. This was primarily because operators were not typically exposed to external light sources and operated under the scotopic visual range. Red lighting was used as it does not disrupt dark adaptation. Based on this, many automobile companies developed their night time concepts in red. However, when driving in an urban setting at night or twilight, the lighting level is described as mesopic. Because of this, a white cockpit colour concept becomes a viable option. In this study, two colour concepts were compared under mesopic vision conditions. The main objective was to assess whether both concepts yielded similar results in terms of interpretability, readability, and differentiability of information. For the experiment, 30 participants performed an occlusion task. A cockpit display with a speedometer in either colour appeared. The cockpit was presented for different duration times to simulate short glances at the speedometer. Statistical tests were performed to examine global response accuracy and mean accuracy for particular presentation times. No significant differences were found. In sum, this paper confirms that a white concept shows no disadvantages relative to a red concept under mesopic lighting conditions.

Introduction

Motivation

Operating a vehicle is visually, cognitively, and physically loading for the driver (Recarte & Nunes, 2003). The coordination of these processes is especially crucial for driving under specific conditions, like at night time or twilight (Eloholma et al., 2006); the driver is continually required to adjust and readapt (e.g. through accommodation and saccades) to items inside the car and on the road. By creating more efficient displays in the vehicle that draws not too much attention away from the primary driving task, will help to decrease workload and provide for a safer driving experience. Neale et al. (2005) present the consequences for a lack of concentration while driving. About 78% of traffic collisions and about 65% of near collisions occurred because of “driving-related inattention to the forward roadway and non-specific eye glance” as well as “secondary task engagement and fatigue”.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Illumination

The human visual system operates over a wide range of luminance from about 10^{-6} up to 10^{6} cd/m². There are three light levels defined for human vision: photopic, mesopic, and scotopic (Table 1). The luminance level for the photopic vision is defined as higher than 3-10 cd/m², the mesopic vision range from 0.001 up to 3-10 cd/m² and the luminance level lower than 0.001 cd/m² is defined as scotopic vision (Boyce & Rea, 1987; Dacheux & Raviola, 2000). It is today acknowledged that the upper luminance limit of the mesopic region cannot be precisely defined (Viikari et al., 2005; Plainis et al., 2005)

Table 1. Functional ranges of visual system capabilities (Boyce, 2006, p. 652)

Name	Luminance Range (cd m^{-2})	Photoreceptor Active	Wavelength Range (nm)	Capabilities
Photopic	>3	Cones	380–780	Color vision, good detail discrimination
Scotopic	<0.001	Rods	380–780	No color vision, poor detail discrimination
Mesopic	>0.001 and <3	Cones and rods	380–780	Diminished color vision, reduced detail discrimination, and a shift in spectral sensitivity as adaptation luminance moves from photopic to scotopic

But the perception of certain stimuli under different luminance conditions does not show equal efficacy. Driving at night time is riskier than during the day. Although only 25% of all traffic is present at night, the number of accidents for night and day is the same (Rumar, 2002). Studies on visual performance while driving with different light levels show that stimuli are perceived differently. With a decreasing luminance level, the reaction time and error rate increases (Alferdinck, 2006).

Mesopic Vision in urban Areas

The first in-vehicle cockpit colour concepts were developed based on the light and luminance used in subaqueous vehicles because operators were deprived of external light sources and operated under the scotopic visual range as described above (Boyce, 2006). The visual system adapts to the dark in order to adjust to certain luminance conditions. This dark adaptation (Purkinje effect) applies to red lighting because the spectral luminous efficiency function shifts to lower wavelengths, where red (700 nm) is not affected (see Figure 1). The rods of the eye do not become saturated, which is one reason why the first colour concepts for in-vehicle cockpit displays were red.

Studies have shown that under low luminance levels and high travelling speeds, more errors occur and additional steering effort is needed (Alferdinck, 2006). Ordinarily, driving at night in rural areas is classified in the scotopic range. However, the light level in urban areas when driving at night or twilight is described as mesopic. The main reason for this is the higher luminance level caused by the signals, signs, street lights, and other vehicles (Stockman & Sharpe, 2006; Viikari, 2008). Additionally, internal light sources also contribute to the overall light level as many driver assistance and information systems, such as navigation or entertainment functions, are completely electronic. Since it is now proven that driving in urban areas does not operate under the scotopic but rather mesopic range, the opportunity is presented for a new vehicle cockpit colour concept.

The main objective of this study was to assess whether a new cockpit colour yielded similar results in terms of the interpretability, readability, and differentiability of information compared to the old concept in red. However, only the interpretability and differentiability are discussed in this paper.

- Interpretability and differentiability: How accurate is the information read from each display? How efficient is the recognition and identification of the target?

Both objective and subjective data was collected. Participants performed two experiments and evaluated the aforementioned criteria with a questionnaire. In this paper, only experiment one is presented where the readability, interpretability and differentiability criteria were tested. The data of the questionnaire was published earlier (Götze et al., 2013). The analysis of the data of the second experiment is still ongoing.

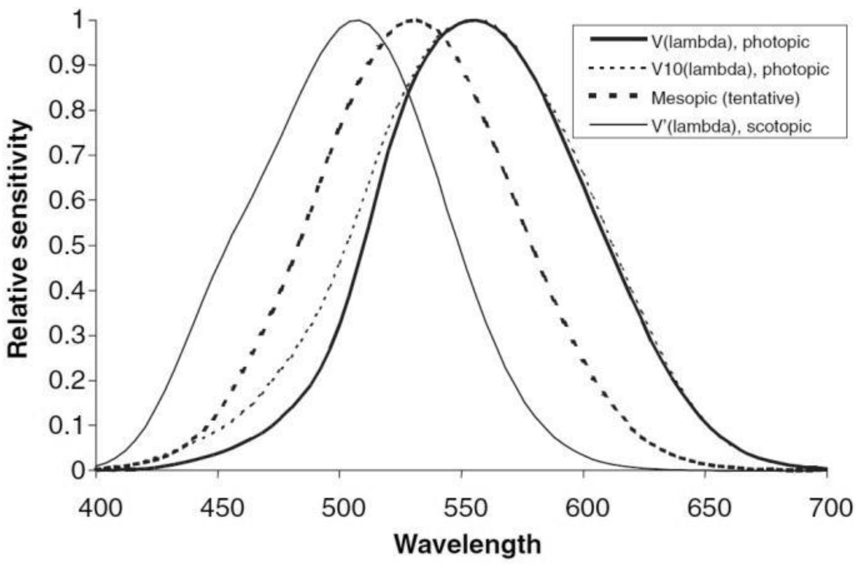


Figure 1. The CIE spectral luminous efficiency functions for photopic vision, V(k) and V10(k), and scotopic vision V'(k) compared with an example of a tentative spectral mesopic function for a typical mesopic light level (Alferdinck, 2006)

Method

Occlusion technique

The occlusion technique (ISO 16673:2007) is used to assess visual demands of a task (e.g. stimulus presented on a display). During the experiment there are two main parameters to be varied: The time duration in which a certain stimulus is visible and the time frame between two stimuli (Baumann et al., 2004). Moreover, there are two different ways to perform an occlusion experiment. The participants either wear shutter glasses or the stimulus on a screen is shut on and off. In this study the latter was used to maintain mesopic vision throughout the experiment.

Framework conditions of the study

Displays

The study compared a currently used cockpit concept (Figure 2) in two different colours. It can be found in some mid-range cars. The size of the whole cockpit display on the screen was 316x117 mm. Both rings, the tachometer and the speedometer, had a diameter of 96.4 mm. These measures correspond to those of a real existing instrument cluster in automobiles. As a baseline colour concept, the red (603.2 nm) colour concept was used compared to white. White was chosen because it is the colour produced by the reflection, transmission or emission of all wavelength of visible light. The luminance level of the cockpit display was tailored for mesopic vision conditions and set to 7.8 cd/m² for the red display and 11.0 cd/m² for the white display.

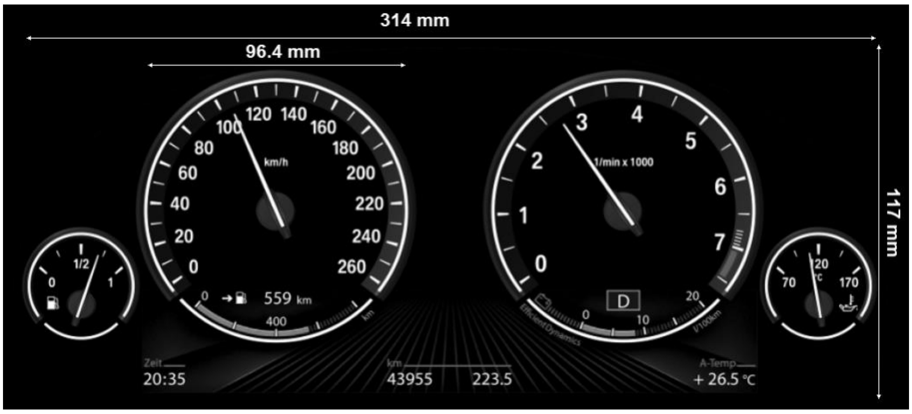


Figure 2. The cockpit concept used in this study with its measurements

Experimental design

The study was conducted in an experimental room under mesopic lighting conditions in a range from 0.01 cd/m² up to 1 cd/m². Participants sat on a chair in front of the screen with their head on a chin rest to ensure a viewing distance of 70 cm (see Figure 3). The Screen was positioned in front of them. Between each presentations a fixation cross, located on the top part of the screen, was presented to simulate a saccade to the driving scene. Furthermore, a small fake camera was placed under the screen in order to motivate the participants not to anticipate the position of where the speedometer would be presented and to stimulate the small saccade of eye movement during the experiment.

Figure 3 Setup of the occlusion task with a screen in front of the participants and a chin rest for the head.

Procedure

At the beginning of the experimental session, all participants performed a visual acuity test (Landolt ring test) to test visual acuity and to ensure that each participant met the minimum acuity for driving (0.5 according to Colenbrander and De Laey (2005)). While performing under mesopic conditions, a visual acuity of at least 0.2 is needed (Uvijls et al., 2001). Additionally, participants also performed a colour vision test. This was especially important considering the nature of the experiment. Afterwards, a demographical questionnaire was administered to participants. Test persons entered the experimental room under mesopic lighting followed by a mesopic adaptation procedure of 20 to 30 minutes (Lamb & Pugh, 2004).

The aim of this experiment was to evaluate under mesopic lighting, whether participants were able to accurately indicate the speed indicated on the cockpit speedometer. Four different display durations and two colour concepts were used for this evaluation. It was additionally important to be able to compare the two colour concepts in order to investigate which concept, if any, would yield better performance in the task. Participants verbally reported the displayed speed to the experimenter. Performance was not based on the quickness of the report; however, the speed was to be reported before the next stimulus was presented.

Pre-experiment
A pre-experiment was run for this study. Six participants performed 58 trials each colour, in addition to a training session in the beginning. The aim of the pre-experiment was to find the perfect occlusion times for the main experiment. Three presentation times were used in this pre-experiment: 300, 350 and 400 ms. The results revealed that the accuracy and misses did not differ much between all three times. Therefore, in the main experiment, the display durations were shifted down in order to test lower times and provoke errors.

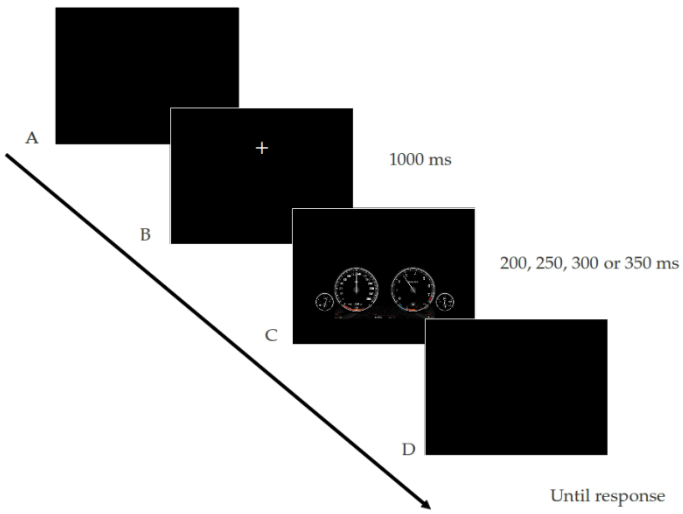


Figure 4 Time sequence of stimuli presented in the experiment. The screen remained black until the participant was ready (A). Afterwards, a white fixation cross on the top of the screen was presented for 1000 ms (B). The cross disappeared and participants were presented a cockpit display on the bottom of the screen in the respective colour of the block, where a certain speed was indicated (C). After the presentation time, the screen turned black again until the next stimuli was presented (D).

Main experiment
Figure 4 shows the procedure of the main experiment. All participants performed a training block before the experiment began. During the training, participants were familiarized with the task and were given feedback on the correctness of their report. No such feedback was given in the main experiment. The experiment was presented to participants as per the sequence shown in Figure 4. The cockpits were displayed for 4 durations: 200, 250, 300, and 350 ms. Possible speeds indicated were 30, 60, 90, 130, 140, 160, 180, 200, and 220 km/h indicated by the speedometer. Participants were not informed that only specific speeds would be shown. The

different distances between speeds were chosen to decrease predictability. Minimum and maximum speeds were excluded from the experiment as they were considered too easy.

Data acquisition

All presentations in this study were prepared and executed with E-Prime 2.0 (Psychology Software Tools, Inc.). Performance metrics (viz. accuracy rates) were also recorded with E-Prime.

Participants

Thirty healthy volunteers participated in this study. The 6 participants that took part in the pre-experiment were not included in the main analysis. Only male subjects were recruited for this study, as several studies show the effects of sex and age on visual performance. Another part of the experiment consisted of a choice reaction time (CRT) task. While simple reaction time (SRT) tasks are associated with age and sex (for example, males perform some SRT tasks faster than females across the life span, whereas women are more accurate (Der & Deary, 2006)), mean CRT tasks tend to be very inconsistent in the considered age range (~25-55). For example, the CRT increases and becomes more variable with age while gender shows no regularity as seen in Figure 5. Additionally, there seems to be different RT patterns for females and males as a function of stimulus location (Adam et al., 1999). Although RT was not relevant in this experiment, this was a crucial factor to the second experiment not reported here. Only males were accepted as participants in order to control the homogeneity of the sample.

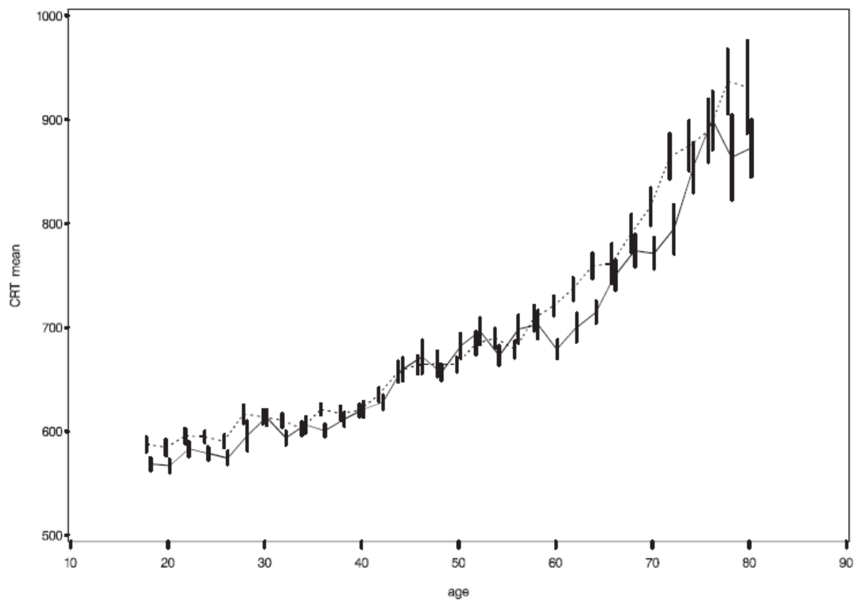


Figure 5. Choice reaction time (CRT) mean of a visual performance study of male (solid) and female (dashed) participants (Der & Deary, 2006)

Results

Participants

All participants were between 24-54 years of age with a mean of 43 years. No participants reported to suffer from any visual or motoric impairment. The driving experience varied from 15-36 years with a mean of 24.6 years and an average driving distance of 25171 km per year (ranging from 8000 to 70000 km).

Occlusion Task

Each participant was presented with 32 training trials and 64 experimental trials, for both colours, red and white. The mean accuracy rate for each presentation duration and the mean global accuracy rate (across all presentation durations) were calculated for all 30 participants. Only accurately reported speeds were considered correct. No response or reporting an inaccurate speed were considered errors.

Mean global accuracy

The mean global accuracy for all participants and all four time frame conditions was calculated. A paired-sampled t-test was performed to examine any difference in global accuracy rate. The mean and SDs for each colour concept can be found in Table 2; Figure 6 graphically depicts these findings. No significant difference between the two colour concepts in terms of global accuracy was found.

Table 2. Mean global response accuracy for two display colour concepts (N = 30)

	White	*Red*
Mean	0.75	0.74
SD	0.13	0.17

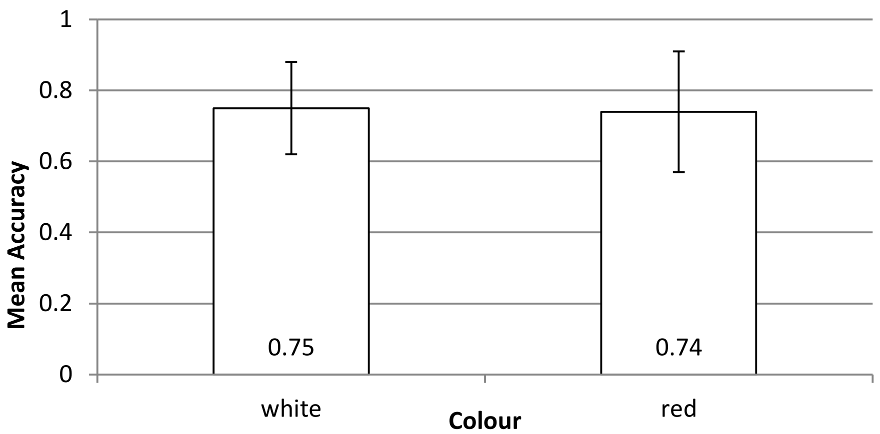


Figure 6. Mean global accuracy with standard deviation for responses across all participants and presentation times for the two colour concepts

Mean accuracy for different presentation times
The mean accuracy for all four different presentation times and both colours was calculated. Accuracy percentages are given in Table 3.

Table 3. Mean accuracy for different presentation times for two display colour concepts

	200 ms	*250 ms*	*300 ms*	*350 ms*
White	0.68	0.75	0.76	0.83
SD White	0.129	0.125	0.152	0.115
Red	0.65	0.73	0.77	0.80
SD Red	0.186	0.190	0.137	0.167

A one-way repeated measures ANOVA was done. Mauchly's test indicated that the assumption of sphericity had been violated, $X^2(5) = 11.35$, $p = .045$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .804$). The results show that there was no significant effect of colour concept on mean accuracy for the different presentation durations $F(1, 29) = .409$, $p = .53$. These results suggest that the colour concept has no effect on the accuracy of reading and reporting a speed from the cockpit (Figure 7).

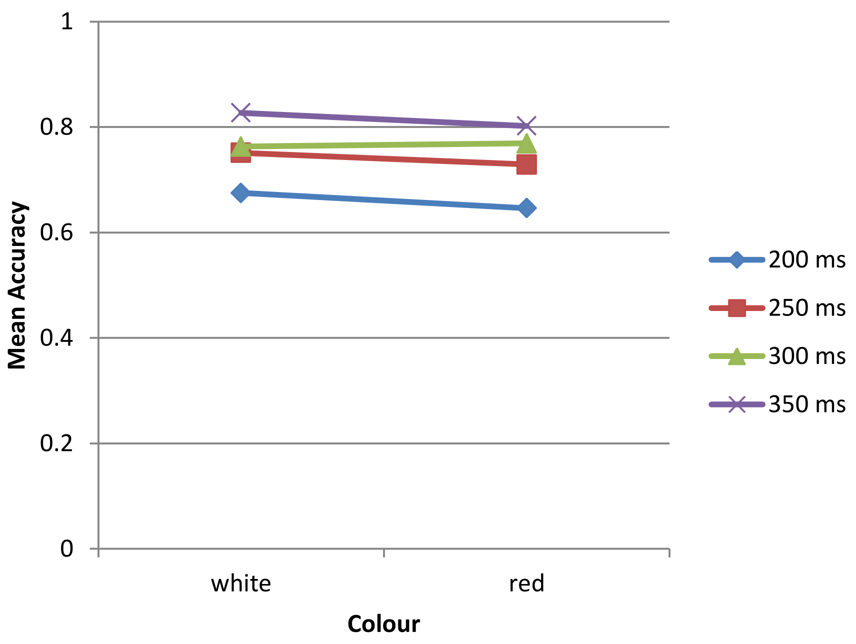


Figure 7. Mean accuracy for all presentation times with each cockpit colour

Additionally, there was a significant main effect of presentation time, $F(2.41, 69.98) = 44.80$, $p \leq .001$, suggesting that presentation time significantly affected mean accuracy (Figure 8).

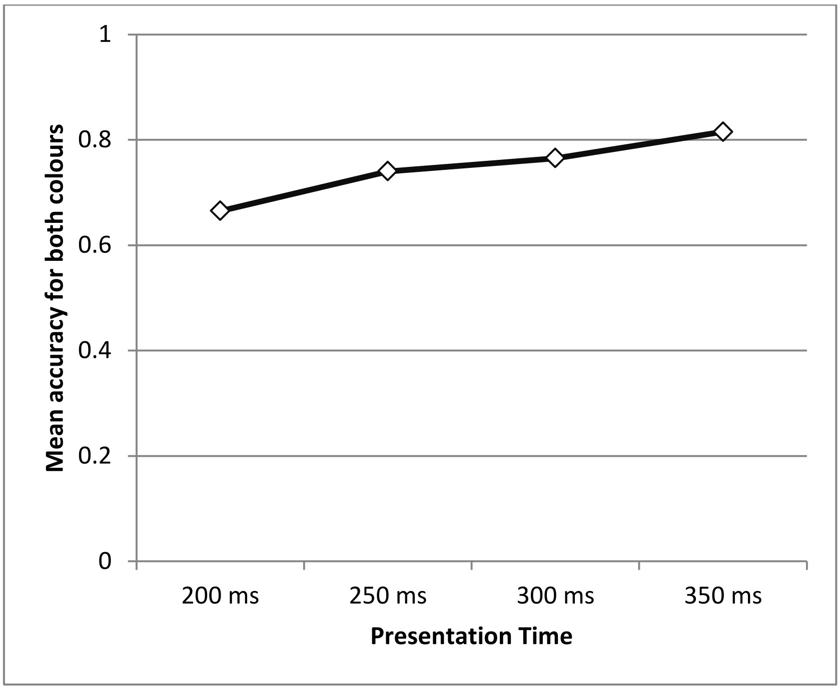


Figure 8. Mean accuracy including both colours for all presentation times

Post-hoc Bonferroni comparisons indicated that all presentation times were significantly different from each other, $p \leq .001$, except for 250ms and 300ms, $p = .12$ where no difference was found. These results suggest that it gets significantly better to read a specific speed accurate in a cockpit display no matter which colour but depending on the presentation time. Shorter presentation times lead to significantly less accurate answers then longer presentation times.

Non-inferiority

Since a main aim of this study was to compare the new colour (white) cockpit concept to the already established one (red), a non-inferiority test was also carried out. With non-inferiority testing, the margin must be smaller than or equal to “the smallest value that would represent a meaningful difference, or the largest value that would represent a meaningless difference. The determination of this margin must be based on both statistical reasoning and [expert] judgment” (Henanff et al., 2006, pg. 1147). It is usually expressed with the confidence interval of:

$$\bar{x} \pm z * \frac{\sigma}{\sqrt{n}}$$

A “meaningful difference” is typically defined case by case. Since 3 of 4 presentation times are already more accurate for the new white colour concept, only accuracy for the presentation time of 300 ms is important in this test. Figure 9 shows the 95% confidence interval for the mean accuracy of each colour and each of the four presentation times. The graph shows that there is no meaningful difference between the both colour concepts.

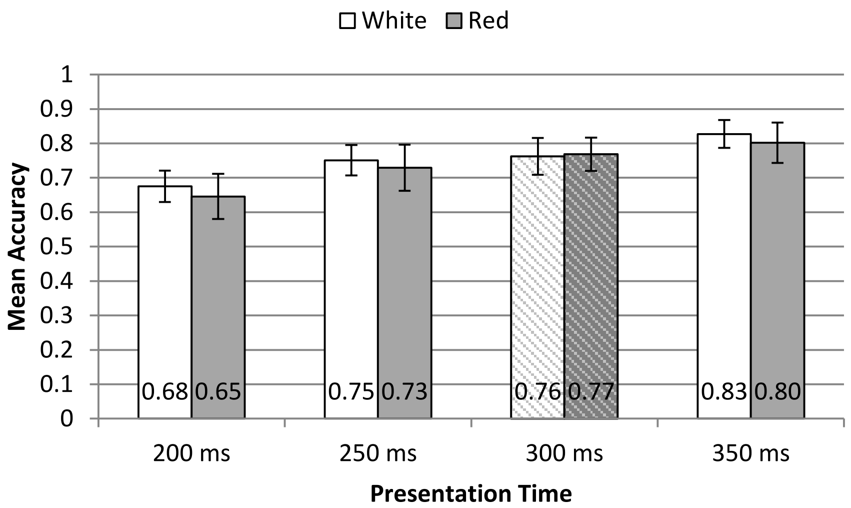


Figure 9. 95% Confidence Intervals for the mean accuracy of both colours and all presentation times

Discussion

In the current study, two car cockpit colour concepts (white and red) were compared in order to examine them in terms of readability of information under mesopic lighting conditions. The occlusion technique was used to facilitate a comparison between the readability of a displayed speed for specific presentation duration. The aim was to investigate whether two display colours yielded the same results for the presentation times 200, 250, 300, and 350 ms. Mean accuracy and global accuracy were included in the analysis. No significant differences between the two colours were found. A significant main effect of the presentation time was found, meaning that the longer the presentation time was, the more correct responses were reported. This result is in line with studies reporting increased visual performance according to an increase in stimulus presentation time (Baumann et al., 2004). A final non-inferiority test showed no meaningful difference between the two colour concepts, implying that both colour concepts are associated with the same performance outcomes. The NHTSA report (Klauer et al., 2010) shows that there is a “2-second rule” which is based on research by Rockwell (1988). His studies showed that the 85th percentile eye glance length was about 1.9s. Still some researchers like Green (1999), Wierwille (1993) or Dingus et al. (1989) suggested a time not greater than 1.5s to look away from the forward roadway. The results of the experiment show, independently of the colour the speed can be read even in very short presentation times. The lighting colour of the cockpit concept is not important under mesopic vision as long as the driver stays adapted to those luminance levels.

Conclusion

In short, the study shows that a new cockpit colour concept, as tested, has no objective disadvantage over the currently used red colour concept under mesopic lighting. These results will help designers of displays or car manufacturers with new possibilities for interior lighting concepts for night-time driving in urban areas.

Acknowledgments

The authors would like to acknowledge the cooperation with the BMW Group on this project. We appreciate the opportunity to have carried out this study.

References

Adam, J.J., Paas, F.G., Buekers, M.J., Wuyts, I.J., Spijkers, W.A., & Wallmeyer, P. (1999). Gender differences in choice reaction time: evidence for differential strategies. *Ergonomics 42*, 327-335

Alferdinck J.W.A.M. (2006). Target detection and driving behavious measurements in a driving simulator at mesopic light levels. *Ophtalmic and Physiological Optics, 26*, 264-280

Baumann, M., Keinath, A., Krems, J.F., & Bengler, K. (2004). Evaluation of In-vehicle HMI using occlusion techniques: Experimental results and practical implications. *Applied Ergonomics, 35*, 197-205

Boyce, P.R, & Rea M. (1987). Plateau and escarpment: the shape of visual performance. Proc. 21st Annual Session of CIE, Venice, Italy, 82-85

Boyce, P.R. (2006). Illumination. In G. Salvendy (Ed.) *Handbook of Human Factors and Ergonomics* (pp. 643-669). Hoboken, NJ, USA: John Wiley & Sons, Inc,

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

Dacheux, R.F., Raviola, E. (2000). Functional anatomy of the neural retina. In D.M. Albert, F.A. Jacobiec, D.T. Azar, and E.S. Gragoudas (Eds.) Principles and Practice of Ophthalmology, 2nd ed. (pp. 1601-1624). Philadelphia: W.B. Saunders Company.

Der, G., & Deary, I.J. (2006). Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. *Psychology and Aging, 21*, 62-73

Dingus, T.A., Antin, J.F., Hulse, M.C., & Wierwille, W.W. (1989). Attentional demand requirements of an automobile moving map navigation system. *Transportation Research Part A, 23*, 301-315

Eloholma, M., Ketomaki, J., Orrevetelainen, P., & Halonen, L. (2006). Visual performance in night-time driving conditions. *Ophthalmic and Physiological Optics, 26*, 254-263

Götze, M., Conti, A.S., Keinath, A., Said, T., & Bengler, K. (2013). Evaluation of a New Cockpit Colour Concept under Mesopic Lighting for Urban Driving. In: D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J.C. Mitchell et al. (Eds.): *Design, User Experience, and Usability. Web, Mobile, and Product Design*, Bd. 8015 (pp. 359-366). Berlin, Heidelberg: Springer Berlin Heidelberg,.

Green, P. (1999*). Visual and task demands of driver information systems* (Technical Report No. UMTRI-98-16). Ann Arbor, MI: UMTRI (The University of Michigan, Transportation Research Institute)

Henanff, A.L., Giraudeau, B., Baron, G., & Ravaud, P. (2006). Quality of Reporting of Noninferiority and Equivalence Randomized Trials. *JAMA 295*

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

Klauer, S.G., Guo, F., Sudweeks, J., & Dingus, T.A. (2010). An Analysis of Driver Inattention Using a Case-Crossover Approachon 100-Car Data: Final Report DOT HS 811 334

Lamb, T.D. & Pugh, E.N. (2004). Dark adaptation and the retinoid cycle of vision. *Progress in Retinal and Eye Research, 23*, 307-380

Neale, V.L., Dingus, T.A., Klauer, S.G., Sudweeks, J., & Goodman, M. (2005). An Overview of the 100-Car Naturalistic Driving Study and Findings. National Highway Traffic Safety Administration, United States

Plainis, S., Murray, I.J., & Charman, W.N. (2005). The role of retinal adaptation in night driving. *Optometry and Vision Science, 82*, 682-688

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

Rockwell, T. (1988). Spare visual capacity in driving revisited. Paper presented at the Vision in Vehicles II. Nottingham, England.

Rumar, K. (2002). Night driving accident in an international perspective. In: Proceedings in the First International Congress "Vehicle and Infrastucture Safety Improvement in Adverse Conditions and Night Driving" VISION (ed. Charlet M), Rouen, Suresnes, France

Stockman, A., Sharpe, L.T. (2006). Into the twilight zone: the complexities of mesopic vision and luminous efficiency. *Ophthalmic and Physiological Optics, 26*, 225-239

Uvijls, A., Baets, R., Leroy, B.P., Kestelyn, P. (2001). Mesopic visual acuity requirements for driving licenses in the EU – research report. *The Bulletin of the Belgian Society of Ophthalmology, 282*, 71-77

Viikari, M., Eloholma, M., Halonen, L. (2005). 80 years of V(**λ**) use: a review. *Light & Engineering, 13*, 24-36

Viikari, M. (2008). Modeling spectral sensitivity at low light levels based on mesopic visual performance. OPTH, 173

Wierwille, W. (1993). Visual and manual demands of in-car controls and displays. In B. Peacock and W, Karkowski (Eds.) *Automotive Ergonomics* (pp. 299-320). Washington, DC: Taylor & Francis

Measuring skill decay in process control - results from four experiments with a simulated process control task

Annette Kluge, Barbara Frank, & Julia Miebach
University of Duisburg-Essen, Duisburg, Germany

Although widely acknowledged in the field of process automation, skill decay of operators and its consequences has not been investigated intensively. In four laboratory experiments (*N*=252 subjects in total), skill decay of a trained start-up procedure of a water purification plant was measured after two (studies 1-3) and three (study 4) weeks with respect to the primary task performance (speed and accuracy of performing the start-up procedure) and secondary task performance to measure mental workload (MW). During the retention interval, experimental groups (EG) received a refresher intervention (RI), while control groups (CG) received no treatment (studies 1-4). The RIs were designed differently based on theories of skill proceduralisation (Practice-RI), the test effect (Skill Demonstration or Procedural Knowledge Test), Mental Practice (Symbolic Rehearsal) and on-the-job training (Work Experience). Results show that substantial skill decay occurs early on but can be countered by induced retrieval effort during RIs, which in turn affects primary task performance and MW differently. MW and variance in performance was lowest for the Practice-RI. RIs based on the test effect supported skill retention but affected MW negatively. Finally, all EGs (except the Practice-RI) and CGs varied in skill decay, which can be substantially attributed to person-related variables (mental ability/memory).

Introduction

The notion of skill degradation and skill decay in highly automated workplaces such as process control is widely assumed to be a noteworthy side effect of automation (e.g. Bainbridge, 1983; Kluge et al., 2012; Onnasch et al., 2013; Parasuraman et al., 2000; Sauer et al., 2000; Wickens & Hollands, 2000). It can potentially become an important safety issue, especially in non-routine situations, e.g. in the start-up or shut-down of a plant (Wickens & Hollands, 2000) or in cases of automation failure (McBride et al., 2013; Onnasch et al., 2013; Parasuraman et al., 2000). A non-routine situation is defined above all by the rarity with which it is performed (Kluge, in press). This rare performance creates a period of non-use, and through this non-use, processes of forgetting as expressed in a very low strength of retrieval occur (Bjork & Bjork, 1992; Bjork, 2009); and consequently, the performance level is not present in the necessary manner.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

The assumption concerning the process of forgetting refers to the *power law of forgetting* (Bourne & Healy, 2012; Wixted & Carpenter, 2007). “With the passage of time and the lack of opportunity to rehearse or refresh acquired knowledge and skill, performance declines and reflects failure to retain information” (Bourne & Healy, 2012, p. 4). The skill degradation is observed in an increased response time or decreased accuracy. The power law of forgetting (Wixted & Carpenter, 2007) can be thought of as the inverse of the power law of practice (Bourne & Healy, 2012). Forgetting is assumed to be a difficulty in recalling knowledge and skill, which is due to the process of losing access (Bjork, 2011). Losing access, in turn, refers to a decrease in retrieval strength, defined as the accessibility at a given point in time (Bjork, 2011; for more details see Bjork & Bjork, 1992, 2006). In particular, procedural tasks are highly liable to forgetting (Martinussen & Hunter, 2010; Farr, 1987).

The amount of forgetting or the decrease in retrieval strength in process control, has seldom been directly addressed and investigated. We therefore conducted a series of four experiments (Kluge et al., 2012; Kluge & Frank, resubmitted; Miebach, 2013) to investigate the parameters within which forgetting becomes apparent in procedural tasks such as standard operating procedures (SOPs). As is known from the law of forgetting, skill decay may lead to a loss of speed and accuracy in performing a task, and may additionally lead to an increase in attentional resources and the mental effort with which the task is performed. In detail this means that the task is not performed skill-based anymore (because it is less proceduralised) but needs to be performed knowledge-based (by using declarative knowledge; Kim, Ritter & Koubek, 2013) with less attentional resources available for e.g. secondary tasks.

In these four studies, we investigated the amount of skill degradation and the impact of various countermeasures. In the present paper, the focus will be on the aspects of skill decay in terms of a) speed loss and its consequences for production outcome, b) increased mental workload (a & b studies 1-4), and c) its correlation with person-related variables (study 4).

The task to be performed

The skill to be maintained and the simulated process control task

In all studies (Kluge et al., 2012; Kluge & Frank, in press) referred to here, the participants’ task was to start up a simulated plant in an eleven-step fixed sequence (Table 1). The start-up of a plant is assumed to be a non-routine task which requires skill retention. The start-up sequence is embedded in the operation of the Waste Water Treatment Simulation (WaTrSim, Burkolter et al., 2009). In WaTrSim, the operator’s task is to separate waste water into fresh water and solvent by starting up, controlling and monitoring the plant. The goal is to maximize the amount of purified water and to minimize the amount of waste water (primary task). This goal is achieved by controlling four main processes in WaTrSim, considering the timing of actions and following fixed sequences (Kluge et al., 2012). The start-up procedure was used to measure skill retention and skill decay, respectively.

Table 2. Sequence of start-up procedure trained. V1 – V4 are the abbreviation for e.g. valve 2, tanks in AWAsim are called Ba, Bb, Bc, Bd, Be, R1 and HB1 as well as heatings are labelled by H1 and K1.

Step #	**Temporal Transfer** (in Initial Training, trained start-up procedure)
Step 1	**Deactivate follow-up control** Operate controller V2 Set the target value from external to internal
Step 2	**Valve V1: Flow rate 500 l/h** Operate controller V1 Set target value 500l/h
Step 3	**Wait until content of R1 > 200 l**
Step 4	**Valve V2: Flow rate 500 l/h** Operate controller V2 Set target value 500l/h
Step 5	**Wait until content R1 > 400 l**
Step 6	**Valve V3: Flow rate 1000 l/h** Operate controller V3 Set target value 1000l/h
Step 7	**Wait until content of HB1 > 100 l**
Step 8	**Switch on heating H1** Operate controller HB1 set from manual to automatic operation
Step 9	**Wait until HB1 > 60°C**
Step 10	**Put column C1 into operation** Operate controller C1 set from manual to automatic operation
Step 11	**Valve V4: Flow rate 1000 l/h** Operate controller V4 Set target value 1000l/h

The general experimental setting of studies 1-3

The studies comprised a between- and within-group design with one or two experimental groups (EGs) compared to a control group (CG) and two measurement times (Figure 1).

In the present paper, we present the findings concerning skill decay in terms of

a) the CGs and their amount of skill decay resulting from speed loss and its consequences for production output, and
b) performance degradation in the secondary task of EGs and CGs (indicator of mental workload).

For studies 1-3 and all groups, each experiment included two measurement times: 1) performance at the end of the Initial Training (IT) and 2) performance at the Retention Assessment (RA) two weeks later. The EGs additionally received a refresher intervention (RI) between IT and RA (for details of the EGs, see Kluge et al., 2012, Kluge & Frank, in press). *CGs were assessed at the IT and the RA only and received no refresher intervention. The CGs' performance served as an indication for "pure forgetting" due to non-use.*

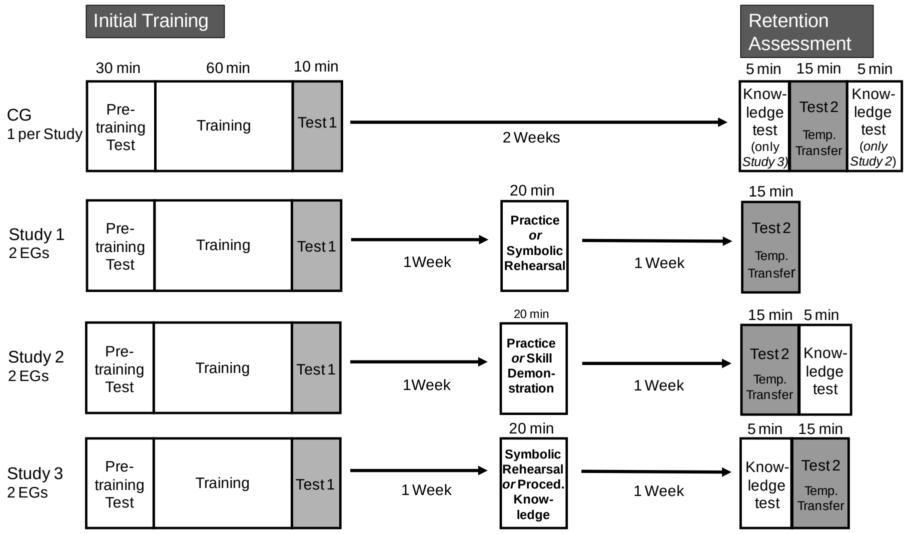


Figure 5. Procedure in studies 1-3; the CG procedure was identical in studies 1-3 and is described in CGs (in study 1, the knowledge test was not applied in RA; in study 2, the knowledge test was applied after the measurement of temporal transfer; in study 3, the knowledge test was applied before the measurement of the temporal transfer).

All participants were recruited through postings in university internet forums, handing out flyers and requests for participants during lectures. To ensure technical understanding, only students from faculties of engineering were recruited. All of them received course credits for their participation. All studies were approved by the ethics committee and subjects were informed about the purpose of the study and told that they could discontinue participation at any time (in terms of informed consent).

The initial training (IT): The IT session lasted for 90 minutes. Participants were welcomed and introduced to WaTrSim. After completing tests concerning person-related variables relevant for the study (described below), participants explored the simulation twice. They were then given information and instructions about the start-up procedure and practised performing the target eleven-step start-up procedure four times. During these first four trials, participants were allowed to use the manual. Following this, they then had to perform the start-up procedure four times without the manual and were told that they were expected to produce a minimum of 1000 litres of purified water. *The fourth and final trial of this series was used as the reference level of performance after training.* Finally, the ability of participants regarding their acquired knowledge was tested (knowledge test, described below).

The retention assessment (RA): The RA took about 20 minutes. After the participants were welcomed, they were asked to start up the plant two consecutive times. *The first trial was used to assess skill retention/decay.* The knowledge test, which was applied in the IT, was also applied at the end of the RA. The test was identical in all studies, except for the position in the RA (Figure 1).

In the IT and the RA, participants were trained and tested in groups containing a maximum of eight persons.

Independent variables

During the retention interval, EGs received a *refresher intervention* (RI) while control groups (CG) received no treatment. The RIs were designed differently based on theories of skill proceduralisation (Practice-RI), the test effect (Skill Demonstration or Procedural Knowledge Test), and mental practice (Symbolic Rehearsal), which are described in detail in Kluge and Frank (resubmitted).

Dependent variables

Skill decay: The start-up procedure (see Table 1) was used for the measurement of skill decay . The start-up procedure incorporates a secondary task which required participants to note the tank level of a specified tank every 50 seconds (scores 0-3) in order to measure mental workload.

Skill decay is therefore operationalised in terms of a) the production outcome, which equals the amount of produced purified water in litres at the measurement point of the RA (two weeks later; Table 2) *and b) secondary task performance.*

Control variables

Person-related variables: *Sex, age, general mental ability* (GMA, using the Wonderlic Personnel Test, 2002), as well as prior knowledge about waste water treatment and basic chemical understanding were assessed at the beginning of the IT (Figure 2). To measure *GMA*, verbal, numerical and spatial aspects of intelligence were assessed. The participants had twelve minutes to answer 50 items. Correct answers were counted (score 0-50 points). The *prior knowledge test* included seven questions about waste water (score 0-7 points).

Knowledge: At the end of the IT and the RA, participants filled in a knowledge test addressing declarative and procedural knowledge about operating WaTrSim. The test includes clozes, questions and diagrams about WaTrSim and background knowledge about waste water treatment (10 questions). Questions included “What are the goals in the start-up procedure in WaTrSim?”, “Which gadget is shown in the diagram?” or “What happens in waste water treatment?” (score 0-24 points).

Table 3. Overview of experimental parts and variables. EGs received a refresher intervention (RI) while control groups (CG) received no treatment

Studies 1-3 **Session Week 1** Initial Training (IT); 90 min	 **Session Week 3** Retention Assessment (RA); 20 min
• Pretraining Test: - Socio-demographic data - General mental abilities - Prior knowledge • Initial Training: - 2x Explore - 4x Start-up with manual • Test 1: - 4x Temporal transfer: Start-up without manual (*performance in fourth of four times was measured*) - Knowledge test	• Test 2: - Temporal transfer: Start-up (further called RA) - Knowledge test*

* Study 1: no knowledge test was applied, Study 2: knowledge test at the end of RA, Study 3: knowledge test at the beginning of RA

Results of studies 1-3

The descriptive statistics of all three studies are displayed in Table 3.

CG: For production outcome, a repeated measures ANOVA showed a significant effect of time and of group but no interaction of time and group (Figure 2, left side). For the secondary task, a significant effect of time and group was revealed, but no interaction of time and group (Figure 2, right side). For the knowledge test, a significant effect of time, no effect of group, but a significant interaction of time and group was shown (Table 4). This means that three control groups showed skill decay from IT to RA and that this skill decay differs substantially between groups, with a medium to large effect size.

EG: A repeated measures ANOVA for the secondary task with EGs revealed a significant effect of time, no significant effect of group, but an interaction of time and group was found (Figure 3). Moreover, a significant effect of time, group and interaction of time and group was found for production outcome (Figure 4). For the knowledge test, no significant effect of time, group or interaction was found (Table 4). That means that skill retention and decay depends on the refresher intervention (Figure 3).

Table 4. Descriptive Statistics of sample size, sex, age, prior knowledge, GMA and performance in IT and RA for CGs; M (SD)

	Study 1			Study 2			Study 3		
	CG 1 (*n*=19)	Practice (*n*=18)	Symbolic Rehearsal (*n*=20)	CG 2 (*n*=24)	Practice (*n*=22)	Skill Demo. (*n*=22)	CG 3 (*n*=24)	Symbolic Rehearsal (*n*=22)	Procedural Knowledge (*n*=22)
Sex	4 female	5 female	5 female	8 female	11 female	11 female	18 female	13 female	12 female
Age	21.42 (2.69)	21.44 (2.59)	21.15 (1.61)	20.88 (1.30)	22.86 (3.36)	21.55 (2.54)	20.58 (1.95)	20.09 (2.09)	20.95 (1.94)
Prior Knowledge (0-7)	5.10 (1.34)	5.56 (1.25)	4.80 (1.24)	4.63 (1.81)	4.86 (1.32)	5.23 (1.23)	5.17 (1.34)	5.50 (1.14)	5.23 (1.31)
GMA (0-50)	28.05 (5.19)	26.78 (4.92)	26.80 (3.94)	25.79 (4.38)	25.00 (3.72)	28.09 (5.80)	24.42 (4.34)	25.27 (4.34)	26.82 (4.87)
Prod. Outcome IT	1142.72 (99.33)	1107.96 (80.49)	1124.72 (77.02)	1092.31 (142.88)	1020.51 (366.32)	1123.55 (178.31)	1026.95 (346.58)	1119.62 (125.23)	1093.74 (150.02)
Sec. Task IT	1.84 (0.69)	2.00 (0.77)	2.20 (0.52)	1.92 (1.06)	1.59 (1.18)	2.00 (0.97)	1.50 (1.06)	1.91 (0.92)	1.86 (1.17)
Knowledge Test IT (0-24)	-	17.67 (3.82)	17.85 (1.60)	14.42 (3.84)	15.09 (3.35)	16.45 (3.36)	15.33 (4.07)	16.00 (3.28)	15.64 (4.29)
Prod. Outcome RA	719.32 (418.69)	1033.43 (196.69)	850.28 (294.05)	630.04 (420.30)	1025.40 (366.03)	968.44 (220.06)	390.03 (440.76)	641.85 (408.41)	665.72 (308.27)
Sec. Task RA	1.05 (0.85)	1.94 (0.54)	1.25 (0.79)	0.75 (1.03)	2.05 (0.99)	1.41 (1.01)	0.17 (0.38)	0.86 (1.04)	0.95 (1.05)
Knowledge Test RA (0-24)	-	-	-	13.67 (4.44)	16.05 (2.66)	16.64 (3.96)	12.75 (3.80)	16.82 (2.36)	15.32 (4.64)

Table 4. Results of repeated measures ANOVA for CG and EGs for studies 1-3

	Production Outcome (skill decay)	Secondary Task (mental workload)	Knowledge Retention
CGs (n=67)			
Time	$F(1,64)=81.67$, $p<.001$, $\eta^2_p=.561$	$F(1,64)=62.25$, $p<.001$, $\eta^2_p=.493$	$F(1,64)=17.81$, $p<.001$, $\eta^2_p=.279$
Group	$F(2,64)=4.33$, $p=.017$, $\eta^2_p=.119$	$F(1,64)=5.11$, $p=.009$, $\eta^2_p=.138$	n.s.
Time x Group	n.s.	n.s.	$F(1,46)=5.39$, $p=.025$, $\eta^2_p=.105$
EGs(n=126)			
Time	$F(1,120)=50.73$, $p<.001$, $\eta^2_p=.297$	$F(1,120)=29.56$, $p<.001$, $\eta^2_p=.198$	n.s.
Group	$F(5,120)=4.33$, $p=.001$, $\eta^2_p=.153$	n.s.	n.s.
Time x Group	$F(5,120)=5.93$, $p<.001$, $\eta^2_p=.198$	$F(5,120)=6.73$, $p<.001$, $\eta^2_p=.219$	n.s.

We also analysed the difference between the variance that occurs between all groups (*SD*) based on Bland (2000), as the variance differs strongly. Significant differences between the variances are displayed in Table 5. The performance of subjects at RA is most similar in the practice and skill demonstration groups.

Table 5. Significant differences in variances between CGs and EGs (Bland, 2000)

	Production Outcome (skill decay)	Secondary Task (mental workload)	Knowledge Retention
CGs vs. EGs	F=1.65, p=.008	n.s.	n.s.
CG1 vs.			
Practice (Study 1)	F=4.53, p=.003	n.s.	-
Symbolic Rehearsal (Study 1)	n.s.	n.s.	-
Practice (Study 2)	n.s.	n.s.	-
Skill Demonstration	F=3.62, p=.006	n.s.	-
Symbolic Rehearsal (Study 3)	n.s.	n.s.	-
Procedural Knowledge	n.s.	n.s.	-
CG2 vs.			
Practice (Study 1)	F=4.57, p=.002	F=3.64, p=.008	-
Symbolic Rehearsal (Study 1)	n.s.	n.s.	-
Practice (Study 2)	n.s.	n.s.	F=2.79, p=.021
Skill Demonstration	F=3.65, p=.004	n.s.	n.s.
Symbolic Rehearsal (Study 3)	n.s.	n.s.	F=3.54, p=.005
Procedural Knowledge	n.s.	n.s.	n.s.
CG3 vs.			
Practice (Study 1)	F=6.75, p<.001	n.s.	-
Symbolic Rehearsal (Study 1)	n.s.	F=4.32, p=.001	-
Practice (Study 2)	n.s.	F=6.79, p<.001	n.s.
Skill Demonstration	F=4.01, p=.002	F=7.06, p<.001	n.s.
Symbolic Rehearsal (Study 3)	n.s.	F=7.49, p<.001	F=2.59, p=.032
Procedural Knowledge	n.s.	F=7.64, p<.001	n.s.

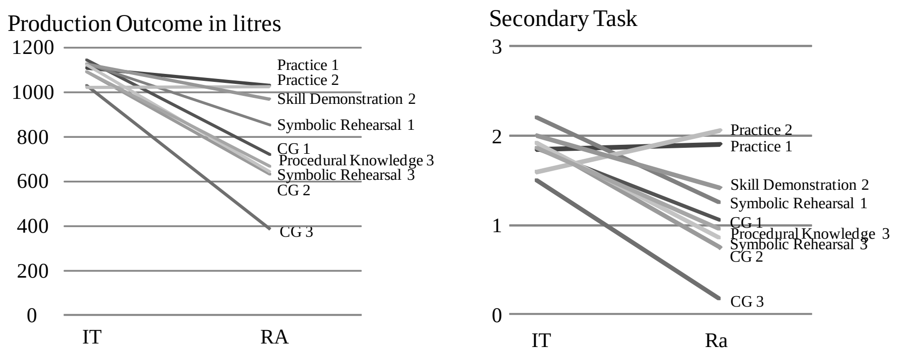


Figure 6. Production Outcome and Secondary Task performance (as indicator of mental workload for primary task) of EGs and CGs of Studies 1-3. Explanation of abbreviation: E.g., Practice 1 means Practice group from study 1

In addition, Pearson correlations for CGs showed significant correlations between GMA and prior knowledge, secondary task in RA, but no significant correlation between prior knowledge and performance were found (Table 6). For EGs, no significant correlations with person-related variables were found (Table 7).

*Table 6. Pearson Correlation between GMA, Prior Knowledge, Production Outcome and Secondary Task in RA for CG of Studies 1-3 (n=57), **p<.01, *p<.05*

	1	2	3	4
GMA(1)	-			
Prior Knowledge (2)	.236	-		
Prod. Outcome RA (3)	.155	.036	-	
Secondary Task RA (4)	.364**	.161	.373**	-
Knowledge Test RA (5)	.383**	.362*	.199	.385**

Results for the CGs show that GMA is significantly related to the mental workload at RA as well as to the amount of knowledge which is retrieved at RA. As the CGs received no RI, this means that the mental workload with which the primary task is performed also depends on the individual's mental abilities.

*Table 7. Pearson Correlation between GMA, Prior Knowledge, Production Outcome and Secondary Task in RA and 2 for EG of Studies 1-3 (n=122), **p<.01, *p<.05*

	1	2	3	4
GMA (1)	-			
Prior Knowledge (2)	.247**	-		
Prod. Outcome RA (3)	.084	-.041	-	
Secondary Task RA (4)	.134	.103	.382**	-
Knowledge Test RA (5)	.341**	.403**	.244*	.325**

Results for the EGs show that GMA affects knowledge retention only and has no effect on mental workload here.

Summary of studies 1-3

In summary, we found that skill decay and loss of retrieval strength has two consequences: loss of speed and accuracy of the performance of a fixed sequence procedure on task performance and the mental workload which accompanies it. Based on the observation that the variance in performance varies to a large extent, we assumed that person-related variables other than GMA may affect the retention of knowledge and skills; thus, in study 4, which is introduced next, we incorporated retentiveness.

Methods study 4

With regard to studies 1-3, criticism emerged that the generalisation of findings is limited because in a real production context, the operators would not be "doing nothing" between IT and a non-routine situation, but would be constantly interacting with the system. We therefore designed a study which simulated a form of "daily routine" for the operators. The study is reported in detail in Miebach (2013). The study conducted by Miebach (2013) addressed the impact of work experience (in which the plant is controlled in routine operation, in which the start-up procedure is not required) on skill retention and skill decay, respectively.

The skill to be maintained here is also the eleven-step fixed sequence as described in section "General Methods" (Table 1).

Experimental procedure of the study

In this study, a between- and within-group design was implemented with one experimental group (EG - work experience/routine operations) compared to a control group (CG) and two measurement times (Figure 3, Table 8).

The two measurement times were as follows: 1) performance at the end of the Initial Training (IT) and 2) performance at the Retention Assessment (RA) *three* weeks later. The EGs additionally received an intervention called Work Experience (WE), which consisted of controlling WaTrSim twice for 30 minutes between IT and RA. The WE took place one week and two weeks after the IT. The CG received only the IT and the RA, which are described in section "General Methods".

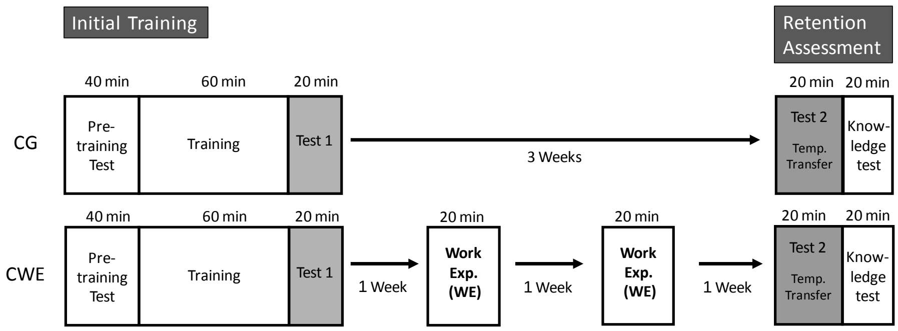


Figure 7. Procedure of study 4; the RA was conducted three weeks after the IT

The dependent variables are identical to studies 1-3

In this study, in addition to sex, age, GMA and prior knowledge, *retentiveness* was measured as a further control variable using subtest I for figural retention and subtest II for verbal retention of the I-S-T 2000R (Liepmann et al., 2007). The dimension of *I-S-T 2000R* measures retentiveness verbally and figurally. After one minute of memorising, the memorised words have to be attributed to presented hypernyms such as “The word with an initial letter B was: a) sport, b) food, c) city, d) job or e) building” (score 0-10). After another minute of memorising, one figure of the pair was presented and the related figure had to be picked: “Please find the right answer” (score 0-13, overall score 0-23).

Furthermore, *knowledge* about WaTrSim was measured with an adapted and extended version of the knowledge test applied in studies 1-3. The test consisted of 13 questions, e.g. “Is it correct that tank R1 has to be filled with at least 100 litres so that the heating HB1 can be turned on?” (score 0-47).

Table 8.Overview of experiment parts and variables.

Study 4 **Session Week 1** Initial Training (IT); 120 min	**Session Week 4** Retention Assessment (RA); 40 min
• Pretraining Test: - Socio-demographic data - General mental abilities - Retentiveness - Previous knowledge • Initial Training: - 2x Explore - 4x Start-up with manual • Test 1: - 4x Temporal transfer: Start-up without manual (*performance in fourth of four times was measured*) - Knowledge test	• Test 2: - Temporal transfer: Start-up - Knowledge test

Results of study 4

The descriptive statistics of the study are shown in Table 9.

CG: A repeated measure ANOVA showed a significant effect of time for the control group in production outcome ($F(1,19)=47.87$, $p<.001$, $\eta^2_p=.716$), in the secondary task ($F(1,19)=39.23$, $p<.001$, $\eta^2_p=.674$) and in knowledge test ($F(1,19)=11.33$, $p=.003$, $\eta^2_p=.374$). This means that substantial skill decay occurred if no WE took place.

EG: A repeated measure ANOVA showed a significant effect of time for the EG in production outcome ($F(1,19)=9.07$, $p=.007$, $\eta^2_p=.323$) but no effects of time in the secondary task (F(1,19)=3.71, p=.069, η^2_p=.164) and in knowledge test (F(1,19)=0.06, p=.805, η^2_p=.003). This means that the WE could not avoid skill decay, but did support the automaticity of skills (indicated by a reduced mental workload with which the production task is performed) and the retention of knowledge.

Table 5. Descriptive Statistics for CG and EG of Study 4; M (SD)

	CG 4 (*n*=20)	WE (*n*=20)
Sex	13 female	12 female
Age	27.05 (4.85)	20.25 (0.85)
Prior Knowledge (0-7)	5.28 (1.22)	5.65 (1.04)
GMA (0-50)	25.90 (5.22)	25.55 (4.67)
Retentiveness (0-23)	19.15 (4.17)	19.55 (3.17)
Prod. Outcome IT	1122.18 (216.63)	1036.51 (307.26)
Sec. Task IT	2.35 (0.67)	2.20 (1.01)
Knowledge Test IT (0-47)	35.55 (4.32)	33.20 (6.65)
Prod. Outcome RA	316.75 (429.03)	810.45 (391.69)
Sec. Task RA	0.75 (1.07)	1.65 (0.99)
Knowledge Test RA (0-47)	29.35 (8.38)	32.85 (5.67)

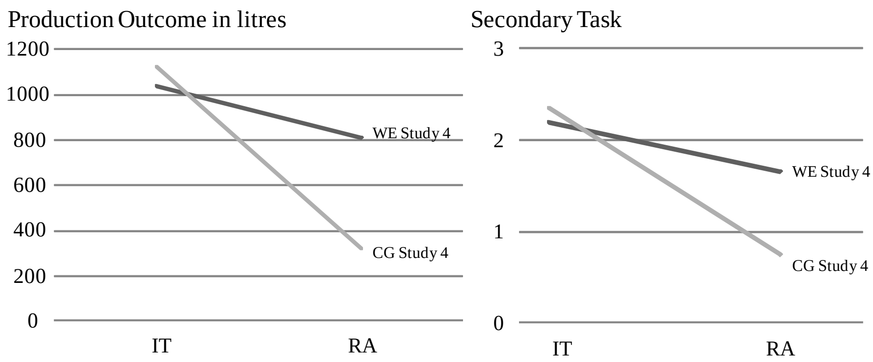


Figure 8. Production Outcome in Litres and secondary task results of WE and CG in IT and RA.

Here too, the difference between variance in performance between groups was analysed, but no significant differences were found (p>.05).

A Pearson Correlation between person-related variables and performance at RA showed no significant correlations either in the CG or in the EG (Tables 10 and 11). This means that in this study, neither GMA nor retentiveness had an impact on knowledge and skill retention.

*Table 6. Pearson Correlation between GMA, Prior Knowledge, Retentiveness, Production Outcome and Secondary Task in RA for CG of Study 4 (n=20), **p<.01, *p<.05*

	1	2	3	4	5
GMA(1)	-				
Prior Knowledge (2)	.318	-			
Retentiveness (3)	-.173	.052	-		
Prod. Outcome RA (4)	.149	.387	-.015	-	
Secondary Task RA (5)	.108	.399	.021	.923**	-
Knowledge Test RA (6)	.427	.397	-.176	.482*	.568**

*Table 11. Pearson Correlation between GMA, Prior Knowledge, Retentiveness, Production Outcome and Secondary Task in RA for EG of Study 4 (n=20), **p<.01, *p<.05*

	1	2	3	4	5
GMA(1)	-				
Prior Knowledge (2)	.280	-			
Retentiveness (3)	.021	-.034	-		
Prod. Outcome RA (4)	.162	.242	.351	-	
Secondary Task RA (5)	.158	-.074	.434	.072	-
Knowledge Test RA (6)	.357	.214	.145	.173	-.273

Discussion

The present paper shows that skill decay as well as the mental workload (with which a task is performed after a longer period of non-use) depends on non-use as well as on the method of refresher intervention (studies 1-3). In addition, it was shown that ordinary work experience also leads to skill decay in certain tasks, although some automaticity in handling the system is acquired (study 4). These results support the theoretical assumptions and empirical findings presented by McBride et al. (2013) and Onnasch et al. (2013). Regarding person-related variables, in studies 1-3, correlations between the person-related variable GMA and the secondary task was found for the CG. In both the EG and CG, a relationship between GMA and knowledge was shown. The fourth study showed no significant correlations, and in particular no correlations between dependent variables and retentiveness. These findings may be explained by the retentiveness test used (IST-2000R), which might not have corresponded to the type of retentiveness required in this study, because it addresses shorter retention intervals in particular. In a further study (Maafi, 2013), we therefore incorporated the Wilde-2 Test (Kersting et al., 2008), which measures retentiveness for longer periods of non-use, and found substantial and medium to large correlations between the retentiveness measured by the Wilde-2 and knowledge and skill retention.

Our studies aimed to show the effects of skill decay and possible countermeasures to support skill maintenance over longer periods of non-use. In summary, results show that different RI methods lead to different performance results and skill decay or maintenance. Even work experience cannot completely counter skill decay with respect to non-routine situations. Organisations should therefore carefully design possible interventions in order to maximise skill retention with a minimum of mental workload.

Acknowledgements

Studies 2 and 3 were funded by the Deutsche Forschungsgemeinschaft (DFG)/German Research Council, with the project number: KL2207/3-1.

References

Bainbridge, L. (1983). Ironies of automation. Increasing levels of automation can increase, rather than decrease, the problems of supporting the human operator. *Automatica, 19*, 775-779.

Bjork, R.A., & Bjork, E.L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. Healy, S. Kosslyn, and R. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ, USA: Erlbaum.

Bjork, R.A., & Bjork, E.L. (2006). Optimizing treatment and instruction: Implications of the new theory of disuse. In L.G. Nilsson and N. Ohta (Eds.), *Memory and Society, Psychological Perspectives* (pp. 109-134). Hove, UK: Psychology Press.

Bjork, R.A. (2009). Structuring the conditions of training to achieve elite performance: Reflections on elite training programs and related themes in Chapter 10-13. In K.A. Ericsson (Ed.), *Development of Professional Expertise. Toward Measurement of Expert Performance and Design of Optimal Learning Environments* (pp. 312-332). Cambridge: Cambridge University Press.

Bjork, R.A. (2011). On the symbiosis of learning, remembering, and forgetting. In A.S. Benjamin (Ed.), *Successful remembering and successful forgetting: a Festschrift in honour of Robert A. Bjork* (pp. 1-22). London, UK: Psychology Press.

Bland, M. (2000). *An introduction to medical statistics*, 3rd ed. Oxford: Oxford University Press.

Bourne, L.E., & Healy, A.F. (2012). Introduction, Training and its cognitive underpinnings. In A.F. Healy, and L.E. Bourne, Jr. (Eds.), *Training Cognition, Optimizing Efficiency, Durability, and Generalizability* (pp. 1-12). New York, NY: Psychology Press.

Burkolter, D., Kluge, A., German, S., & Grauel, B. (2009).Wastewater Treatment Simulation (WaTrSim): Validation of a new process control simulation tool for experimental training research. In *Proceedings of Human Factors and Ergonomics 2009* (pp. 1969-1973). San Francisco: Human Factors & Ergonomics Society.

Ebbinghaus, H. (1885*). Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Leipzig: Verlag von Duncker & Humber.

Kersting, M., Althoff, K., & und Jäger, A.O. (2008). *Wilde-Intelligenz-Test 2 (WIT-2).* Göttingen: Hogrefe.

Kim, J. W., F. E. Ritter, & R. J. Koubek. 2013. “An Integrated Theory of Skill Acquisition and Retention in the Three Stages of Learning.” *Theoretical Issues in Ergonomics Science* 14: 22-37.

Kluge, A., & Frank, B. (in press). Counteracting skill decay: Four refresher interventions and their effect on skill retention in a simulated process control task, *Ergonomics*.

Kluge, A., Burkolter, D., & Frank, B. (2012). “Being prepared for the infrequent”: A comparative study of two refresher training approaches and their effects on temporal and adaptive transfer in a process control task. In *Proceedings of Human Factors and Ergonomics 2012, 56* (pp. 2437-2441). Santa Monica: Human Factors & Ergonomics Society.

Liepmann, D., Beauducel, A., Brocke, B., & Amthauer, R. (2007). *IST 2000 R*. Göttingen: Hogrefe.

Martinussen, M., & D.R. Hunter (2010). *Aviation Psychology and Human Factor*s. Boca Raton: CRC Press.

Maafi, S. (2013). *Trainieren, merken, abrufen! Der Einfluss von personenspezifischer Variablen auf die Trainings- und Transfereffekte in der Prozesskontrolle [Train, remember, recall! The impact of person-related variables on training and transfer effects in process control]*. Master thesis, University of Duisburg-Essen, Duisburg, Germany: Chair of Business and Organisational Psychology.

McBride, S.E., Rogers, W.A., & Fisk, A.D. (2013). Understanding human management of automation failure. *Theoretical Issues in Ergonomics Science*. Online first 17 June 2013.

Miebach, J. (2013). *The impact of work experience on skill retention of complex work tasks in high risk working environments - an experimental study in a simulation environment*. Master thesis, University of Duisburg-Essen, Duisburg, Germany: Chair of Business and Organisational Psychology.

Onnasch, L., Wickens, C.D., Li, H., & Manzey, D. (2013/online first). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors*, published online, 20th August 2013.

Ormerod, T.C., & Shepherd, A. (2004). Using task analysis for information requirements specification: The sub-goal template (SGT) method. In D. Diaper, and N.A. Stanton (Eds.), *The handbook of task analysis for human-computer interaction* (pp. 347-365). Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2000). A Model for Types and Levels of Human Interaction with Automation. *IEEE Transactions on Systems, Man, and Cybernetics- Part A: Systems and Humans, 30*, 286-297.

Sauer, J., Hockey, G.R.J., & Wastell, D. (2000). Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43*, 2043-2064.

Wixted, J.T., & Carpenter, S.K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science, 18*, 133-134.

Wonderlic Inc. (2002). *Wonderlic Personnel Test*. Libertyville: Wonderlic Inc.

An Experimental Study on the Effect of Group Interaction on Creativity − Ideation novelty

Denis A. Coelho, Maria de Fátima Lanhoso Vieira
Universidade da Beira Interior
Portugal

Abstract

The effects of group dynamics on creativity are investigated, through a design briefing and the generation of concepts by two groups of industrial design students. Seventeen voluntarily participating students were randomly assigned to two conditions: a control group and a treatment group. The groups were given the same design brief and drawing materials, simultaneously in two classrooms. The treatment group was asked by the moderator, after 20 minutes of initial individual silent concept generation, to discuss through a structured discussion the design brief. At the same time the control group was asked by another moderator to write in a sheet of paper answers to the same questions that were given orally to the treatment group. This intermediate stage of the parallel simultaneous sessions lasted 15 minutes, at the beginning of which a different colour pen was exchanged for the one initially given to the subjects. Finally, the subjects in the two groups were given an additional 20 minutes to complete ideation silently and individually. The underlying hypothesis for this study is that group dynamics promotes increase in effectiveness of individual ideation. Results were analysed in terms of maximum novelty of individual ideation, showing a positive post-treatment effect, departing sharply from the novelty degradation in the control group in the second phase of the experiment. Future data analysis work contemplates analysis of the ideation outcomes of the experiment using a variety metric, in order to complement the findings obtained for novelty, in characterizing the effect of group interaction in creative ideation.

Introduction

Given the widespread belief that creativity is essential to compete in industry, research has focused on the process by which creative ideas are generated and on the processes that contributes to creative outcomes for individuals in organizations. Research on creativity has also focused on tangible creative outcomes (those ideas or products which meet the criteria of being novel and useful) as well as on each individual’s perception of the creativity occurring (Kurtzberg & Mueller, 2005).

Decades of creativity research have produced several major theories of creativity (e.g., Amabile, 1996; Ford, 1996; Simonton, 1999; Woodman, Sawyer, & Griffin, 1993), and each notes that some forms of interpersonal interaction can play a role in

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

the creative process. Hence, research on the development of original ideas that are useful or influential (creativity) has evolved from a perspective focused on individuals (Mumford & Gustafson, 1988; Sternberg & Lubart, 1999) to studying group creativity in organizations (West, 2000) and collaborative learning (Johnson, Johnson, & Smith, 2007) as well as micro-creativity in group processes (Chiu, 2008) within piecemeal problem solving.

Isaksen (1998) reviewed 40 years of empirical research on brainstorming, one of the most well-known tools for creative thinking. Seminal experimental studies performed by Taylor, Berry and Block (1958) suggest that group participation when using brainstorming inhibits creative thinking. Dunnette, Campbell and Kay (1963) found that brainstorming effectiveness of individuals is greater when they are brainstorming individually, than when brainstorming is done in groups. Moreover, the superiority of individual brainstorming over group brainstorming was found to be relatively higher when it was preceded by group participation. These authors concluded that group participation is accompanied by certain inhibitory influences even under conditions (e.g., brainstorming) which place a moratorium on all criticism. Isaksen (1998) noticed in the review that measurement of the effectiveness of idea generation should be improved in future studies comparing individual and group idea generation with the use of improved measures of the ideation outcomes.

Design research has developed a series of measures of ideation effectiveness, focusing on outcome based metrics, rather than process based metrics, due to the perceived difficulty in relating the occurrence of cognitive processes to the effectiveness of an idea generation method (e.g. Coelho & Versos, 2011). Shah et al. (2003) proposed four objective measures of ideation effectiveness for assessment of ideation outcomes in engineering design: novelty, variety, quality and quantity. The latter is defined simply as the total number of ideas generated by a group or individual during a designated amount of time. Generating several ideas increases the chances of occurrence of better ideas (Cross, 1996). The quality of an idea is an independent measure based on a physical property or ratio related to the performance of the artefact (e.g. time, weight, energy – Shah et al., 2003).

To measure novelty, the design problem is first decomposed into its key functions and characteristics. The idea produced is then analysed by first identifying which functions it satisfies and describing how it fulfils these functions at the conceptual and/or embodiment level (for a description see Pahl & Beitz, 1996). Each description is finally graded for novelty, which, combined by applying the weights to each function and stage, yields the novelty score for each idea. The novelty metric proposed by Shah et al. (2003) had also been used by psychologists to measure creativity (Jansson & Smith, 1991).

To measure variety (Shah et al., 2003), one examines how each function is satisfied. A variety rating applies to an entire group of ideas, not a single idea. Ideas are grouped based on how different two ideas are from each other. The use of a different physical principle to satisfy the same function makes two ideas very different. On the other hand, if two ideas differ only in some secondary construction level detail, say a dimension value, the ideas are only slightly different. Verhaegen et al. (2013) identified three shortcomings of the aforementioned variety metric, proposing a matching set of refinements. Their variety metric is more sensitive to changes in distribution of ideas over nodes on a given abstraction level in the genealogy structure, accounts for the degree of uniformness of the distribution of ideas and it exhibits a monotonically increasing behaviour from higher to lower abstraction levels.

Ideally, variety and novelty should be considered simultaneously, because novelty is concerned with how well the ideation outcomes expand the design space, while variety is concerned with how well the ideation outcomes explore the (more or less expanded) design space.

Aims

The aim of the experimental study reported in this paper is to ascertain whether the 50 year old findings of Dunnette, Campbell and Kay (1963) can be verified in actuality applied to ideation in industrial design, with ideation effectiveness assessed using more recently developed metrics. Novelty (Shah et al., 2003) was chosen as the metric to assess effectiveness of ideation. It is a measure of how unusual or unexpected an idea is compared to other ideas. The underlying overarching hypothesis for the study is that the effectiveness of individual ideation in industrial design is improved by group participation. In particular, ideation effectiveness is assessed in this paper, using the novelty metric presented by Shah et al. (2003), in an experimental comparative study that sought testing the following specific hypotheses:

- H1 - The individual maximum novelty scores of the ideation outcomes in the control group are lower than the ones in the treatment group, after the same amount of time following group-interaction in the treatment group.
- BH - Individual maximum novelty scores are not statistically different between the control and treatment groups prior to the group-interaction (baseline hypothesis) after the same amount of time of ideation.

A limitation of this paper is that it does not report on the variety scores (this analysis is still under preparation), which when applied to an idea set provides a measure of how much the design space has been explored, contemplating the variation brought by several ideas.

Method

Participants

The effect of group dynamics on creativity was investigated through a briefing and design concepts generation by two groups of Portuguese students of industrial design. Seventeen volunteer participants were randomly assigned to two conditions: a control group and a treatment group. Participants were rewarded with a 5 Euro gift card from a local department store.

Procedure

Both groups received the same design briefing (shown in appendix - Wilson et al., 2010) and the same drawing materials simultaneously in two classrooms. The briefing chosen was sufficiently open to foster the generation of novel and varied concepts and had been tested before, and was translated to Portuguese by the authors. The subjects in the treatment group were requested by the moderator, after 20 minutes of silent generation of individual initial concepts, to analyse the project briefing through a structured group discussion. At the same time, individuals in the control group were asked by the chairman to write down individual responses to the same questions that were verbalized orally to the treatment group, but they were only presented in written form to the control group. This intermediate step of the parallel simultaneous sessions lasted 15 minutes, at the beginning of which the pen which had been initially provided was exchanged for a different colour pen in both groups. Finally, the subjects in both groups had an additional 20 minutes to complete the ideation silently and individually.

Dependent Measures

The results were analysed using the approach presented by Shah et al. (2003), focusing on the key functions and characteristics of each idea to determine the effect of group interaction on the maximum score of individual ideation novelty by comparing the results in the two groups.

The total novelty of each idea may be calculated from equation (1).

$$M_1 = \sum_{j=1}^{m} f_j \sum_{k=1}^{n} S_{1jk} p_k \qquad (1) \text{ (Shah et al. 2003)}$$

M_1 - novelty metric value for an idea with *m* functions and *n* stages;

f_j - the weight that is given according to the importance of the function or essential characteristics;

n – number of stages of product development;

S_{1jk} - the various attributes of a particular function or essential feature of stage *k* of product development;

p_k - the weight given according to the importance of each product development stage.

Because in the experiment, only the concept generation stage was included, *n* equals 1 and p_k is given the value 1, simplifying equation (1) resulting in:

$$M_1 = \sum_{j=1}^{m} f_j S_{1j}$$

Calculation of S_1 is done afterwards, given the frequency of occurrence of the attribute in the universe of ideas for that function, using equation (2):

$$S_{1jk} = \frac{T_{jk} - C_{jk}}{T_{jk}} \times 10 \quad (2)$$

T_{jk} - total number of ideas generated for a function (*j*) for each stage (*k*);

C_{jk} - number of ideas generated for the current attribute a function (*j*) for each stage (*k*);

Multiplying by 10 is meant to standardize the result.

Hence S_1 depends on the universe of ideas produced for a certain function, in each phase of the experiment, and also depends on the number of ideas produced for the current attribute.

Each participant case was considered as a replication of the experiment in a shared context, considering that the novelty metric of ideation effectiveness is a measure of how much the limits of design space are pushed farther, expanding the design space, considering all ideation outcomes from all participants altogether to set a context for the relative novelty of each idea. As participants may have had more than one idea in each experimental phase, the participant's idea that achieved the highest novelty score is taken as the novelty performance for that participant in the particular condition. The novelty metric can only be applied to an entire group of ideas, and not to an isolated idea. Ideas were grouped based on how they were different from each other across clusters. For example, using a different physical principle to fulfil the same function makes two ideas very different, increasing the variety of the results of ideation. On the other hand, if two ideas differ only in a few details of construction or on a secondary level, such as a dimensional value (or an additional feature of the material), the ideas are only slightly different.

An example of one idea, which obtained a novelty score of 4.32 from equation (2), is shown in Fig.1. Table 1 shows the general function categories and the particular attribute analysis made for the idea shown in Figure 1.

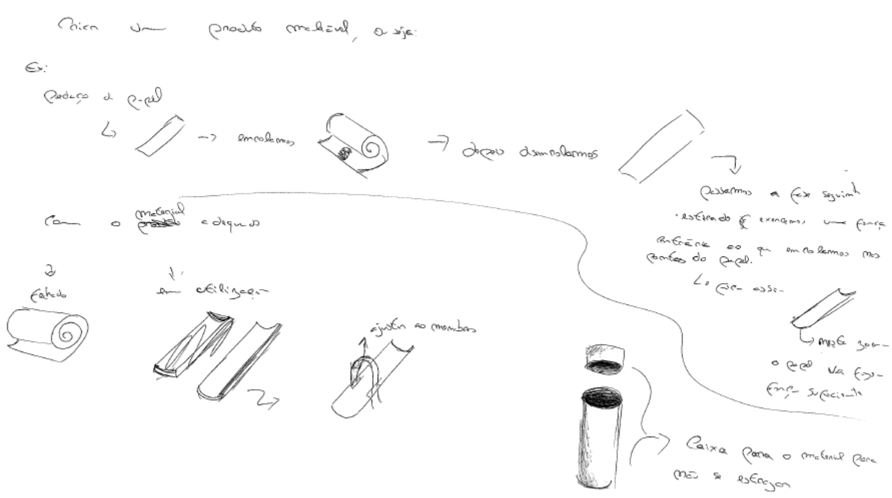


Figure 1. Example of an idea resulting from the experiment (participant in control group, second experimental phase- post-questionnaire)

Table 1. Analysis of the function attributes of the idea depicted in Figure 1 for the process of assessing its novelty score (the novelty score obtained for this idea was 4.32)

Function	*Attribute*
1. Variation in size for storage / immobilization	Curl
2. Treatment of injured member	Splint
3. Material used	Not very rigid
4. Comfort and health of patient / guide safeguarded	No

Results

The results were analysed statistically considering the maximum value of the novelty of each participant in the first and second part of the experiment. The novelty scores of the ideation outcomes are shown in Table 2, in the form of the mean and standard deviation of the maximum novelty score of the ideas generated per participant for the four experimental conditions (two groups in two experimental phases). With this arrangement, testing of hypothesis 1 and the baseline hypothesis ensued.

Table 2. Mean and standard deviation of the maximum idea novelty score per participant

Phase / Group	*Treatment (n=9)*	*Control (n=8)*
First (pre-questionnaire: 20 min)	*5.40 (1.00)*	*5.06 (1.14)*
Second (questionnaire and post-questionnaire: 15 + 20 min)	*5.49 (2.24)*	*3.99 (1.82)*

Due to small sample sizes, non-parametric statistics (Siegel & Castellan, 1988) are used. While the distribution of first phase maximum novelty is not statistically different across groups (Independent Samples Mann-Whitney U Test, U=42.5, p=0.541), the same does not apply for the second phase (U=59.5, p=0.021). The distribution of novelty scores is statistically different comparing the treatment and the control groups in the second phase of the experiment. The related samples Wilcoxon Sign Ranked Test was applied to the differences between phase one and phase two, with the following results: Treatment – W=17.0 , p=0.607; Control – W=4.0 , p=0.208. Focusing on the means of the first and second phase for both groups shows a decrease in the control group, but this effect is not statistically proven. After questionnaire administration, ideation novelty shows stability in the treatment group overall, while it is degraded on average in the control group, without achieving significance (p-value is greater than 0.05).

The total quantity of ideation outputs differed, albeit not in a statistically significant way, between both groups in both the pre-questionnaire phase and the second phase of the experiment (Table 3). Although randomly selected, the treatment group showed on average higher ideation productivity in terms of sheer idea quantity in both phases of the experiment compared to the control group. The average number of ideas generated per participant decreased 10.5% in the treatment group and 25% in the control group. Statistical tests applied to the data described in Table 3 did not show significant differences for both horizontal and vertical comparisons in the Table. However, the independent samples Mann-Whitney U Test applied to the number of ideas in the second phase across groups marginally approached significance with p=0.059, and U= 55.5.

Table 3. Mean and standard deviation of number of ideas generated per participant

Phase / Group	*Treatment (n=9)*	*Control (n=8)*
First (pre-questionnaire: 20 min)	*2.11 (0.78)*	*1.50 (0.53)*
Second (questionnaire and post-questionnaire: 15 + 20 min)	*1.89 (1.27)*	*1.13 (0.64)*

Analysis of the results reported in Tables 2 and 3 inspired a question: “Are the participants who achieve the highest novelty scores in each phase also the ones who are more prolific in terms of quantity of ideas generated?”. To answer this question the approach to correlations reported by Coelho et al. (2013) is applied resulting in the selection of Spearman’s rank order coefficient. As experimental conditions were virtually the same for the two groups of participants in the first phase of the experiment, the two groups are joined in one (n=17) yielding rho=0.066 (p=0.802) for the correlation between the maximum novelty score and the quantity of ideas generated per individual. Hence no correlation whatsoever is found for the first part of the experiment. For the second part of the experiment, groups are maintained separate in the correlation analysis, as conditions between the treatment and the control group were effectively different. Both groups display high correlation. The treatment group shows rho=0.852 (p=0.04; n=9) with significance, while the correlation in the control group yields rho=0.843 (p=0.09, n=8) approaching significance. Hence, while achieving high novelty scores in the first 20 minutes of the experiment did not depend on the quantity of ideas generated in that phase, in the second part of the experiment (involving 15 minutes with individual / group questioning and 20 minutes for silent ideation only), the more each participant in either condition generated ideas, the higher the maximum novelty score of the individual's ideation outcome.

Discussion

Statistical analysis of the novelty results shows that in the first part of the experiment the two groups were undifferentiated regarding the maximum novelty of participant's individual ideation set. The effect of the group dynamics in the treatment group appears to lead to maintaining in the second part of the experiment the average value of novelty reached in the first part of the experiment. In the control group, the average novelty per participant decreased, albeit not in a statistically significant way, from the first to the second phase of the experiment.

Comparing the first and second experimental phases, suggests that ideation quantity displays a tiredness effect over time. This effect may explain the fact that the idea novelty and quantity of the treatment group did not increase in the second phase, while the means in the control group decreased. Moreover, increased standard deviation in the second phase for the treatment group denotes increased variance over time, possibly because some subjects followed inspiration and detailed few ideas, while others kept exploring the solution space with new alternative but very conceptual solutions.

In future analysis of results, ideation outcomes in both groups (treatment and control) are to be analysed using the approach presented by Verhaegen et al. (2013), to determine the effect of group interaction on the degree of variety of ideation. To assess the variety of each set of ideas generated by each of the two groups (control and treatment) in the first and second parts of the experiment, an examination of how each function is satisfied in each idea is to be carried out. The experimental hypothesis is that the variety of ideation in each group, normalized with respect to the size of the group, is not different for the first part of the experiment in the control and treatment groups, differing however in the second part of the experiment. The second part of the experiment also aims to test if the variety of ideas of the treatment group, standardized for the size of the group, is higher than the variety of the ideas of the control group, or, if variety is lower for the treatment group than for the control group.

Once the overall results of the experimental study are extracted (combining the novelty and variety metric analyses) authors shall seek to determine if group dynamics influences the effectiveness of ideation evenly for the novelty and variety metrics. This analysis has been done both at the individual level for the novelty metric and is to be performed at the group scale only for the variety metric. Because of the relatively short duration of the phases of the experiment which did not give rise to many ideas per participant in each phase, with a few exceptions, the direction of the effect of group dynamics is only to be seen by comparing the variety results between both groups, each one viewed as a whole.

References

Amabile, T.M., Whitney, D., Weinstock, J., Miller, L., & Fallang, C. (1997). *What really happens in creative projects: Event sampling through electronic data collection*. Working Paper, Harvard Business School.

Chiu, M.M. (2008). Effects of argumentation on group micro-creativity: Statistical discourse analyses of algebra students' collaborative problem solving. *Contemporary Educational Psychology, 33,* 382-402.

Coelho, D.A., & Versos, C.A. (2011). A comparative analysis of six bionic design methods. *International Journal of Design Engineering*, *4*, 114-131.

Coelho, D.A., Harris-Adamson, C., Lima, T.M., Janowitz, I., & Rempel, D.M. (2013). Correlation between Different Hand Force Assessment Methods from an Epidemiological Study. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *23*, 128-139.

Cross, N. (1996). Creativity in design: not leaping but bridging. In *Creativity and Cognition 1996: Proceedings of the Second International Symposium*. Loughborough, LUTCHI.

Dunnette, M.D., Campbell, J., & Jaastad, K. (1963). The effect of group participation on brainstorming effectiveness for 2 industrial samples. *Journal of Applied Psychology*, *47*, 30-37.

Ford, C.M. (1996). A theory of individual creative action in multiple social domains. *Academy of Management Review, 21*, 1112-1142.

Isaksen, S.G. (1998). *A review of brainstorming research: Six critical issues for inquiry*. Creative Research Unit, Creative Problem Solving Group-Buffalo.

Jansson, D.G., & Smith, S.M. (1991). Design fixation. *Design* Studies, 12, 3-11.

Johnson, D., Johnson, R., Smith K. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, *19*, 15–29.

Kurtzberg, T.R., & Mueller, J.S. (2005). The influence of daily conflict on perceptions of creativity: A longitudinal study. *International Journal of Conflict Management*, *16*, 335-353.

Mumford, M.D. & Gustafson, S.B. (1988). Creativity syndrome: Integration, application and innovation. *Psychology Bulletin*, *103*, 27–43.

Pahl, G., Beitz, W. (1996). *Engineering Design – A systematic approach*, 2nd ed. [Translation], London: Springer Verlag.

Siegel, S., & Castellan, N.J.J.(1988). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Book Company, New York.

Simonton, D.K. (1999). *Origins of genius: Darwinian perspectives on creativity*. New York: Oxford University Press.

Shah, J.J., Smith, S.M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design studies*, *24*, 111-134.

Sternberg, R.J., Lubart, T.I. (1999). The concept of creativity. In R.J. Sternberg (Ed.), *Handbook of creativity*, Cambridge University Press, New York, pp. 3–15.

Taylor, D.W., Berry, P.C., & Block, C.H. (1958). Does group participation when using brainstorming facilitate or inhibit creative thinking? *Administrative Science Quarterly*, 23-47.

Verhaegen, P.A., Vandevenne, D., Peeters, J., & Duflou, J.R. (2013). Refinements to the variety metric for idea evaluation. *Design Studies*, *34*, 243-263

West, M.A. (2000). Reflexivity, revolution, and innovation in work teams. *Advances in Interdisciplinary Studies of Work Teams*, *5*, 1–29

Wilson, J.O., Rosen, D., Nelson, B.A., & Yen, J. (2010). The effects of biological examples in idea generation. *Design Studies*, *31*, 169-186.

Woodman, R.W, Sawyer, J.E. & Griffin, R.W. (1993). Toward a theory of organizational creativity. *Academy of Management Review, 18*, 293-321.

Appendix

Design Briefing (Wilson et al., 2010)

Mountain-TREK

Mountain-TREK (MTREK) is an outdoor wilderness company that organizes backpacking trips to the mountains throughout the year. During these trips, MTREK utilizes trip guides to lead a group of participants through these wilderness expeditions. For safety reasons, MTREK requires each of its guides to carry emergency kits containing an assortment of medical supplies. These kits contain items that can be used in the case of sickness, insect bites, wounds, trauma, etc. Due to the other items outfitted in the guides' packs, available space is limited. In extreme hiking conditions, MTREK has noticed a significant risk of leg and ankle dislocations and fractures.

Design challenge

Due to the potential for leg injuries, MTREK is now requiring guides to carry additional supplies to treat these injuries. In this design challenge, MTREK has hired you to design a device that can be used to immobilize a joint or limb in case of an extreme injury. This device must (1) be as light and small as possible when stored in the guides' packs but (2) rigid enough and large enough to immobilize the leg of an average-sized male.

Portuguese version (translated and adapted from Wilson et al., 2010)

Monte-TREK

A Monte-TREK (MTREK) é uma empresa de atividades ao ar livre que organiza expedições às montanhas ao longo de todo o ano. Nestas atividades, a MTREK recorre a guias de montanha que lideram o grupo de participantes nestas expedições. Por razões de segurança, a MTREK exige que cada um dos seus guias transporte consigo um conjunto de itens, para utilizar se se tornar necessário, que inclui um estojo de primeiros socorros com material para fazer curativos e alguns medicamentos. Este estojo contém itens que podem ser usados em caso de enjoo ou mal-estar, picada de insetos, feridas, traumatismo, etc. Estes itens limitam o espaço disponível na mochila dos guias. Em condições extremas de escalada, a MTREK notou que existe um risco significativo de ocorrência de fratura da perna e de deslocamento do tornozelo.

DESAFIO DE DESIGN – Geração de conceitos alternativos

Devido ao potencial acrescido de lesão e ferimento na perna e no tornozelo, a MTREK vai passar a exigir que os seus guias levem consigo itens adicionais para tratar este tipo de lesões e ferimentos. Neste desafio de design, a MTREK contratou-o(a) para criar um dispositivo que possa ser usado para imobilizar uma articulação ou um dos membros inferiores em caso de ocorrência de lesão ou ferimento extremo. Este dispositivo tem de ser tão leve e tão pequeno quanto possível quando armazenado na mochila dos guias, mas tão rígido e tão grande quanto necessário para imobilizar a perna de um homem adulto de estatura média.

The Effect of Task Set Instruction on Detection Response Task Performance

Antonia S. Conti, Carsten Dlugosch, & Klaus Bengler
Institute of Ergonomics, Technische Universität München, Garching, Germany

Abstract

Detection response tasks (DRTs) have been suggested as a way to measure online levels of cognitive workload. DRTs have been used in various applied settings under different task instructions. Especially in situations where multiple tasks are performed simultaneously (e.g. a driver operating a vehicle and using a navigation device), how someone allocates attention and other cognitive resources is also based on the given task instruction. Task instruction is important and can affect the strategy one takes on to complete a task, affecting the metric used to gauge task performance (e.g. reaction times). This, of course, can affect the results and the conclusions drawn based on these results; which, in an applicative setting, has real-life outcomes (e.g. law making, car manufacturers, etc.). As the DRT method is in the process of being standardized, the effect of various instruction types on task performance is to be investigated. The present paper reports on a between-group study where participants performed a triple-task scenario, involving a simulated driving task, under four different instructions. No difference in mean DRT RTs was found according to task instruction, suggesting that the complexity of a triple-task condition subjugates any influence that different task instructions could have on the performance metric.

Introduction

Varieties of detection response tasks (DRT) have been implemented for many years in driver distraction research (e.g., Bengler et al., 2012; Conti et al., 2012; Engström et al., 2005; Jahn et al., 2005; Merat & Jamson, 2008; Olsson & Burns, 2000; van Winsum et al., 1999). Over time, this method has matured and is currently being used to measure the attentive effects of cognitive load (Bruyas & Dumont, 2013; Conti et al., 2013; Engström et al., 2013; Harbluk et al., 2013; Young et al., 2013). The DRT presents continuous stimuli (visual, auditory, or tactile) to which a participant responds via button press. The performance metrics of the DRT, reaction times (RTs) and misses, can be interpreted as the degree to which the other tasks required cognitive or attentional resources. The DRT is used according to the secondary task method (Knowles, 1963; Ogden et al., 1979; De Waard, 1996; Wickens & Hollands, 2000) and is performed in addition to other tasks.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

As part of the progress of the DRT method, the DRT is being standardized under ISO/CD 17488. Essential to this process is the standardization of the instructions to be used. Especially because the DRT is performed in addition to other tasks, a standardized task set instruction is needed to be able to advise the participant on how to prioritize task performance when they are required to perform multiple tasks at a time.

Instructions for an experiment are typically standardized with the intent to optimally orient participants to an experiment. In addition to serving as an introduction and explanation, it is reasonable to consider that the specific instructions a participant receives can influence his or her task performance. However, previous research on this topic both support and oppose this.

For example, Haider and Frensch (1999) found that task performance metrics, reaction times (RTs) and errors, as well as the extent of information processing, were sensitive to different instructions (participants were told to optimize accuracy, speed, or a combination of the two). The authors concluded that this finding could be due to the setting of a response criterion based on the given instructions. Task instructions have also been observed to influence the mental representations of task responses, such that the instructed response labels (colours and location, in this case) in a dual-task setting affect the coding of responses and task performance (Wenke & Frensch, 2005). Trottier and Pratt (2005) found that saccadic RTs were reduced when the task instruction was to gather information on a scene (an instruction relevant to the function of a saccade), rather than to be as fast as possible. Another performance modulation due to instructions was reported by Ansorge and Newmann (2005), where priming effects were eliminated by instructing the prime as task irrelevant.

On the contrary, the effect of task instruction has also been reported to have a minimal effect task performance. In 2001, Levy and Pashler evaluated whether dual-task slowing could be altered through instructions that varied task priority. They found that manipulating task instructions did not alone suffice to eliminate dual-task interference. Additionally, the effect of task instructions seems to diminish for difficult tasks compared to easier ones (Parker & Barber, 1964). In this study, participants who received task motivating instructions performed significantly better on a simple digit task than those who did not receive such instructions. However, for more difficult memory and reasoning tasks, no such differences in performance were found. The authors concluded that difficult tasks are not easily influenced by instructions. Finally, in an article by Koch (2008), an investigation of the role of task instructions in task repetitions and switches is reported, where it was found that the task instruction effect was present in the former and not the latter.

The current article reports an experiment where the effect of task instruction on DRT RT performance was assessed. Whereas most previous research investigated the effect of task instruction on the performance of one or two tasks, this experiment required participants to perform the DRT in addition to two other tasks; namely, a simulated driving task and either a cognitive task (n-back) or a primarily visual-manual task (Surrogate Reference Task; SuRT). As the DRT is of particular interest in in-vehicle testing, the simulated driving task was included in the study. The secondary tasks were chosen based on their relevance in the current DRT standardization process; precisely because of this, experience with these tasks and their effects according to the DRT has already been established.

Methods

Design

This experiment used a 4 (task set instruction) x 6 (task set load) mixed-model design. The between-subjects variable was task set instruction and the within-subjects variable was task set. Four participant groups were differently instructed on how to perform the same task sets. Participants performed the head-mounted version of the DRT (HDRT) in addition to two other tasks. The HDRT was tested under 6 types of task set load, including 2 baselines: HDRT performed alone and HDRT performed with the simulated driving task; and 4 experimental trials with 2 secondary tasks, the n-back and SuRT, in 2 levels of difficulty: HDRT + simulated driving task + n-back (0), HDRT + simulated driving task + n-back (2), HDRT + simulated driving task + SuRT easy, and HDRT + simulated driving task + SuRT hard.

Participants

Participants were separated into four groups reflecting the number of task set instructions tested. A total of 79 participants were tested, 71 of which qualified for analysis (2 were excluded due to ambiguous license status; 3 due to red-green colour blindness; 3 because of a technical error occurring in the data recording phase). All participants held a valid driver’s license and reported to have normal or corrected-to-normal vision. Table 1 describes these participants in more detail.

Table 1. Participant group demographics (N = 71).

Participant Group	*N*	*Mean Age*	*Minimum Age (years)*	*Maximum Age (years)*	*n Male*	*n Female*	*Left Handed*	*Right Handed*
1	18	24	17	33	16	2	2	16
2	18	24	19	33	15	3	1	17
3	18	24	19	28	13	5	3	15
4	17	23	18	29	12	5	4	13

Apparatus

This experiment was conducted in a fixed simulator at the Institute of Ergonomics, Technische Universität München. The driver’s seat was centrally located behind an active steering wheel (reconfigurable active yolk from Wittenstein). A large LCD monitor displayed the simulated driving scene (SILAB; Veitshöchheim, Germany), in front of the driver. To the right of the driver, a separate screen displayed the visual-manual task (Surrogate Reference Task, SuRT).

Head-Mounted Detection Response Task (HDRT)

The DRT used in this experiment is a USB device developed at the Institute of Ergonomics. The HDRT was implemented. The HDRT is a construction with one red LED mounted to a baseball-type cap and placed on the participant's head. The LED was viewed at 18 cm (measured along the centre line of the hat's brim, starting from where the brim is joined to the cap). Signals (viz. LED turning on) were randomly presented every 4000 – 5000 ms (signal onset to onset) and remained on for 1000 ms or until button press. The response button was fastened to the left index finger with a Velcro strip and positioned to facilitate a response by pressing the button against the steering wheel. Participants were instructed to respond as quickly and accurately as possible to these signals via button press.

Simulated driving task

The driving-like task simulated a two-way highway with two lanes in each direction. This simulated driving task required participants to drive along the highway scene in the middle of the right-hand lane, maintaining a constant speed of 80 km/h. The vehicle was automatic and required no gear shifting.

Secondary Tasks: n-back and SuRT

The n-back and the SuRT were used in both easy and difficult variants (n levels 0 & 2; Mehler et al., 2009; and SuRT as per ISO/TS 14198:2012) to induce different levels of workload. The n-back task is a system-paced task where pre-recorded numbers are dictated to the participant and he or she is to repeat these numbers. The "n" indicates the number the participant needs to repeat: either the number just said, n = 0, or the number said two steps prior to the current number, n = 2. In this task, participants were dictated 20 numbers and responded according to the current "n" of the task. Participants were to be as accurate as possible and responses were given verbally and recorded. The SuRT task required participants to find a pop-out target circle (larger circle) in a display filled with distracters (smaller, same sized circles). Task difficulty increased as the difference between the target and distracters was less prominent. Participants were to perform this task as quickly and accurately as possible. Navigation through and responses to the SuRT were performed via a numeric keypad located under the screen on the participant's right side.

Task set instructions

In the current experiment, a distinction between task instruction and task set instruction is made: task instructions indicated how to perform each individual task (given to all participants); task set instructions specified how to prioritize the performance of each task within the task set (given to all but one group, group 4 in Table 2). Each participant group (4 groups in total) was given different task set instructions with the goal of explicitly directing participants' attentive and cognitive resources. Table 2 presents the priorities specified by the task set instructions per participant group. These instructions were derived based on previous DRT literature and reasonable permutation. For safety reasons, the simulated driving task was always the primary priority for task set instructed groups.

Table 2. Instructed task priority per participant group.

<table>
<tr><th>Participant Group</th><th>First Priority</th><th>Second Priority</th><th>Third Priority</th></tr>
<tr><td>1</td><td rowspan="3">Simulated driving task</td><td>DRT</td><td>Secondary task</td></tr>
<tr><td>2</td><td colspan="2">No priorities specified</td></tr>
<tr><td>3</td><td>Secondary task</td><td>DRT</td></tr>
<tr><td>4</td><td colspan="3">No priorities specified</td></tr>
</table>

Procedure

At the beginning of the experimental session, participants watched a standardized multimedia instructional presentation that introduced and explained how to perform each task (viz. HDRT, simulated driving task, n-back, and SuRT) and how to prioritize the whole task set (ex. Driving + HDRT + either n-back or SuRT). Participants were able to ask questions and practice the tasks until they felt comfortable. Participants then performed the experiment, beginning with either a baseline or experimental block. A synchronization program ran all task software. Each trial began with allowing the participant to adjust their lane position and speed to 80 km/h. After performance on the simulated driving task was stable, the HDRT and secondary task began to run, respectively, within seconds of each other. Trials were recorded for 1 minute and included the simultaneous performance of all three tasks. Performance on the simulated driving-task and secondary tasks were also recorded, however, as the focus of the current article is on the DRT RT, other task performances are not discussed.

Dependent variables and hypotheses

HDRT RTs (ms) to HDRT signals were measured. Signal responses were classified as a hit between 100 ms and 2000 ms post-signal onset. Only the HDRT RTs of hits qualified for analysis. The RT value for a task set or baseline was calculated as the mean HDRT RT during this 1 minute segment. Means and standard deviations were then calculated across participants per task or baseline. The authors hypothesized that both task set load and task set instructions would both affect mean HDRT RTs. It was expected that task set instructions prioritizing the HDRT would yield faster mean RTs than those instructions where the HDRT was given a lower priority. For groups where task priority was less restricted (viz. group 2 and 4), similar DRT performance for these groups was expected due to a possible ingrained ascription of importance for the driving task. Additionally, it was hypothesized that task set would affect mean HDRT RTs, showing the sensitivity of the HDRT to different task loads.

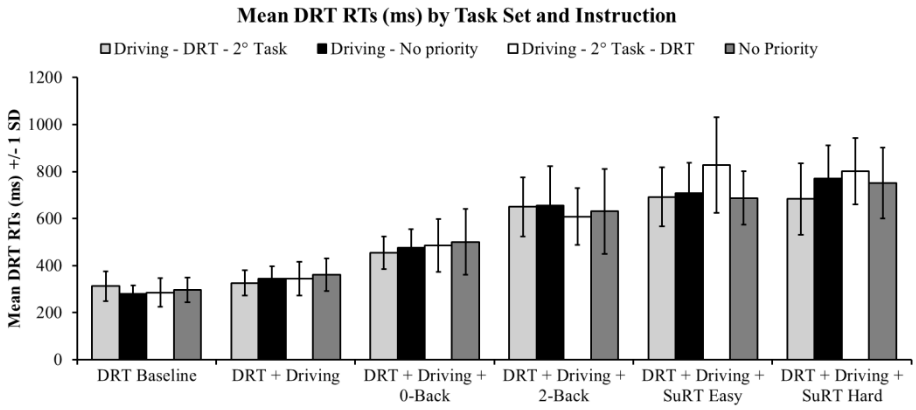


Figure 1. Mean HDRT RT (ms) +/- 1 SD per task set per task set instruction group. "DRT Baseline" indicates the performance of the HDRT as a single task.

Results

In Figure 1, the mean HDRT RTs (ms) across conditions per task set and instruction can be found. For all conditions, the hit rates were above 70%. It can be seen that in the baseline trials, mean HDRT RTs were overall faster than in the experimental trials. HDRT RT values, means (M) and standard deviations (SD), can be found in Table 3.

Data was analysed using a 4 (task set instruction) x 6 (task set load) mixed design ANOVA. Mauchly´s test indicated that the assumption of sphericity had been violated for task set load, $\chi 2$ (14) = 131.95, p < .01. Therefore the Greenhouse-Geisser correction (ε = .68) was used. No significant main effect of task set instruction was found on mean HDRT RTs.

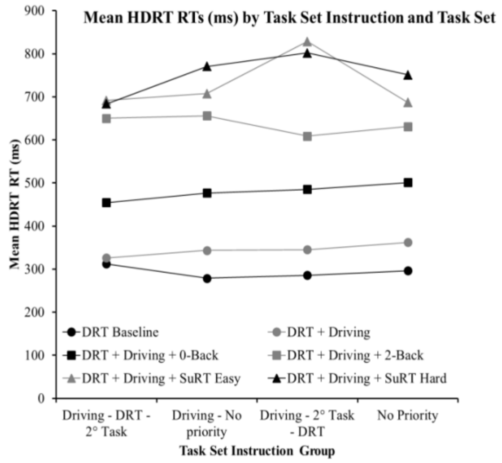


Figure 2. HDRT performance changes across task set instruction. Mean HDRT RT (ms) in each task set, except for the SuRT, was similar across task set instruction groups.

A significant interaction between task set load and task set instruction was found, $F(10.23, 228.46) = 2.74$, $p < .01$, $\eta_p^2 = .11$, meaning that depending on the task set load, the task set instruction had a different effect on mean HDRT RTs (see Figure 2). Simple effect tests, using Bonferroni adjusted alpha levels, indicated that for participant group 1, mean HDRT RTs did not significantly differ between the two baselines, DRT baseline and DRT + driving. Additionally, in groups 1, 2, and 4, mean HDRT RTs were not significantly different across task sets with 2-Back and SuRT in both difficulty levels. In group 3, no significant difference was found in sets with SuRT, easy and hard. All other comparisons were significant, $p < .01$. Pairwise comparisons were also carried out for each task set across participant groups. Task sets always yielded similar HDRT performance despite task instruction except for the task set with SuRT easy, under which mean HDRT RTs were significantly different between participant group 1 and 3, and 3 and 4, $p < .05$.

In terms of main effects, task set load was found as significant, $F(3.4, 228.46) = 312.43$, $p < .01$, $\eta_p^2 = .83$, meaning that the HDRT was sensitive to load induced by the different task sets. Specifically, the Bonferroni post hoc test revealed significant differences between HDRT RTs in all conditions, $p < .01$, except for the two task sets containing SuRT Easy and SuRT Hard.

Table 3. For each participant group, the specific task instruction is indicated as well as the mean (M; upper value) HDRT RTs (ms) and SDs (lower value).

	Task Instruction	Baselines		DRT + Driving +			
		DRT Alone	DRT + Driving	n-Back (0)	n-Back (2)	SuRT Easy	SuRT Hard
1	Driving+DRT+2°Task	312.32	326.32	454.67	650.33	691.52	683.67
		62.74	53.88	68.86	125.50	125.46	151.63
2	Driving+No Priorities	278.80	343.77	476.67	655.90	707.66	770.80
		36.85	52.79	79.39	167.98	129.13	139.86
3	Driving+2°Task+DRT	285.85	345.08	485.08	608.77	828.21	801.84
		60.18	71.93	112.07	120.22	203.08	140.86
4	No Priorities	296.67	362.20	501.18	631.02	687.03	750.77
		53.35	69.19	139.84	180.36	113.30	150.94

Discussion

The aim of the current experiment was to investigate whether different task set instructions affected mean DRT RTs. The HDRT was used for this evaluation, as were a simulated driving task and two additional secondary tasks: n-back and SuRT, in two levels of difficulty. It was hypothesized that in addition to the task set load affecting the DRT metric, task set instructions prioritizing the HDRT would yield faster RTs from participants relative to those instructions where the HDRT had a lower priority. Additionally, when no task priority was instructed, similar DRT performance for these groups was expected. Although task set load did affect mean DRT RTs, no significant effect of task instruction was found on mean HDRT RTs.

Generally, the data trends indicate that for all task sets with lower load (i.e. easier task variants), except for the SuRT, faster HDRT RTs were yielded. These HDRT RTs increase with task demand. A significant interaction between task set instruction and task set was found and revealed that mean HDRT RTs differed according to both task set load and instruction. By visually inspecting the graphs in Figure 1 and 2, it can be seen that no systematic influence seems to affect the data. The source of this interaction stems from the SuRT task. According to the HDRT RTs, SuRT easy did not always lead to faster RTs than for SuRT hard, in fact, sometimes this relationship was reversed. Since the DRT method proposes to measure the attentional effects of cognitive workload, it is not surprising that such effects are found for a primarily visual-manual task, which differs more in visual demand than it does in cognitive (Young et al., 2013). Additionally, because the SuRT is self-paced, participants could have used a different performance strategy relative to other tasks, which were system paced.

The significant main effect of task set load shows that the HDRT was sensitive to the changes in load corresponding with each task set. In accordance with this finding, it can be assumed that the participants were, in fact, performing the tasks in all conditions, else, no significant difference would be expected. As discussed above, the differences in HDRT performance during the SuRT easy and difficult was not significant.

The main effect of task instruction was not significant, implying that the task instructions used in this experiment did not affect mean HDRT RTs. At this time it cannot be confirmed whether task instruction, therefore, has no effect on the performance of a task set involving multiple tasks. First, an in depth evaluation of the other performance metrics, including an analysis of the performance strategies used in solving the tasks, needs to be carried out. It is possible that the difference in task performance manifests in another performance metric. The evaluation of the additional task performances included in this study is currently on-going.

In terms of the current experiment and the manipulation used, the mean HDRT RTs were not significantly affected by altering task instructions. Alternative hypotheses for this occurrence could be that because participants had to perform multiple tasks at once, the task set was too complex and participants were not able to accurately attend and perform as instructed (supported by Parker et al., 1964). As there were many tasks included within this experiment, it is also possible that some sort of task switching effect affected the performance metric, similar to the finding described by Koch (2008). The exact dynamic present here needs to be further investigated. Future investigations should also examine if the way the DRT and other tasks were technically operated (see Methods, each trial began with the driving simulation, then the DRT and secondary task started within seconds of each other) affects the performance outcome. This could indicate that giving different task instructions is not enough to alter task performance (supported by Levy et al., 2001) and that participants were not able to or did not adapt their strategy to the instructions given. An additional subjective questionnaire on how participants felt they were able to follow the instructions given could help clarify any inexplicable trends in the data. Future reproductions of this study could also include additional incentives for participants to ensure that the tasks are performed as close to the instructed priority as possible.

Conclusion

In terms of the current experiment, it has been demonstrated that task set instruction does not significantly affect mean HDRT RTs. Further evaluation is required to definitively confirm that task set instructions, therefore, have no measureable effect on task performance in a multiple task setting.

Acknowledgements

The authors would like to thank B. Helfer for his assistance in collecting the data, and A. Eichinger and M. Körber for their experimental and statistical insight.

References

Ansorge, U., & Neumann, O. (2005). Intentions Determine the Effect of Invisible Metacontrast-Masked Primes: Evidence for Top-Down Contingencies in a Peripheral Cuing Task. *Journal of experimental psychology. Human perception and performance*, *31*, 762–77.

Bengler, K., Kohlmann, M., & Lange, C. (2012). Assessment of cognitive workload of in-vehicle systems using a visual peripheral and tactile detection task setting. In: Work: A Journal of Prevention, Assessment and Rehabilitation 41 (Supplement 1/2012), S. 4919–4923.

Bruyas, M.P., & Dumont, L. (2013). Sensitivity of Detection Response Task (DRT) to the driving demand and task difficulty. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 64–70). Iowa City, USA: University of Iowa.

Conti, A.S., Dlugosch, C., Schwartz, F., & Bengler, K. (2013). Driving and speaking: revelations by the head-mounted detection response task. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 362–368). Iowa City, USA: University of Iowa.

Conti, A.S., Dlugosch, C., Vilimek, R., Keinath, A., & Bengler, K. (2012). An assessment of cognitive workload using detection response tasks. In *Advances in Human Aspects of Road and Rail Transportation* (pp. 735–743). CRC Press.

De Waard, D. (1996). *The Measurement of Drivers' Mental Workload*. (Doctoral Thesis). University of Groningen. Haren, The Netherlands: University of Groningen, Traffic Research Centre.

Engström, J., Åberg, N., Johansson, E., & Hammarbäck, J. (2005). Comparison between visual and tactile signal detection tasks applied to the safety assessment of in-vehicle information systems. In *Third International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 232–239). Iowa City, USA: University of Iowa.

Engström, J., Larsson, P., & Larsson, C. (2013). Comparison of static and driving simulator venues for the tactile detection response task. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 369–375). Iowa City, USA: University of Iowa.

Merat, N., & Jamson, A.H. (2008). The Effect of Stimulus Modality on Signal Detection : Implications for Assessing the Safety of In-Vehicle Technology. *Human Factors*, *50*, 145–158.

Haider, H., & Frensch, P.A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied*, *5*, 129–151.

Harbluk, J.L., Burns, P.C., Hernandez, S., Tam, J. & Glazduri, V. (2013). Detection response tasks: Using remote, headmounted and Tactile signals to assess cognitive demand while driving. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 78–84). Iowa City, USA: University of Iowa.

ISO/CD 17488: Road vehicles - Transport information and control systems – Detection-Response Task (DRT) for assessing selective attention in driving, 2013-09, International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/catalogue_detail.htm?csnumber=59887

ISO/TS 14198:2012 Road vehicles -- Ergonomic aspects of transport information and control systems -- Calibration tasks for methods which assess driver demand due to the use of in-vehicle systems, 2012, International Organization for Standardization, Geneva, Switzerland. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=54496

Jahn, G., Oehme, A., Krems, J., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*, 255–275.

Knowles, W.B. (1963). Operator Loading Tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *5*, 155–161.

Koch, I. (2008). Instruction effects in task switching. *Psychonomic Bulletin & Review*, *15*, 448–452.

Levy, J., & Pashler, H. (2001). Is dual-task slowing instruction dependent? *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 862–869.

Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. *Transportation Research Record: Journal of the Transportation Research Board*, *2138*, 6–12.

Ogden, G.D., Levine, J.M., & Eisner, E.J. (1979). Measurement of Workload by Secondary Tasks. *Human Factors : The Journal of the Human Factors and Ergonomics Society*, *21*, 529–548.

Olsson, S., & Burns, P. C. (2000). Measuring driver visual distraction with a peripheral detection task. Retrieved December 13, 2013, from http://www-nrd.nhtsa.dot.gov/departments/Human Factors/driver-distraction/PDF/6.PDF

Trottier, L., & Pratt, J. (2005). Visual processing of targets can reduce saccadic latencies. *Vision research*, *45*, 1349–54.

Van Winsum, W., Martens, M. H., & Herland, L. (1999). *The effects of speech versus tactile driver support messages on workload, driver behaviour and user acceptance* (TNO-Report No. TM-99-C043). Soesterberg, Netherlands: TNO Human Factors.

Wenke, D., & Frensch, P.A. (2005). The Influence of Task Instruction on Action Coding: Constraint Setting or Direct Coding? *Journal of experimental psychology. Human perception and performance*, *31*, 803–19.

Wickens, C.D., & Hollands, J.G. (2000). Attention, Time-Sharing, and Workload. In N. Roberts & B. Webber (Eds.), *Engineering Psychology and Human Performance* (Third Edit., pp. 439–479). Upper Saddle River, New Jersey: Prentice-Hall Inc.

Young, R.A., Hsieh, L., & Seaman, S. (2013). The Tactile Detection Response Task: Preliminary validation for measuring the attentional effects of cognitive load. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design* (pp. 71–77). Iowa City, USA: University of Iowa.

Visual search strategies of child-pedestrians in road crossing tasks

Hagai Tapiro, Anat Meir, Yisrael Parmet, & Tal Oron-Gilad
Ben-Gurion University of the Negev, Israel

Abstract

Children are overrepresented in road accidents, often due to their limited ability to perform well in road crossing tasks. The present study examined children's visual search strategies in hazardous road-crossing situations. A sample of 33 young participants (ages 7-13) and 21 adults observed 18 different road-crossing scenarios in a 180° dome shaped mixed reality simulator. Gaze data was collected while participants made the crossing decisions. That was used to characterize their visual scanning strategies. Results showed that age group, limited field of view, and the presence of moving vehicles affect the way pedestrians allocate their attention in the scene. Therefore, we can deduce that adults tend to spend relatively more time in further peripheral areas of interest than younger pedestrians do. It was also found that the oldest child age group (11-13) demonstrated more resemblance to the adults in their visual scanning strategy, which can indicate a learning process that originates from gaining experience and maturation. Characterization of child pedestrian eye movements can be used to determine readiness for independence as pedestrians. The results of this study emphasize the differences among age groups in terms of visual scanning. This information can contribute to promote awareness and training programs.

Introduction

Over a third of road traffic deaths in low and middle-income countries are among pedestrians. Even in more developed countries the rates of pedestrian deaths do not fall much behind; in Europe and in the Americas, the pedestrian death rate is 27% and 23%, respectively (WHO, 2013). Accident statistics consistently show that children are overrepresented as a group in pedestrian accidents. Children aged 15 and younger are accounted for 7% of the pedestrian fatalities in 2009 and 25% of all pedestrians injures in traffic crashes (NHTSA, 2009). The NHTSA report revealed that pedestrians are fully or partially accounted for most of the accidents, and that children are more likely than adults to be the cause of the accident. In an earlier study, pedestrians under the age of 10 were found to be responsible for 77% of the motor-pedestrian accidents, children aged 10 to 14 were the guilty party in 60% of the accidents, and those aged 15 or older were to blame in 33% of the accidents (Hunter et al., 1996). Thus, children more than adults, are at risk as pedestrians, often due to their own actions.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Much of the success in making the crossing decision relies on the pedestrian's ability to focus attention on relevant elements in the traffic environment and to ignore irrelevant stimuli. It has been recognized in several studies that pedestrians' selective attention abilities are linked to their performance in road crossing tasks (Dunbar et al., 2001; Foot et al., 1999; Tabibi & Pfeffer, 2007). It was also demonstrated that children have limited attention abilities compared to adults, and that those improve with age developing during middle childhood (Akhtar & Enns, 1989; Lin, et al., 1999). Overall, the literature suggests that due to their inferior attention abilities, young children are less competent to act in traffic than older children and adults (Dunbar et al., 2001; Tabibi & Pfeffer, 2007; Whitebread & Neilson, 2000).

Children suffer from inadequate visual searching behaviour that exemplifies the importance of efficient visual search strategy for safe road-crossing. Early literature review revealed that while the proportion of children that failed to look before stepping down to the road was significant but inconsistent, age was always an effecting factor (van der Molen, 1981). More recent studies estimate that over 50% of the children are not looking at the traffic before stepping into the road and the younger they are, the less likely they do so. When looking does occur it is often restricted to a single observation, often in the wrong direction (Zeedyk et al., 2002). Analysis of accident data demonstrates the consequences of this behaviour, and thus it is estimated that 39% of child-pedestrian victims do not look for approaching vehicles at all before crossing (Grayson, 1975). Nevertheless, it is not just that children do not 'look'. Studies have shown cases where children that have looked in the right direction still failed to detect the approaching vehicle. For example, Grayson (1975) estimated that in 31% of the cases, children had 'looked but failed to see' the hitting vehicle. This kind of failure might be difficult to explain, as it is reasonable to assume that when looking at an object directly one will 'see' it and respond accordingly. However, this phenomenon is not so surprising. For example, one can stare at the right direction but attend to something else like sounds or thoughts, a phenomenon known as the "inattentional blindness" (Mack, 2003). It is also possible that one is not interpreting the scene correctly or quickly enough. Just like in the case of inexperienced drivers, that delayed in processing the representation of a still-image traffic scene when a potentially hazardous object is present in it, but able to fixate upon the object just as fast as an experienced driver (Huestegge et al., 2010). In either way, no matter what cause children to fail spotting the approaching vehicle the results of the above studies strongly suggest that gaze-related behaviour is part of the reason for child-motor accidents.

Children's search process is affected by conspicuous parts in the visual field, and with age, this process becomes more systematic, exhaustive, focused and rapid (Day, 1975). Other studies exemplify additional aspects of failures in children's visual search; children before age 9 have limited ability to adapt their visual search strategy to fit different tasks (Hall, 1985), and young children's visual attention is more adversely affected by the presence of distractors than older children and adults (Pastò & Burack, 1997). Age-related differences also exist in a more simple aspects of eye-movements; six-year-old children make shorter fixations, and shorter and more rapid eye movements in comparison to adults (Mackworth & Bruner, 1970).

Whitebread and Neilson (2000) examined the relationships between pedestrian skills and visual search strategies. According to their findings, major changes in strategy occurred around the age of 7-8 years; this change expressed in the frequency and pattern of looking at different directions, having a sophisticated 'last-minute' checking approach, exhaustive visual search strategy, and the speed of making the crossing decision.

The work of Whitebread and Neilson (2000) justifies the need to further examine child pedestrian visual search behaviour among age groups. Meir et al. (2013) contrasted the crossing behaviour of three different child pedestrian age groups and an adults group. The current work adds the layer of visual search behaviour to their results. The aim of the current study is to get better understanding of the pedestrian visual behaviour and visual attention distribution in the environment in various roadside situations, with regard to the age of the pedestrian. The research hypothesis was that age has a direct link to visual behaviour and attention distribution; meaning that the older the children are they will show more resemblance to the experienced-adults group.

Method

Participants

Fifty-four participants were recruited, 21 experienced-adult participants aged 20-27 (mean age=25.3, SD=1.8), fourteen 7-to-8 year-olds (mean age=7.8, SD=0.7) eighteen 9-to-10-year-olds (mean age=9.6, SD=0.3) and nine 11-to-13 year-olds (mean age=11.5, SD=0.9). Children completed the experiment in exchange for an educational compensation equivalent of 30 NIS (approx. $10). Adults received the monetary compensation or bonus credit in an introductory course. Participants signed an informed consent form. Parental consent was given for participants under the age of 18.

Eye tracker and Dome facility

The head mounted ASL eye tracking system HS-H6 was used to perform correct measurement of pupil diameter and gaze direction (see Figure 1). After a short 9-points calibration process, the eye tracker accurately tracks the subject's eye movements and gaze direction by sampling the eye at 60Hz. The HS-H6 head mounted eye tracker was designed to track gaze direction over approximately a 30-35 degree vertical visual angle and a 40-45 degree horizontal visual angle, with 0.5 degrees precision and a 0.1 degrees resolution. The head mounted helmet is equipped with a front camera recording the same scene viewed by the participant. The video from the scene camera was recorded with an overlay of the participant's gaze location (Figure 2).

The 3D Perception™ Dome consists of a 180 degrees cylindrical screen (radius of 3.5 meters) aligned with a very accurate projection system of three projectors. This setting allows measurement of the participants when watching pre-designed simulated scenarios of real life situation from the roadside environment without the risk of harm (Figure 3).

Figure 1. The ASL Head-Mounted HS-H6 Eye Tracking System.

Figure 2. Capture from the scene video with the gaze location layer marked by a black cross.

Procedure

Participants arrived at the Dome facility for an hour-long session. Participants went through a stage of eye calibration, after which their eye-movements were recorded via the eye tracker. Each participant observed the 18 scenarios in a random order and engaged in a crossing decision by pressing a response button each time quickly when it was safe to cross the road. Two practice scenarios were used to familiarize with the experimental task. For each scenario and for each participant a video with the superimposed eye gaze data was saved on file using the scenario and the participant number as identifier. At the end of the session, participants were asked to fill a computerized demographic questionnaire, then received the compensation and left.

Stimuli

Eighteen typical urban simulated scenarios, each one lasting 10-45 seconds, shown from a pedestrian’s point of view, as seen in Figure 4. The scenarios consisted of a structured combination of elements: (1) Traffic Movement (no moving vehicles, one-way street where moving vehicles are traveling in one direction, two-way street where moving vehicles are traveling in two direction), (2) obscured Field of View (unrestricted, partially obscured by the road’s curvature on the left, or partially obscured by parked vehicles, see Fig. 4), and (3) Presence of Zebra-crossing (with zebra-crossing, without zebra-crossing), aiming to decipher participants’ responses to those situations.

Eye-data collection

A fixation happens when the eye focuses on a specific detail, and during this time, the eye collects information. Fixation duration varies and is estimated to be between 150-to-600ms, as 100ms is the minimum time required for the eye to receive information. In order to collect the fixation data from the videos, a manual coding method was applied. A trained research assistant did the manual process of fixation-identification.

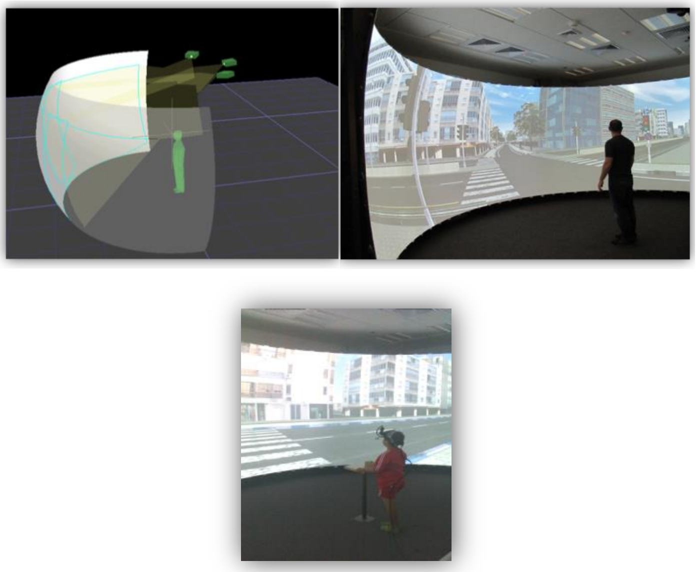


Figure 3. Dome projection facility at BGU Ergonomics complex.

In each frame in the video, a dynamic black cross was observed symbolising the participant gaze location in the scene at that time. The pinnacle digital video capture card recording at 25 FPS. Fixation was defined as a sequence of at least three consecutive video frames (120ms) during which the participant gaze's illustrated by the black cross remained focused on a specific detail in the scene. The manual coder watched each video frame-by-frame and manually documented each fixation (only those that fell in one of the areas of interest) in an Excel table, registering fixation duration and location.

Areas of Interest (AOIs)

For each scenario, five areas of interest (AOIs) were defined (Figure 5). The close range central area was defined as the 10 meters of road in each side from the pedestrian's point of view (AOI 3). Then symmetrically areas to the right of the centre and to the left were defined. The middle right/left range (AOIs 2/4) was the part of the road distant at least 10 meter to the right/left of the point of view but less than 100 meters away. The far right/left range (AOIs 1/5) was the part of the road at least 100 meter or more to the right/left of the pedestrian point of view.

Measures of visual search behaviour

Gaze distribution [%]: This is the distribution of fixation-time, that was directed by each participant in each scenario, between the defined AOIs.

Fixation duration [milliseconds]: The mean duration of fixations made by each participant in each scenario.

Dwells [seconds]: The mean duration of dwells made by each participant in each scenario. A single dwell is defined as the time during which a contiguous series of one or more fixations remains within an AOI.

Frequency of fixations [fixations/min]: This is a calculation of the mean number of fixations, made by each participant in each scenario, per minute.

Figure 4. The Field of View factor as displayed in the virtual scenarios: (1) Unrestricted (above); (2) Partially obscured by the road's curvature (middle); (3) Partially obscured by parked vehicles (below).

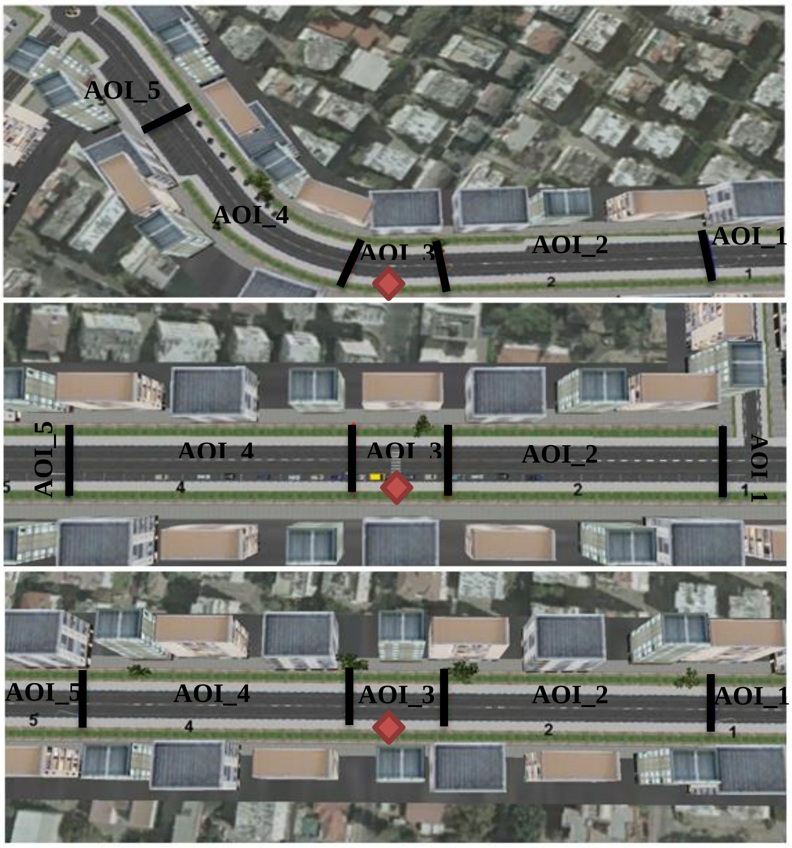


Figure 5. The Field of View factor as displayed in the virtual city terrain from a 2D perspective: (1) Unrestricted (below); (2) Partially obscured by the road's curvature (middle); (3) Partially obscured by parked vehicles (above). The division into areas of interest (AOIs) marked 1-5 is shown in each scene. The diamond shape icon represents the pedestrian point of view.

Results

Gaze distribution

For each participant and scenario, the Gaze distribution over the five AOI's sums up to one and therefore Gaze distribution is compositional data i.e., non-negative proportions with unit-sum. These types of data arise whenever we classify objects into disjoint categories and record their resulting relative frequencies, or partition a whole measurement into percentage contributions from its various parts. Therefore, attempts to apply statistical methods for unconstrained data often lead to inappropriate inference. Dirichlet regression suggested by Hijazi and Jernigan (2009) is more suitable. Dirichlet regression is a regression model that was design to deal with compositional data and analyse the five AOIs simultaneously under the constrained that they sum-up to one. The Dirichlet regression model was fitted using

DirichletReg package, in R Language. Applying a backward elimination procedure found the best fitting model has three significant main effects. The dependent variable was the vector of AOIs and the independent variables were Age-group, Traffic-Movement (TM) and Field of View (FOV); all of them to be statistically significant (p<0.05). Results from the Dirichlet regression shown in Table 2 revealed differences among age groups and among the crossing conditions. Table 3 shows the Gaze distribution for each category.

Table 2. The Dirichlet regression estimated coefficients for the set of five AOIs.

Variable	*Category*	*Far-left*	*Mid-left*	*Centre*	*Mid-right*	*Far-right*
Intercept		-1.90***	-1.76***	-1.63***	-1.72***	-1.76***
Age-group+	Aged 9-10	0.03	0.17	0.4***	0.26*	-0.06
	Aged 11-13	0.06	0.13	0.02	0.03	-0.02
	Adults	0.19*	0.01	-0.23*	0.02	-0.02
Traffic Movement++	One-way traffic movement	0.14	0.32***	0.47***	0.23**	0.16*
	Two-way traffic movement	0.11	0.17*	0.57***	0.63***	0.29***
Field of View+++	Restricted by parked vehicles	0.07	0.25**	0.54***	-0.22**	-0.15
	Restricted by road curvature	-0.31***	0.42***	-0.21**	-0.41***	-0.23**

+*Reference aged 7-8.* ++*Reference "no-traffic".* +++*Reference unrestricted field of view.*
** p < 0.05, ** p < 0.01, *** p < 0.001.*

Table 3. The mean Gaze distribution [%] for each of the Age-group, Traffic-movement, and Field of view categories.

Variable	*Category*	*Far-left*	*Mid-left*	*Centre*	*Mid-right*	*Far-right*
Age-group	Aged 7-8	9.3	29.9	32.9	15.0	12.9
	Aged 9-10	7.3	29.8	37.2	16.9	8.8
	Aged 11-13	9.5	34.7	27.8	15.8	12.2
	Adults	17.0	31.5	21.9	15.5	14.1
Traffic Movement	No-traffic	13.5	34.3	30.0	12.1	10.1
	One-way traffic movement	13.1	37.7	26.2	11.5	11.5
	Two-way traffic movement	9.7	23.2	27.7	23.4	16
Field of View	Unrestricted field of view	15.2	21.7	21.7	22.7	18.7
	Restricted by parked vehicles	16.7	25.0	38.8	11.9	7.6
	Restricted by road curvature	4.4	48.4	22.8	12.7	11.7

Children aged 9-10 spent relatively more time than all other age groups looking at the close range (AOI_3), while the adults showed a significant opposite, by spending relatively less time in the close area than all other groups. Adults spend relatively more time looking to the far left range (AOI_5) than all child-groups, thus they allocate more attention to the cars travelling on the close lane (i.e., in their perspective from left to right).

Traffic movement also had an effect on the participants' visual search distribution. The most evident phenomenon was that traffic movement, either one-way or two-way caused participants to spend relatively more time looking at the close proximity area than in scenarios where no travelling vehicles were present. With regard to FOV, it led to statistically significant distribution differences evident in both restricting conditions. When restricted on both sides by parked vehicles, participants spent more time looking at the close proximity area than in the unrestricted condition. It seems that when encountered with blocked field of view on both sides, participants tend to spend more time on the part of the scene that is not restricted, hence the immediate close area; and attempt to allocate more visual attention to the left. When restricted by a road curve on the left, statistically significant effects were shown in all AOI's. The curve, that limited participants from seeing the oncoming vehicles, was to the left of the participant point-of-view; once the vehicle coming from the curve became visible, it was less than 6 seconds until the vehicle arrived into the participant path of crossing. As a result participants spent more time looking at the mid-range left side (AOI_4) and less time at all the other AOIs than they did in the unrestricted scenarios.

Fixation duration

The mean fixation duration for each age group reveals a constant incline of average fixation duration with age (Table 4). While the youngest group had the shortest fixations in all AOIs, the adults had the longest average fixation in all AOIs but AOI_5 (Far-left). Fixation duration increases as are farther from the centre, almost twice as long in some cases. Although in far peripherals AOIs the difference between groups was larger, it was not statistically significant; which means that as fixations got longer, variance increased as well. Although not statistically significant, the longer average fixations for all groups were at the far-left area, about 500ms for the 3 oldest groups and 378ms for the youngest group.

Dwells duration

The results show that all groups had longer dwells in peripheral areas than in the centre (Table 4). Age differences were again very clear; adults' dwellings in peripheral areas were much longer than the two youngest age groups, especially on the left where the adults' average dwell was 1.54 s. in comparison to 1.03 s. for the children aged 7-8 and 1.1 s. for the children aged 9-10. The oldest group of children were again the most resembling to the adults.

Frequency of looks

The number of fixations each participant made in each scenario inside any of the AOIs was divided by length of the scenario, which gave the frequency of looks. The highest frequency was viewed in the youngest age group (Mean=52.2), and this score was significantly different than of the oldest child group aged 11-13 and the adults that scored 41.9 and 45.4 fixations/minute, respectively. Age group 9-10 was similar (2 fixation/minutes less) than that of the youngest group. Overall, two clusters of groups were visible, that of the two youngest age groups and that of the oldest child and adults.

Table 4. Mean scores and one-way ANOVA for visual search strategy variables

	AG1 (Aged 7-8)	*AG2 (Aged 9-10)*	*AG3 (Aged 11-13)*	*AG4 (Adults)*	*F ratio*	*Sig. group differences*
Fixation duration (ms)						
Far-left	378	515	493	499	0.59	
Mid-left	290	333	356	374	7.12***	AG1< AG3/AG4
Centre	236	260	254	271	4.47***	AG1 < AG4
Mid-right	266	325	303	340	3.99***	AG1 < AG4
Far-right	369	408	413	451	1.11	
Dwells (Seconds)						
Far-left	1.03	1.10	1.23	1.54	2.94*	
Mid-left	1.06	1.35	1.48	1.49	2.84*	AG1< AG4
Centre	0.76	1.04	0.92	0.76	2.89*	
Mid-right	0.75	1.08	1.21	0.99	2.76*	AG1<AG3
Far-right	1.13	0.98	1.50	1.43	1.75	
Frequency of fixations (fixations/min)	52.2	50.3	41.9	45.4	6.34***	AG1> AG3/AG4 AG2>AG3

*In this table, the F ratio is a measure of the overall significance of the differences between the four age groups. *p < .05, ***p<.001. The significance of differences between pairs of groups, reported in the last column, was tested with the Tukey HSD at level of 0.05.*

Discussion

The present experiment studied children's visual search strategies in hazardous road crossing situation by examining their gaze data while they were engaged in crossing decision tasks. The hypothesis stated that age has a direct link to visual behaviour and attention distribution, that is, that the older the child, the more his or her performance will resemble that of an experienced-adult.

Results indicated that age, limited field of view and the presence of moving vehicles affect the way pedestrians allocate their visual attention in the scene; and that age is associated with several measures of visual search. Indeed, adults were found to spend relatively less time in the centre area than children, shifting more visual attention to the far left area. It was also shown that the oldest child group (aged 11-13) showed partial resemblance to the adults, by spending relatively less time viewing the centre area than the younger children and relatively more time in the mid-left area. The measure of visual attentiveness, outside the scope of age reveals that pedestrian's visual search strategy is flexible and versatile giving the environment. Thus, pedestrians were more attentive to the road when driving vehicles were present than in scenarios with no traffic, and more importantly they accommodated their visual attention to restrictions such as parked vehicles and curvatures in the roadway.

This trend was apparent in all other gaze related measures. Fixation duration extends with age and as they moved from the centre to the peripheral areas, especially to the left, which coincides with the way they allocate their attention. The reoccurring diverse results the left side is not random, it may reflect a calculated manner of estimating the source of risk. Vehicles coming from the left are traveling in the close lane to the pedestrians and therefore constitute the immediate threat when stepping into the road. The dwells, much like the fixation, extends with age and adults seem to dwell longer in comparison to children, especially than the two youngest groups. It was also visible in all groups that peripheral dwells were longer than those in the centre. The frequency of fixations measure demonstrates again the age related differences; adults and the oldest group of children have a lower frequency of fixations than the younger children aged 7-10. It seems from the results that younger children tend to do more and shorter fixations; possibly, reflecting that younger children scan the environment in a more hectic manner.

Meir et al., (2013) showed that both 9–10-year-olds and 11–13-year-olds presented a less decisive performance reflecting upon higher attentiveness towards potential hazards along with better prediction of the upcoming events of adults. Taken together with previous behavioural findings (Meir et al., 2013) that crossing behaviour is age related. The pattern of the results presented here, serves as an additional indication for distinguishing between the middle age groups, a learning process that originates from gaining experience. The present research showed that gaze data might serve as a tool for differentiating between pedestrians with varied age and experience levels in a dynamic simulated environment. Findings serve applicable meaning, that is, further development of the characterization of children and adults' eye-movements patterns may be used as a methodology to determine potential pedestrians' readiness for crossing independence.

Acknowledgment

This research was supported by the Ran Naor Foundation, Tsippy Lotan, Ph.D, technical monitor.

References

Akhtar, N., & Enns, J.T. (1989). Relations between covert orienting and filtering in the development of visual attention. *Journal of experimental child psychology*, *48*, 315–334.

Day, M.C. (1975). Developmental trends in visual scanning. Advances in child development and behavior, 10, 153-193.

Dunbar, G., Hill, R., & Lewis, V. (2001). Children's attentional skills and road behavior. *Journal of experimental psychology: Applied*, *7*, 227–234.

Foot, H., Tolmie, A., Thomson, J., McLaren, B., & Whelan, K. (1999). Recognizing the hazards. *The Psychologist*, *12*, 400-402.

Grayson, G.B. (1975). The Hampshire child pedestrian accident study (Report No.668). Transport and Road Research Laboratory.

Hall, L.C. (1985). Searching high and low: the development of efficient visual search in hearing, deaf and learning disabled children. In R. Groner, GW McConkie, and C. Menz (Eds.), *Eye Movements and Human Information Processing* (pp 124–143). North-Holland.

Hijazi, R., & Jernigan, R. (2009). Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & statistics*, *4*, 77–91.

Huestegge, L., Skottke, E.-M., Anders, S., Müsseler, J., & Debus, G. (2010). The development of hazard perception: Dissociation of visual orientation and hazard processing. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*, 1–8.

Hunter, W.W., Stutts, J.C., Pein, W.E., & Cox, C.L. (1996). *Pedestrian and Bicycle Crash Types of the Early 1990s* (Report FHWA-RD-95–163. Washington, DC, US: Federal Highway Administration.

Lin, C.C., Hsiao, C.K., & Chen, W.J. (1999). Development of sustained attention assessed using the continuous performance test among children 6-15 years of age. *Journal of Abnormal Child Psychology*, *27*, 403–12.

Mack, A. (2003). Inattentional blindness: looking without seeing. *Current Directions in Psychological Science*, *12*, 180–184.

Mackworth, N.H., & Bruner, J.S. (1970). How Adults and Children Search and Recognize Pictures. *Human Development*, *13*, 149–177.

Meir, A., Parmet, Y., & Oron-Gilad, T. (2013). Towards understanding child-pedestrians' hazard perception abilities in a mixed reality dynamic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, *20*, 90–107.

NHTSA's National Center for Statistics and Analysis (2009). *Traffic safety facts, 2009: Pedestrians* (Report DOT-HS-811-394). Washington, DC, US: National Highway Traffic Safety Administration.

Pastò, L., & Burack, J. (1997). A developmental study of visual attention: Issues of filtering efficiency and focus. *Cognitive Development*, *12*, 523–535.

Tabibi, Z., & Pfeffer, K. (2007). Finding a safe place to cross the road: the effect of distractors and the role of attention in children's identification of safe and dangerous road-crossing sites. *Infant and Child Development*, *16*, 193–206.

Van der Molen, H.H. (1981). Child pedestrian's exposure, accidents and behavior†. *Accident Analysis & Prevention*, *13*, 193–224.

Whitebread, D., & Neilson, K. (2000). The contribution of visual search strategies to the development of pedestrian skills by 4-11 year-old children. *The British Journal of Educational Psychology*, *70*, 539–57.

World Health Organization (WHO). (2013). Global status report on road safety 2013: supporting a decade of action.

Zeedyk, M.S., Wallace, L., & Spry, L. (2002). Stop, look, listen, and think? What young children really do when crossing the road. *Accident; Analysis and Prevention*, *34*, 43–50.

Traffic Light Assistant – Can take my eyes off of you

Michael Krause, Verena Knott, & Klaus Bengler
Institute of Ergonomics, Technische Universität München
Germany

Abstract

A traffic light assistant on a smart phone is assessed with an eye tracker in real traffic. The system is displayed on three different screen sizes on a nomadic device and gaze durations are measured. Another condition of the experiment includes an acoustic click when the display content changes to reduce glance frequency. The acoustic click as an auditory hint does not reduce the frequency of gazes, as expected. The gaze durations can get shorter as display size increases, but this does not necessarily reduce the percentage of time an in-vehicle information system (IVIS) is looked at. Subjective ratings indicate that display contents can even be shown too big. Overall, the gaze durations are in line with current limits, even when displayed on a small screen. A set of gaze histograms and calculated gaze metrics is provided to enable comparison with other experiments and IVIS.

Introduction

Stops at traffic lights and related braking and acceleration are potential aims for optimization. In the Bavarian pilot project KOLIBRI, which stands for Kooperative Lichtsignaloptimierung – Bayerisches Pilotprojekt [Cooperative Traffic Light Optimization – Bavarian Pilot Project], four partners evaluated the feasibility and chances of traffic light optimization and driver information on arterial roads. The project was managed by TRANSVER GmbH, a traffic engineering company. The Board of Building and Public Works in the Bavarian Ministry of the Interior was responsible for the two rural roads, which served as test tracks. BMW AG built up a demonstration car. The Institute of Ergonomics at the Technische Universität München handled the human factors and ergonomics issues within the project, especially the information for the driver about upcoming traffic signal states with smart phones.

The final information system is based on already installed GSM and UMTS networks. The traffic lights on the two test tracks - one in northern Munich with seven traffic light controlled intersections on seven kilometres, the other one near Regensburg with eight traffic light controlled intersections on a length of five kilometres - were equipped with radio data modems. A central server receives the process data from the traffic lights and tries to estimate upon historical data their behaviour within the next minutes. This probability forecast is send to the demonstration car or smart phones. These devices calculate a speed

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

recommendation, or give other appropriate information about the traffic light phases to a driver. Due to the in-vehicle use, special care must be taken for suitability while driving (ESoP, 2006).

Related work – Traffic Light Assistance

Early contribution to the field of traffic lights assistance and driver information can be found in the *Wolfsburger Welle* project from VW AG (Hoffmann, 1991; Voy et al., 1981). About the same time, Australian traffic engineers experimented with driver information and traffic lights, especially with dynamic speed recommendations signs along the road (Trayford et al., 1984; Bell 1984; van Leersum, 1985; Trayford & Crowle, 1989). In the CVIS project, a green wave in-vehicle information system was compared to dynamic signs along the road, in a simulator study (Duivenvoorden, 2007; Pauwelussen et al., 2008).

Thoma (2010) compared different human machine interfaces (HMI) for a traffic light information on-board system. The idea, to feed traffic light information into the adaptive cruise control, can also be found in Thoma (2010) and in the KOLINE project (Bley et al., 2011). Chen et al. (2009) gave a review on signal countdowns and showed the risks of using traffic light signal countdowns. Ferreira et al. (2010) dealt with the idea of virtual traffic lights (VTL) in the car, to replace the roadside infrastructure. Olaverri et al. (2012) evaluated HMIs on the head up display for the VTL in a simulator. The TRAVOLUTION project from AUDI AG (Braun et al., 2009) used WiFi to bring the traffic phase information to the car. This approach was also used in the German *aktiv* project (Hoyer, 2012), which used a mobile device to inform the driver. Data for WiFi connection ranges under different vehicle speeds can be found in (Iglesias et al., 2008). A traffic phase information system was also included in the German simTD project (2008). The SignalGuru project (Koukoumidis et al., 2011) used smart phones and image processing to estimate traffic light phases.

KOLIBRI differs from other projects in its heavy reliance on already installed mobile communication networks, instead of dedicated short range communications, with their coverage drawbacks and maintenance costs. The information presentation on the wide spread smart phones makes it possible to equip every car, regardless of make and year. The most important distinction is the user- and usability-centered approach. Its use while driving needs special attention. Many of the preceding or running projects showed a lack of consideration for user needs, thus focusing on technical feasibility only.

This paper will focus on a sub-aspect of "suitability while driving": The gaze behaviour.

Related work – Gaze behaviour

Gaze behaviour in transportation is a long-researched topic. All major guidelines for in-vehicle information systems (IVIS) (i.e. JAMA, 2004; AAM, 2006; ESoP, 2006) address the visual demand with some gaze metrics (occlusion metrics or glance durations). For instance, the AAM states: "[…] single glance durations generally should not exceed 2 seconds”. And the ISO 15005:2002, which is referenced in the ESoP, introduces a maximum dwell time of 1.5s to capture information.

There is also a research tradition on the determination of optimal font size (Beymer et al., 2008) and even the interaction between font size, display size and remembering was assessed (Sanchez & Goolsbee, 2010). Tullis (1984) used metrics like grouping and density, which are indirectly connected to display size, to predict search times and subjective ratings on alphanumeric displays. The well-known model SEEV (Wickens et al., 2001) has a special focus on visual attention allocation. The Distract-R approach (based on ACT-R) currently tries to expand this field with salience features (Lee et al., 2012). The advantages of an improved SEEV, called NSEEV, can be found by Steelman et al. (2011). NSEEV especially uses extended salience features and can be used for more than just steady-state distributions.

In the experiment presented here two topics related to visual attention are addressed:

- Given that smartphone screen sizes on the market are heterogeneous, how does the display size of presented graphical information on small devices affect gaze duration?
- Can an auditory hint (a click for new information on the display) reduce glances to the designed traffic light assistance IVIS and thus glance frequency?

Accordingly, it is possible to state the hypothesis that gaze duration should get shorter when the display size of small devices increases because the graphical content should be easier to perceive. On the other hand, this should only be true on a small scale, since the eye has only a narrow spot (fovea) for sharp seeing. So, if the display gets larger, one has to look around on the screen. This has been addressed by many usability studies (e.g. for websites, Nielsen 2006).

The discussion of the results of this experiment will try to connect the outcomes to the SEEV model. “SEEV derives its name from the four forms of attentional influence that it posits: channel *s*alience, the *e*ffort needed to move attention between channels, the operator’s *e*xpectancy of the signal on each channel, and the task criticality or *v*alue of the information in each channel” (Steelman et al., 2011).

Method

Test environment

Test track

The individual test runs of the experiment were performed on federal road 13 (B13) in northern Munich. Over a length of about seven kilometres, the test track has seven traffic-light-signal-controlled intersections. Moreover, it is a section of the B13 with two lanes in each direction, and in addition, there are turning lanes in front of the individual traffic lights. A guardrail separates the carriageways structurally from each other. Under dry road conditions, the speed limits on the test track are mostly limited to 100 km/h, and reduced to 70 km/h in front of every intersection. In the various test runs used in the experiment, this route was driven either from the south (48° 14’ 48 N 11° 36’ 7 E) to the north (48° 17’ 56 N 11° 34’ 36 E), or in the opposite direction. At both ends of the test track, two turning points were used during the experiment to evaluate the performed test run subjectively by questionnaires and to explain the object of the following test run. The experiments avoided the rush hours, which have highly directional traffic load. Thus, the traffic density was about 500 cars/h, for each direction. In the experiment reported here, the traffic lights were acting on a coordinated scheme (green wave). The predetermined switching times are programmed into a nomadic device to inform the driver.

Test vehicle

For the study on the real test field, a BMW X5 (modified US model) was used as vehicle. In the experiment, a Samsung Galaxy Tab GT-P1000 (display: 7.0 inch, 1024 x 600, Android-Firmware-Version 2.3.6) was used to display the mobile KOLIBRI-traffic light assistant. The nomadic device was fixed to the right of the driver with a mount in the ventilation slots of the centre console of the test vehicle. The automated analysis of gaze data, which were generated by the Dikablis Eye-Tracking System (Ergoneers GmbH, Manching), also required the attachment of markers on the dashboard of the test vehicle (see Figure 1).

Figure 1. Attachment of the nomadic device and the markers in the test vehicle

Human Machine Interface (HMI)

After initial experiments regarding various HMIs in the static driving simulator (Krause & Bengler, 2012a – 2012b), the optimized concept in Figure 2 is used for the experiment in real traffic.

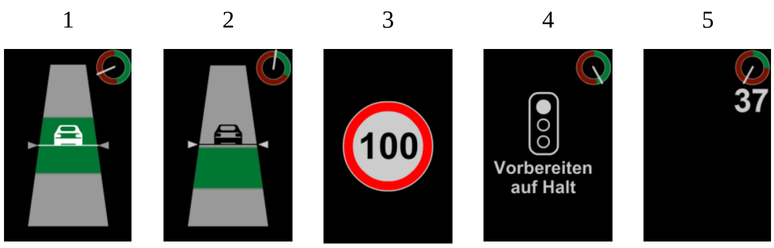


Figure 2. Display states of the KOLIBRI-traffic light assistant

The favoured display concept of the KOLIBRI-traffic light assistant on the nomadic device, which was used in this study, includes a speed recommendation in form of a speed carpet. The green area represents the speed recommendation that supports the driving behaviour for a green wave (Figure 2-1 and Figure 2-2). The illustrated vehicle position corresponds to the current speed. The white vehicle indicates that the driver is within this recommendation while driving at current speed. The black car of Figure 2-2 points to the contrary. The position of the pointer of the "Heuer" traffic light in the right upper corner of the application shows the current cycle state - green or red light - of the next traffic light. The speeding display appears when the speed limit is violated (> 10 km/h), see Figure 2-3. If the vehicle is outside the criteria for calculating a speed recommendation (i.e. if the driver has to move too slowly or too fast to reach the green wave), Figure 2-4, called "Vorbereiten auf Halt" ("Prepare to Stop"), is displayed. It shows the driver that the next traffic light will be red at arrival. The combination of the Heuer traffic light and the countdown appears at a speed less than 5 km/h in front of a red light signal system (Figure 2-5). This function displays the waiting time, until the traffic light changes to green again.

Independent variables: variation of size and click

This experiment used variations of the KOLIBRI-traffic light assistant in the display size. Furthermore, a variant was used with an acoustic signal. Figure 3 illustrates the relative proportions of the screen sizes of the display concepts of the assistance system in the individual test runs. In the previously conducted studies in the static driving simulator, the traffic light assistant was presented on a smartphone (Samsung Galaxy Ace S5830) with a 3.5-inch diagonal (Krause & Bengler, 2012a – 2012b). So the experiment could be compared in real traffic with the appropriate simulator study, one display of the variations on the tablet should correspond to the smartphone display in size. The display concept HMI_M has the characteristic of a 3.5-inch diagonal. The large display HMI_L was adapted in such a way that the screen of the tablet is utilized optimally (width of 900px). Derived from these dimensions generated for the HMI_L display, the exact sizes of the other KOLIBRI-displays result from the specified scaling factors.

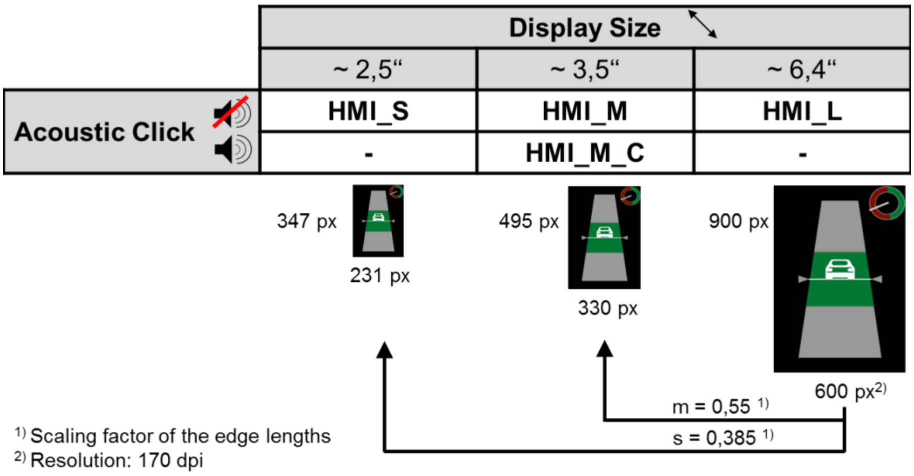


Figure 3. Relative proportions of the screen sizes

The smallest display concept HMI_S corresponds to half of the display surface of the HMI_M-system. The fourth variant of KOLIBRI-assistance, indicating a change between different states by an acoustic click (HMI_M_C), is identical in size to HMI_M-display concept. The HMI prototype was implemented in Adobe AIR. The size scaling was performed inside AIR (vector graphics). There was nothing like a border around the resized graphic. The resized graphics were just centred on the black screen. The short click was played when the display changed between one of the states of Figure 2-1 to Figure 2-5 (for instance, if the vehicle travelled too fast and Figure 2-3 appeared or when the car drove out of or into the speed recommendation, e.g. from Figure 2-1 to Figure 2-2). The volume of the tab was set to maximum. The display content was recalculated for every new GPS value, typically at 1Hz.

Gaze tracking

Within the scope of this study, the Dikablis Eye-Tracking System (Ergoneers GmbH, Manching) was used for experimental recording of the gaze data and D-Lab 2.1 for analysis. Three areas of interest were defined (Figure 4): "windscreen" (or "front"), "speedometer" and "smartphone" (or "tablet"). The "smartphone" area had a fixed size and was not adjusted when display content was scaled.

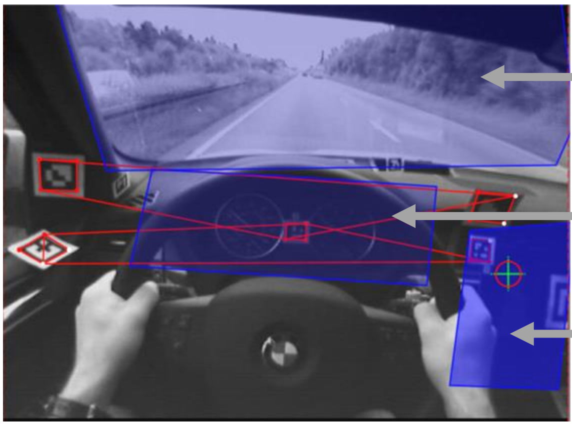


Figure 4. Areas of Interest (AOI)

Questionnaires

In addition to the questionnaire for collecting demographic information, the following standardized written survey instruments were used in this study:

NASA-Task Load Index

The method for collecting of ratings of the workload, called NASA-TLX, was developed by the Human Performance Group at NASA Ames Research Center (Hart, 1986; Hart & Staveland 1988). In this study, we used a German translation (Seifert, 2002). An Overall Workload Index (OWI), based on a non-weighted average of the ratings of the subscales, was used for comparison (Hart, 1986).

System Usability Scale

The System Usability Scale (SUS) represents a reliable and cost-effective questionnaire with ten items (each on a 5-point-Likert-scale) for a global rating of the usability of interactive systems (Brooke, 1996). In the experiment, the SUS was used after each full run combined with the KOLIBRI-traffic light assistance. A careful German translation was used, combined with the English original text. There is some effort in Germany (Reinhardt, 2012) to get a uniform and verified translation. The translation used here was involved in different experiments at our institute, so we published the translated items and condensed SUS results from these studies (Krause & Bengler, 2013) to make comparisons possible.

Test procedure

After the test subjects were welcomed and informed about the KOLIBRI project, they indicated their consent to participate in the study. In addition, a demographic questionnaire collected personal data. Afterwards, the explanation of the maximum 120 minutes ongoing trial followed. After successful calibration of the eye tracking system and vehicle-specific instructions, the explanation of the KOLIBRI-traffic light assistant with its functions followed.

After these preparations, the acclimatisation test run took place using the KOLIBRI-traffic light assistance, which included the acoustic click. In addition, five test runs in randomised order were realised in the BMW X5 on the test track B13. A baseline run (BL) without the use of a traffic light assistant was compared with four other runs combined with various KOLIBRI systems that differed in their display sizes (HMI_S, HMI_M, HMI_L). Moreover, in another test run (HMI_M_C), the KOLIBRI assistant with a 3.5-inch diagonal and an acoustic click was tested on the test track. The signal happens at each change in display to illustrate a display change to the driver.

The individual test runs were randomly done in one direction (south-north or north-south). To assess the subjective workload during the test runs, both the NASA-TLX and the SUS questionnaire (except for the baseline run) had to be completed by the test subject after finishing the run.

Participants

The sample of this study included 22 test persons (5 females; mean age = 28.6 years; st. dev. = 8.4). All test subjects had a valid driving license. No test person of the study had a colour deficiency or important eye disease for the test procedure. However, 32% of the subjects reported the use of a visual aid. Nine percent of the participants were not familiar with driving an automatic car and two persons of the KOLIBRI study did not know the real test track in northern Munich. With a proportion of 45 percent, almost half of all participants of the experiment had already participated in a KOLIBRI study in the driving simulator. 63 percent of the subjects had participated in a previous study that was carried out with the use of the traffic light assistant in real traffic. In this study, reaction times were collected by handling a tactile detection task.

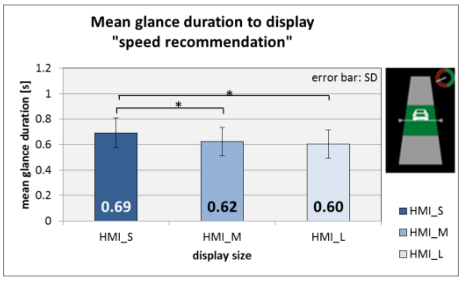
Results and discussion

Glance duration and display size

The traffic lights assistant reported the current display state to the eye tracking system. So it is possible to split the gaze analysis up. An average trial run in the “HMI_L”, “HMI_M” and “HMI_S” conditions lasted 417 seconds (SD: 40s). From this time, only sections were analysed when the vehicle was moving. This was achieved by considering the display states “speed recommendation” (see. Figure 2-1 and Figure 2-2) which was, on average, active for a total time (summed up segments) of 200 seconds (SD: 53s) during a test run. And “prepare to stop” (Figure 2-4), which was, on average, active for a total time of 124 seconds (SD: 66s). Figure 5 shows the results.

A single factor (display size), repeated measurement ANOVA for “speed recommendation” reported a significant difference $F(2;42) = 13.4$, $p < 0.001$, whereas the same test for the state “prepare to stop” was not significant $F(2;42) = 0.646$ $p = 0.529$. A Bonferroni corrected post-hoc test revealed the data in Table 7.

1



2

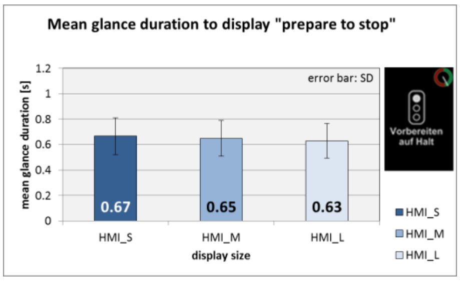


Figure 5. Mean glance durations with different display sizes

*Table 7. Bonferroni corrected post hoc test for glance durations in display state "speed recommendation" (*p < 0.05)*

HMI_S – HMI_M	p = 0.004*
HMI_S – HMI_L	p < 0.001*
HMI_M – HMI_L	p = 0.923

For the "speed recommendation" display state, the same ANOVA (display size) was carried out on glance frequency and glance percentage of time. Both were not significant (average glance percentage of time to the phone during "speed recommendation" was HMI_S: 12.1%; HMI_M: 10.2%; HMI_L: 11.2%; average glance frequency during "speed recommendation" was HMI_S: 0.174Hz; HMI_M: 0.163Hz; HMI_L: 0.186Hz). It seems like a larger screen can reduce glance durations up to some point, where information is easy accessible and additional increasing has only a minor effect (maybe a further enlargement can even worsen glance durations). At the same time, a larger information presentation can slightly increase the glance frequency, which even counteracts potential benefits. In terms of SEEV, the top-down features (*e*xpectancy and *v*alue) should not be affected by resizing. The *e*ffort to move the gaze to the HMI should also not be highly affected (same location). The slightly increasing gaze frequency could be explained by increasing *s*alience with display enlargement. What is hard to explain and model with SEEV would be the decreasing glance duration. In NSEEV, the parameters "inhibition-of-return", or the fixation duration maybe could be used. An easier approach would be to add a bottom-up parameter, which accounts for information retrieval costs within one area of interest.

Glance frequency and acoustic click

For the analysis of glance frequencies, the data are analyzed similarly to the antecedent section and treated separately for the display states of "speed recommendation" and "prepare to stop". Figure 6 shows the results. The hypothesis was that the acoustic click could reduce check gazes and lower glance frequency.

However, a one-tailed t-test revealed no significance in both cases. For the state of “prepare to stop”, even the direction of the hypothesis is wrong.

Table 8. One-tailed t-tests for the reduction of glance frequencies by acoustic click

speed recommendation	t(21) = 0.577; p = 0.285 >0.05
prepare to stop	t(21) = -0.414; p = 0.658 >0.05

A two-tailed t-test for paired samples between the states of “speed recommendation” and “prepare to stop” (i.e. Figure 6-1 versus Figure 6-2) reveals: t(43) = 4.7; p <0.001. So the system design, and information presentation (to show in some situations “prepare to stop”), reduces the glance frequency significantly.

One explanation with SEEV for the inefficiently working acoustic help could be top-down *e*xpectancy, together with a sufficient bottom-up visual *s*alience. The display Figure 2-1 contains the Heuer sign in the upper right corner, an item with low informational bandwidth (it behaves like a clock and is predictable) and the dynamic speed recommendation with a medium to high information bandwidth. The display of Figure 2-4 also contains a Heuer clock and a static message with extremely low informational bandwidth. Maybe, if visual *s*alience meets the requirements imposed by *e*xpectancy, there is not much room for improvement by the acoustic sound.

1 2

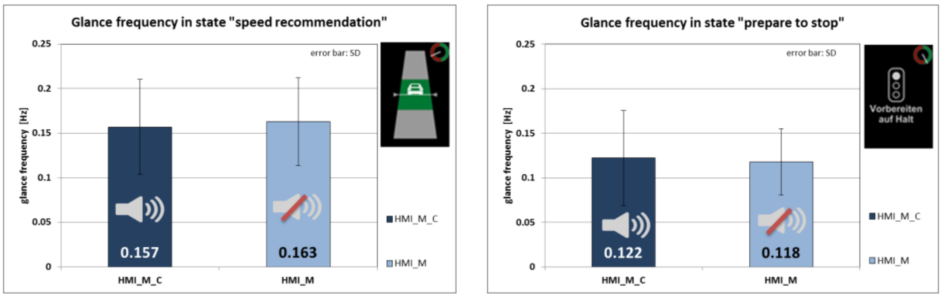


Figure 6. Mean glance frequency with and without acoustic click

System Usability Scale

The SUS scores for the different system variants are reported in Figure 7. Single factor, repeated measurement ANOVA showed no significance: F(3;63)=1.207, p=0.314.

Looking at the absolute values HMI_M seems to be the favorite. According to Bangor et al. (2009), a score of 86.4 can be connected to the adjective “excellent”. There seems to be a drop in the SUS-Score for the HMI_L. In terms of SEEV, this could be a mismatch between top-down parameters (*e*xpectancy and *v*alue) and the bottom-up parameter *s*alience. The subjects feel it is too large.

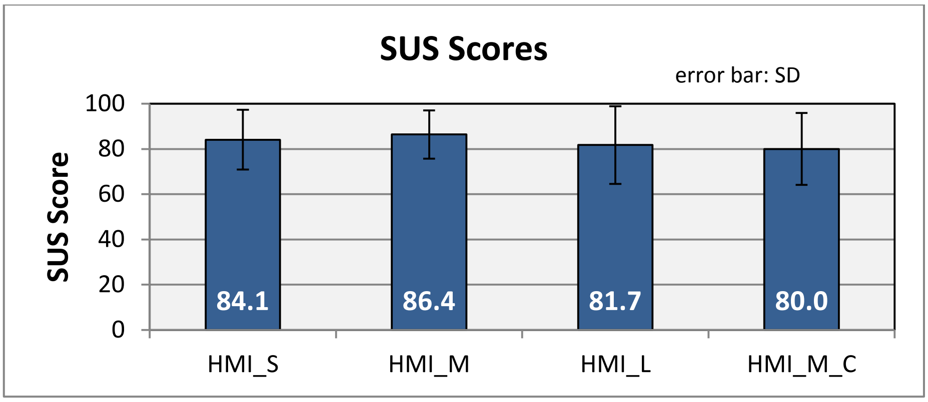


Figure 7. SUS Scores of the different system variants

NASA-TLX

Figure 8 shows the results of the NASA-TLX questionnaire. Single factor, repeated measurement ANOVA (Greenhouse Geisser corrected, due to Mauchly test) revealed $F(2.9;60.7)=4.577$, $p=0.006$. A Bonferroni corrected post-hoc test was significant for: baseline/HMI_S ($p=0.004$), baseline/HMI_M_C ($p=0.048$) and HMI_S/HMI_M ($p=0.047$). Looking at the absolute values, HMI_M imposes the lowest additional subjective demand compared to the baseline run.

Figure 9 depicts the results from the NASA-TLX sub-dimensions. The most subjective impact for the smartphone information comes from the "mental" and "temporal" dimensions.

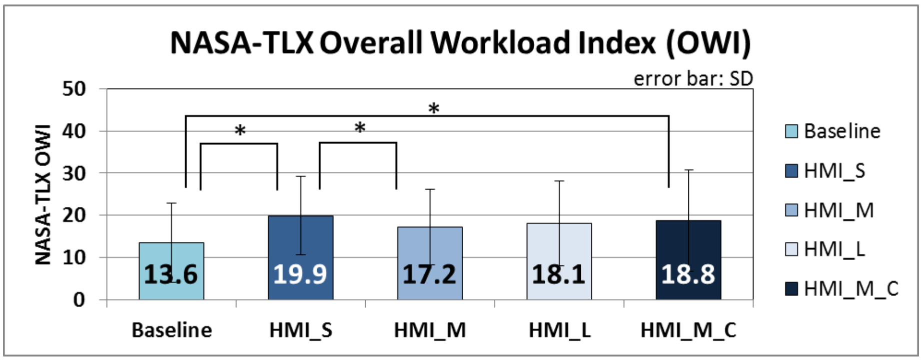


Figure 8. NASA-TLX Overall Workload Index (OWI)

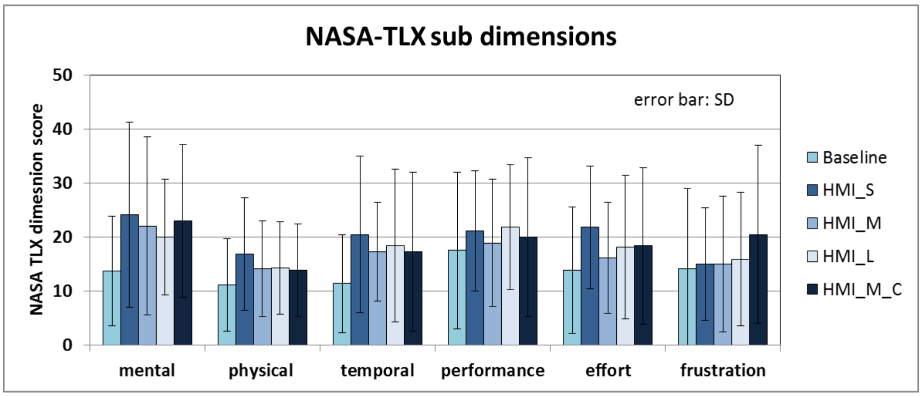


Figure 9. NASA-TLX sub-dimensions

User preferences

The test subjects were asked which variant they preferred (forced choice).

Table *9* contains the results. This led to a split decision between HMI_M and HMI_M_C.

Table 9. Number of users that prefer a specific system variant

HMI_S	3
HMI_M	8
HMI_M_C	8
HMI_L	3

The comments of the participants were written down and grouped. Table 10 lists the remarks expressed most often. The most frequent answer reveals that display components can be even subjectively too large. The often sounding signal can be a result of driving style: Some participants drove around the border of the speed recommendation and got a lot of clicks when they drove in and out of the velocity carpet (despite an implemented hysteresis).

Table 10. Top three comments with number of occurrence

Presentation of HMI_L is too large	7
Acoustic click is a good idea	6
Acoustic signal sounds too often	5

General gaze histograms and metrics

For the following histograms, the gaze data were exported from D-Lab and processed by Matlab. An additional minimal threshold of 120ms for single glances was introduced. The histograms only show data when the smart phone was in the "speed recommendation" display state (Figure 2-1, Figure 2-2). For the baseline condition, the display was blanked (black), but the smartphone reported to the eye tracking system if a speed recommendation would have been shown to get comparable road sections. Figure 10 shows the distribution of gazes to the smart phone. The distribution data in the upper right corner reveals that the 85th percentile of all gazes from all participants is 0.88 seconds.

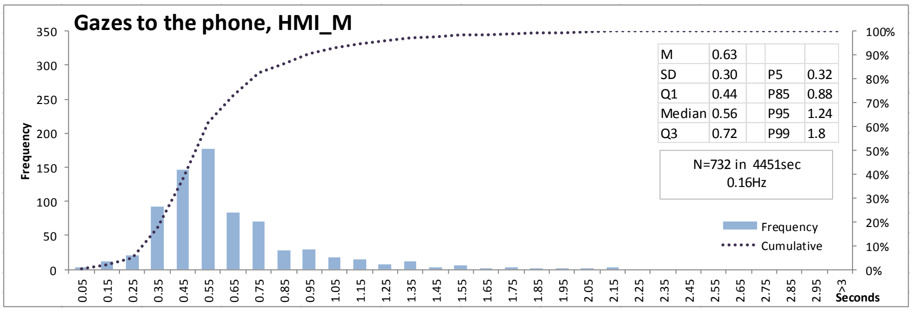


Figure 10. Gazes to the phone, condition HMI_M

The smartphone shows a speed recommendation and can induce people to combine gazes to the phone with gazes to the speedometer. So Figure 11 holds the times, participants do not look to the area "windscreen" (called "Eyes off the road"). Figure 11 shows that the 85th percentile value of this metric (1.12 seconds) is also below the limit of current major guidelines. Figure 12 holds the same metric for the baseline condition.

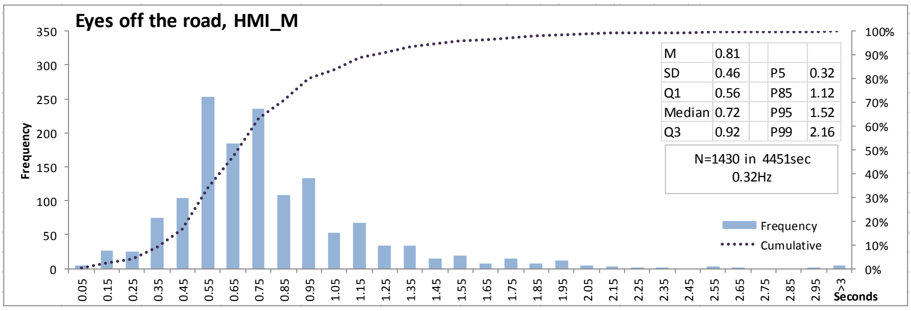


Figure 11. Eyes off the road, condition HMI_M

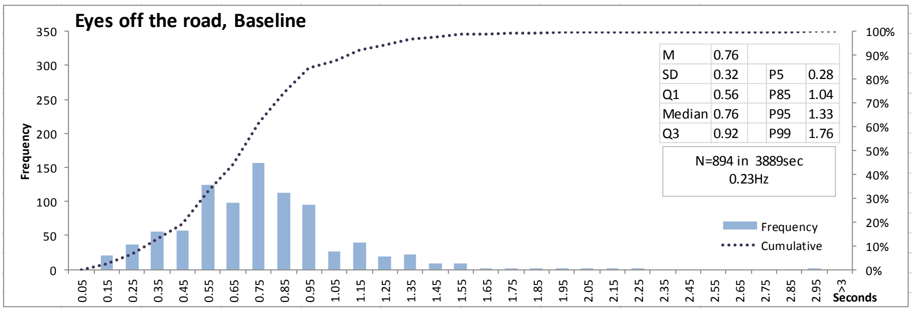


Figure 12. Eyes off the road, condition Baseline

Figure 13 and Figure 14 hold the distribution of gazes through the windscreen for the conditions HMI_M and baseline (both under the condition that the display shows a speed recommendation). Both figures show an exponential decay. For modelling, the lowest and highest bin from the figures were excluded. An exponential curve fitting led to:

*occurrence(t) = 233*e^-(t/2.28s)* for Figure 13 with a fitting of R^2=0.945 and

*occurrence(t) = 144*e^-(t/2.82s)* for Figure 14 with a fitting of R^2=0.932.

The constant (233 or 144) adjusts for the relative differences, but the difference in mean lifetime (2.28s or 2.82s) should be an interpretable gaze metric of a system. The baseline has a shallower curve (mean lifetime 2.82s) with relatively longer gazes to the front. The more often used term with exponential decay is *half-life* (= ln(2)*mean lifetime), which is 1.95 seconds for Figure 14. The visual demand of a system lowers the mean lifetime and, accordingly, the half-life of front gazes. Or to put it another way: Gazes through the windscreen are a kind of eyes-off-the-system glances.

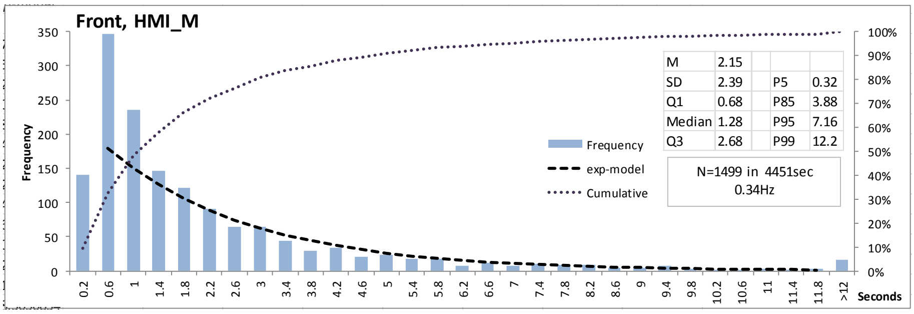


Figure 13. Gazes through the windscreen; condition HMI_M

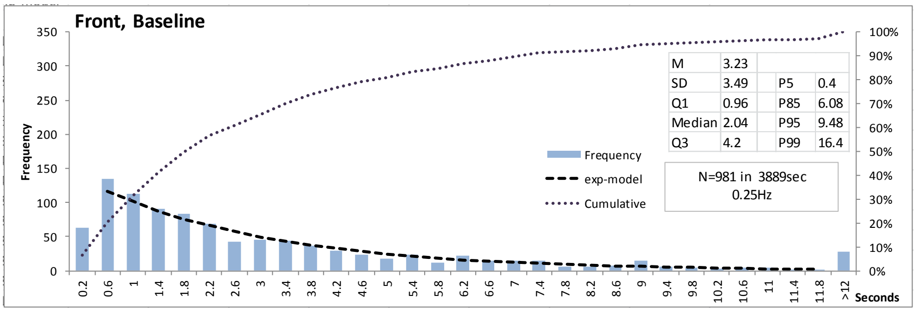


Figure 14. Gazes through the windscreen; condition Baseline

Krause and Bengler (2012b) used an artificial metric to compare different HMI variants: As measure of *severity*, the average gaze frequency was multiplied with the

85th percentile value (both values are derived from the data pool of all participants, not with an intermediate step of individual averages). Thus, the severity measure for the “natural” behaviour (e.g. speedometer checks) in Figure 12 would be 0.23Hz x 1.04s = 0.24. The severity with speed recommendations from the smart phone (Figure 11) would be: 0.32Hz x 1.12s = 0.36. We interpret the difference of 0.36 – 0.24= 0.12 as an additional visual load imposed by the smartphone. A similar value can be obtained by calculating the severity of Figure 10: 0.16Hz x 0.88s = 0.14.

The multiplication of frequency by mean glance duration gives a percentage amount of time. For instance, according to the data in Figure 10, the participants show 0.16Hz x 0.63s = 0.1, so 10% of the time to the phone. The multiplication of frequency by the 85th percentile of glance duration maybe can include some information about how heavy the tail of the glance distribution is. The long gaze durations in the tail are important for driving safety (Horrey et al., 2007).

Conclusion

An increase in display size can reduce glance duration, but it does not necessarily reduce the percentage of time an IVIS is looked at. Furthermore, subjective ratings indicate that display contents can even be shown too big. The gaze durations for the traffic light assistant are in line with limits of major guidelines, even when displayed on a small smartphone or must be scaled by height due to landscape mode.

An acoustic click (as hint for display changes) does not reduce glance frequency as expected. Subjective ratings are divided. The click will be implemented in the final system and can be enabled by the user. This will also include a function to suppress fast successive clicks for some driving styles. Appropriate information presentation design has shown to reduce glance frequency for this system more effectively than acoustic feedback.

The subjective ratings from NASA-TLX dimensions indicate that participants feel a slight mental and temporal demand when using the system. This was assessed in another experiment with a tactile detection task (TDT), to get objective data.

The different comments of the participants motivated to carry out some minor tweaking of the system. With a view of the already excellent SUS-scores, it is clear that, with each iteration of the engineering process, it gets much more difficult to score higher.

A set of gaze histograms and calculated gaze metrics was provided to enable comparison with other experiments and IVIS, as well as discussion with researchers.

Outlook

From the results of this experiment one can argue: Maybe the interface and information are within the limits for gaze behaviour, but how strong is the mental demand on the driver? Another experiment was carried out to test this and measure the mental demand of the information system with a tactile detection task (TDT).

The TDT is a tactile variant of the detection response tasks (DRT), which are currently in a standardization process by *ISO TC22 SC13 WG.8*.

In the experiment reported here, the traffic lights were acting on a predetermined and coordinated scheme (green wave). Afterwards, the traffic lights on the test roads were equipped with data modems to transmit their current state and were switched back to a traffic-adaptive control scheme. Based upon the transmitted process data, a central server tried, based upon historical data, to estimate how the traffic light probably will act within the next couple of minutes. The server sent this forecast to a demonstration car and smart phones. In the last field experiment, we tested this improved app and examined if the estimated switching times can be used to inform a driver.

Acknowledgement

The authors would like to thank Philipp Schneider and Johannes Henel for preparing the gaze data.

References

AAM (2006). Driver Focus-Telematics Working Group: Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems.

Bangor, A., Kortum, P., Miller, J. (2009). Determining What Individual SUS Score Mean: Adding an Adjective Rating Scale. *JUS Journal of Usability Studies 4(3)*, 114–123.

Beymer, D., Russell, D., Orton, P. (2008). An eye tracking study of how font size and type influence online reading. In: *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction - Volume 2* (pp. 15-18). British Computer Society, Swinton, UK.

Bell, A. (1984). If you always caught the green light. *ECOS Magazine(Issue 40).*

Bley, O., Kutzner, R., Friedrich, B., Saust, F.; Wille, J., Maurer, M., Niebel, W., Naumann, S., Wolf, F., Schüler, T., Bogenberger, K., Junge, M., Langenberg, J. (2011). Kooperative Optimierung von Lichtsignalsteuerung und Fahrzeugführung. *12. Braunschweiger Symposium AAET – Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel, Braunschweig.*

Braun, R., Busch, F., Kemper, C., Hildebrandt, R., Weichemeier, F., Menig, C., Paulus, I., Preßlein Lehle, R. (2009). TRAVOLUTION - Netzweite Optimierung der Lichtsignalsteuerung und LSA-Fahrzeug-Kommunikation. *Straßenverkehrstechnik 2009(6)*, 365–374.

Brooke, J. (1996). SUS - A quick and dirty usability scale. In: *Usability evaluation in industry* (pp. 189–194). Taylor & Francis, London ; Bristol, PA.

Chen, H., Zhou Huaguo, Hsu, P. (2009). What Do We Know About Signal Countdown Timers? *ITE Journal on the web July 2009,* 72–76.

Duivenvoorden, K. (2007). *Roadside versus in-car speed support for a green wave: a driving simulator study*. Master's Thesis, University of Twente.

ESoP (2006). European Commission Recommendation of 26 May 2008 on safe and efficient in-vehicle information and communication systems: update of the

European Statement of Principles on human-machine interface (2008/653/EC).

Ferreira, M., Fernandes, R., Conceição, H., Viriyasitavat, W., Tonguz, O.K. (2010). Self-organized traffic control. In: Proceedings of the seventh ACM international workshop on VehiculAr InterNETworking (pp. 85-90). ACM, New York, NY, USA.Mulder, L.J.M. (1992). Measurement and analysis methods of heart rate and respiration for use in applied environments. *Biological Psychology, 34*, 205-236.

Hart, S.G. (1986). NASA Task Load Index (TLX). Volume 1.0; Paper and Pencil Package.

Hart, S.G., Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: *Human mental workload*. North Holland Press; Elsevier Science Pub. Co., Amsterdam.

Hoffmann, G. (1991). Up-to-the-minute information as we drive—how it can help road users and traffic management. *Transport Reviews 11(1)*, 41–61.

Horrey, W., Wickens, C. (2007). In-Vehicle Glance Duration: Distributions, Tails, and Model of Crash Risk. *Transportation Research Record Vol. 2018*, 22 – 28.

Hoyer, R. (2012). Verkehrliche Potenziale des vorausschauenden Fahrens an kooperativen Lichtsignalanlagen. *5. Tagung Fahrerassistenz München.*

JAMA (2004). Guideline for In-vehicle Display Systems — Version 3.0

Iglesias, I., Isasi, L., Larburu, M., Martinez, V., Molinete, B. (2008). I2V Communication Driving Assistance System: On-Board Traffic Light Assistant. In: *Vehicular Technology Conference, 2008 (*pp. 1–5*)*. IEEE VTC 2008.

Koukoumidis, E., Peh, L.-S., Martonosi, M.R. (2011). SignalGuru: leveraging mobile phones for collaborative traffic signal schedule advisory. In: *Proceedings of the 9th international conference on Mobile systems, applications, and services* (pp. 127-140). ACM, New York, NY, USA.

Krause, M., Bengler, K. (2012a). Traffic Light Assistant – Driven in a Simulator. In *Proceedings of the 2012 International IEEE Intelligent Vehicles Symposium Workshops.*

Krause, M., Bengler, K. (2012b). Traffic Light Assistant Evaluation of Information Presentation. In: *Advances in Human Factors and Ergonomics 2012. Proceedings of the 4th Ahfe Conference* (pp. 6786-6795). CRC Press.

Krause, M., Bengler, K. (2013). Subjective Ratings in an Ergonomic Engineering Process Using the Example of an In-Vehicle Information System. In Masaaki Kurosu (Ed.): Vol. 8005. *Lecture Notes in Computer Science, Human-Computer Interaction. Applications and Services* - 15th International Conference, HCI International 2013, Las Vegas, NV, USA, July 21-26, 2013, Proceedings, Part II (pp. 596–605). Springer.

Lee, J., Lee, J.D., Salvucci, D.D. (2012). Evaluating the distraction potential of connected vehicles. In: *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 33-40). ACM, New York, NY, USA.

van Leersum, J. (1985). Implementation of an advisory speed algorithm in TRANSYT. *Transportation Research Part A: General 19(3)*, 207–217.

Nielsen, J. (2006). F-Shaped Pattern For Reading Web Content. *Jakob Nielsen's Alertbox*.

Olaverri Monreal, C., Gomes, P., Krüger Silvria, M., Ferreira, M. (2012). In-Vehicle Virtual Traffic Lights: a Graphical User Interface. *7th Iberian Conference on Information Systems and Technologies, CISTI'2012, Madrid, Spain*, 1–6.

Pauwelussen, J., Hoedemaeker, M., Duivenvoorden, K. (2008). CVIS Cooperative Vehicle-Infrastructure Systems. *Deliverable D.DEPN.4.1b Assess user acceptance by small scale driving simulator research.*

Reinhardt, W. (2012). Crowdsourcing the translation of SUS. Newsblog: *http://isitjustme.de/2012/01/crowdsourcing-the-translation-of-sus/* last check 27th December 2012.

Sanchez, C.A., Goolsbee, J.Z. (2010). Character size and reading to remember from small displays. *Comput. Educ 55(3)*, 1056-1062.

Seifert, K. (2002). *Evaluation multimodaler Computersysteme in frühen Entwicklungsphasen. Ein empirischer Ansatz zur Ableitung von Gestaltungshinweisen für multimodale Computer-Systeme*. PhD Thesis, Technische Universität Berlin.

simTD (2008). Sichere Integrierte Mobilität – Testfeld Deutschland. German C2X project, *http://www.simtd.de*

Steelman, K.S., McCarley, J.S., Wickens, C.D. (2011). Modeling the Control of Attention in Visual Workspaces. *Human Factors: The Journal of the Human Factors and Ergonomics Society 53(2)*, 142–153.

Thoma, S. (2010). *Mensch-Maschine-Interaktionskonzepte für Fahrerassistenzsysteme im Kreuzungsbereich*. Dissertation, Technische Universität München.

Trayford, R.S., Doughty, B.W., Wooldridge, M.J. (1984). Fuel saving and other benefits of dynamic advisory speeds on a multi-lane arterial road. *Transportation Research Part A: General 18(5-6)*, 421–429.

Trayford, R., Crowle, T. (1989). The ADVISE traffic information display system. In: *Vehicle Navigation and Information Systems Conference* (pp. 105–112) Conference Record.

Tullis, T.S. (1984). *Predicting the usability of alphanumeric displays.* PhD Thesis, Rice University

Voy, C., Zimdahl, W., Mainka, W., Stock, F. (1981). Einrichtung zur Verkehrsführung längs einer Fahrbahn. German Patent DE 3126481C2.

Wickens, C.D., Helleburg, J., Goh, J., Xu, X., & Horrey, W.J. (2001). Pilot Task Management: Testing an Attentional Expected Value Model of Visual Scanning (*Technical Report ARL-01-14/NASA-01-7*). Savoy, IL: University of Illinois, Aviation Research Lab.

A model for an innovative Lane Change Assistant HMI

Daniele Pinotti[1], Fabio Tango[2], Maria Giulia Losi[1], & Marco Beltrami[3]
[1]RE:Lab s.r.l., Reggio Emilia
[2]Centro Ricerche Fiat, Orbassano
[3]AEB Technologies, Cavriago
Italy

Abstract

Millions of people are killed every year in road accidents, caused mostly by human errors. Therefore, it is of paramount importance to support the driver by improving the interaction between humans and machines. This paper presents a research aimed at developing a Cooperative Lane Change Assistant Human Machine Interface (HMI). The system checks whether the adjacent lane is free: if positive, it will suggest the lane-change manoeuvre via a recommendation HMI; if negative, it will show a warning-based HMI. In addition, an acoustic warning will be provided if a critical event is detected. As preliminary test, 20 users carried out two sequences of 18 lane-changes on a static driving simulator: the first run served as a baseline while the second added a secondary cognitive task to induce distraction. Eye-tracking and vehicle data were collected throughout the test and will be used to develop a Driver Cognitive Distraction Model. This classifier will be included into the HMI strategy; further tests have been planned to evaluate the efficiency and the acceptability of this assistant HMI, which can be regarded as a co-driver for the driving lateral task. This work is in the interest of the ARTEMIS Joint Undertaking under the number 269336 (www.d3cos.eu).

Introduction

The lane change manoeuvre is considered one of the most difficult driving tasks and special attention is needed: it is estimated that crashes resulting from improper lane changes constitute almost 8% of all car crashes (Fitch & Hankey, 2012). A common kind of dangerous situation may occur if the driver underestimates the speed of an approaching vehicle or overlooks it at all. It has been shown that drivers' perception of inter-vehicle distances are often insecure, especially at high velocities (Roelofsen et al., 2010). Lane Change Assistant (LCA) systems may help human drivers in avoiding severe accidents. A LCA monitors adjacent lanes and keeps the driver informed of the presence of other vehicles nearby. Typically, these systems don't perform any automated action to prevent a lane change, though such a feature is expected in the future (Visvikis et al., 2008).

Furthermore, in the event of conflicting resources (i.e. a section of a lane aimed by two vehicles) a cooperative LCA could considerably improve the handling of

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

resources, boosting a beneficial collaboration between drivers (Heesen et al., 2012). Such cooperation could be achieved via a Vehicle-to-Vehicle (V2V) communication, which would allow an exchange of information between involved vehicles or also between their assistance systems. Moreover, cooperation among road users should also be associated with an efficient cooperation between the LCA system and the human driver (Flemisch et al., 2008), meaning that a proper HMI needs to be built. The system could also provide additional support if it was able to improve the involved drivers' situation awareness (Endsley, 1996). Many experiments with driving simulators assess the impact of cognitive distraction on the driving performance. Strayer and Johnston (2001) found that drivers engaged in phone conversations were more likely to miss road signs, plus their reaction time was superior with respect to the ones that weren't on the phone. Reported effects are similar whether mobile phones were held in hand or in hands-free configuration (speakerphone): this indicates that the decrease in the driving performance wasn't due to a physical limitation, but rather to a different cognitive load induced by a phone conversation. Then, the field of research was extended to the larger domain of the in-vehicle interfaces: Lee et al. (2001) reported a 300 ms delay in the braking of the vehicle for drivers who were using a voice-based email system while performing a primary driving task on a simulator. Recently, the growing number of In Vehicle Information Systems (IVIS) on the market, such as radios, GPS navigation systems etc., threatens to increase dangerously the sources of distraction. Therefore, to allow the drivers to take advantage of the IVIS without reducing safety becomes a critical topic (see also Liang et al., 2007).

In this paper, an innovative Cooperative-LCA HMI is presented: it embeds elements to support the driver in performing the lane change manoeuvre as well as a cognitive distraction model to infer a possible state of distraction. The key elements of the proposed system are:

- A cooperative driver model, a set of cooperation algorithms for collision avoidance
- A warning/recommender HMI, a multichannel interface which delivers the information to the driver in the most suitable way
- A Driver Cognitive Distraction Model (DCDM), a binary classifier which infers the cognitive state of the driver

This research is performed within the project D3CoS (Designing Dynamic Distributed Cooperative Human-Machine Systems), whose objective is to develop methods, techniques and tools (MTTs) for system engineers and to embed them in industrial system development processes, to support affordable development of highly innovative cooperative human-machine systems.

Background

Cognitive distraction and Machine Learning approach

As aforementioned in the introduction, despite the complexities of the driving task, it is not unusual to see drivers engaged in various other activities while driving, including talking to passengers or by a cellular-phone, listening to the radio and even reading. Concerns with the use of the IVIS while driving are also becoming increasingly common. All in all, any activity that distracts the drivers or competes for their attention with the driving primary task has the potential to degrade performance and have serious consequences for road safety. In this context, one of the main goals of this research is to support drivers by means of cooperation algorithms for collision avoidance, integrated with classifiers of users' state and in particular of cognitive distraction. Hence, the following research questions are investigated:

- Is it possible to develop appropriate driver's distraction classifiers, which are then used together with D3COS cooperative applications, in order to support users and to make such systems even "smarter"?
- Which techniques and algorithms can be regarded as the most "appropriate" (meaning enough accurate, precise, reliable and feasible)?
- With respect to the current state of the art, how can we improve the methodology and the results possibly achievable?

In order to deal with these topics, many methods have been proposed to evaluate and classify driver's distraction (Tango & Botta, 2009; Mattes, 2003; Young et al., 2003). In particular, Machine Learning (ML) techniques seem to be very appropriated for this type of classification problem. In fact, ML is the technology of searching large volumes of data for unknown patterns. It has been successfully applied in business, health care and other domains (Tan et al., 2005; Baldi & Brunak, 2001). So, the rationale to use ML techniques copes with two aspects:

- From a more "philosophical" point of view, one of the most ambitious goals of automatic learning systems is to mimic the learning capability of humans and their capability of driving is widely based on experience, particularly on the possibility to learn from it.
- From a more technical point of view, collected data about vehicle dynamics and external environment are definitely non-linear. Several studies have proved that in such situations ML approaches can outperform the traditional analytical methods. Moreover, also human's driver mental and physical behaviour is non-deterministic. So, since mental state of the drivers is not observable, no simple measure can index visual and cognitive distractions precisely (Zhang et al., 2004; Liang et al., 2007).

Hence, these types of technology can be applied to build a discrimination model able to capture the differences between drivers' normal and distracted behaviour. This phase is currently under development in the D3COS project, where the different distraction models are prepared and then implemented in the C-LCA application.

Lane Change Assistant systems

Designing a proper interface for a LCA system should comprise two different tasks: how the proper suggestion is presented to the driver and how the state of the system is communicated (Visvikis et al., 2008). Standards recommend that warnings are used to tell which side of the surrounding area is already occupied by another vehicle. Moreover, visual warnings should be clearly distinguishable from any other in-vehicle signal and should be positioned so that the driver is prompted to use mirrors.

LCA systems usually provide a two-level warning: the first level is reserved to cautionary warnings, when the likelihood of a collision is relatively low. The second level is instead a more imminent kind of warning, when dangerous situations are more likely to occur (Visvikis et al., 2008). Visual, acoustic or haptic channels may be used, but visual warnings are mostly suitable for a low priority information since they rely on the fact that the driver is looking at the display. On the other hand, acoustic warnings can draw the attention of the driver regardless of where he/she is looking and are therefore more appropriate for urgent warnings.

Furthermore, the driver might assume that the LCA system is always active throughout the journey. It is possible, tough, that the system is occasionally inactive, for example if the driver forgot to activate a manual system or if the criteria for an automatic activation were not met. Therefore, the driver should be constantly kept aware of the state of activation of the system in order to prevent any misinterpretation; likewise, it is also important to notify the driver if the system is temporarily unavailable, for example in case a sensor is covered with dust or snow.

Concept

The system presented in this paper is an innovative Cooperative Lane Change Assistant HMI, aimed at providing the driver with a twofold information: on one hand the system will give a warning if it detects a dangerous situation or a state of distraction of the driver; on the other hand it will suggest a lane change manoeuvre if a gap on a more suitable lane is available, according to a cooperative driver model. The driver model is built on the bases of a highly cooperative approach utilizing methods from the field of distributed artificial intelligence and multi-agent systems. This model enables to plan the trajectory of the vehicles moving on a highway and cooperatively checks them for conflicts: if a collision was detected, the trajectories of the vehicles would be adjusted to be conflict-free. A detailed description of this model is beyond the scope of this paper and can be found in Vokrinek et al. (2013).

The development of the HMI logic and images was driven by the research background reported in the previous section and was supported by pilot tests on a driving simulator, which are described in the next sections. The visual interface is integrated in the simulator dashboard, between the speedometer and the RPM needles, in order to minimize the deviation of driver's gaze from the road. The acoustic interface is played by the simulator speakers.

Figure 1. Sample pictures of the visual HMI: a) "lane change" recommendation; b) "keep your lane" warning; c) LCA inactive.

In case the driver is proceeding on the lane suggested by the system only a white straight arrow is displayed, whereas a red box is added in case the indicator is turned on (fig. 1b) in order to alert the driver that he/she is about to perform an unsafe lane change; it is important, in fact, to prevent needless or distracting warnings during normal driving (Visvikis et al., 2008), so for those cases a minimum warning level was chosen. If instead the system is recommending a lane change, a green box will highlight a free gap onto the suggested lane (fig. 1a): this particular recommendation doesn't need any trigger or input from the driver (proactive behaviour).

The acoustic channel is instead launched if one of the following conditions becomes true:

- Two vehicles are about to collide (mostly because one of the drivers isn't following the shared plan, for example by attempting a lane change towards an occupied lane)
- A state of distraction is detected, in which case a sound is emitted with the purpose of alerting the driver

In particular, the state of cognitive distraction will be inferred by the classifier which is being developed by means of ML techniques: whenever a "TRUE" condition is detected, the HMI strategy is consequently adapted, emitting a sound specifically designed to warn a driver e.g. on a wide frequency range in order to be easily audible and clearly distinguishable from background noise.

The cooperative model constantly computes an updated plan and, at the minimum, the visual channel is always active so that the driver has at any moment the possibility to check his/her lane positioning. However, if necessary input data are not available, then an HMI-OFF image is displayed (fig. 1c) in order to keep the driver aware of the status of the system.

Experimental design

Description of the simulator

The experiments whose setup is presented in this paper were carried out using a static driving simulator owned by Reggio Emilia Innovazione (http://www.reinnova.it/en) and equipped with a SCANeR™ simulation engine (http://www.scaner2.com). The system is a fixed based simulator that comprises a mock-up of a car with real driving controls, specifically a seat, steering wheel, pedals, gear, handbrake and a digital simulated dashboard displaying a traditional instrumental panel, with RPM, speedometer and vehicle subsystem lamps. The scenario is projected on a frontal screen at the driver point of view, together with the rear mirror displayed on the top of the field of view and the wing mirrors on the sides. Environmental sounds, such as the engine rumble, are provided through the loudspeakers.

Distraction Model

A set of experiments on the driving simulator was carried out, involving 20 users in the age between 20 and 40 (Mean = 28.9, SD = 3.8 years). Participants were 14 males and 6 females, with a minimum driving experience of two years and at least 10,000 Km per year of travelling. The simulated scenario represented a highway and, after a five minutes warm-up period to let the user acquire confidence with the simulator, participants were instructed to:

- keep a fixed speed (130 Km/h)
- stay on starting lane, whenever possible

For the primary driving task, the reference test protocol is the Lane Change Task by Mattes (2003). With respect to this protocol a modification of the triggering event has been applied, corresponding to the introduction of one vehicle that appears in front of the Ego vehicle instead of a sequence of road signs suggesting a lane change. The reason of this adjustment was to ensure consistency with the scenario that will be used to test the C-LCA. A secondary task was then introduced to induce cognitive distraction: while driving, participants were asked to count backwards by 7 starting from a given number, suggested by the interviewer (e.g. 100, 93, 86, 79…). Every participant performed two driving session, each consisting of 18 LC manoeuvres:

- Run 1 (A1): LC without secondary task, serving as baseline for the driver cognitive distraction model
- Run 2 (A2): LC with secondary task

The time interval between two subsequent LC manoeuvres (15-40 s) was randomized between the participants according to 8 different combinations of the 18 manoeuvres, both for A1 and A2. Each participant had a different combination in A1 and A2. Randomization was aimed to minimize learning and interaction effects among the different sessions.

Figure 2. The driving simulator used for the experiments.

Throughout the test, vehicle data (such as speed, lateral position, steering wheel…) were collected by the Recorder module of the simulation software at a sample rate of 20 Hz. In addition, users were equipped with a Tobii® Glasses eye-tracker system (http://www.tobii.com) to collect gaze position and in particular three different Areas Of Interest (AOI) were defined: dashboard, rear mirror, central area (fig. 3).

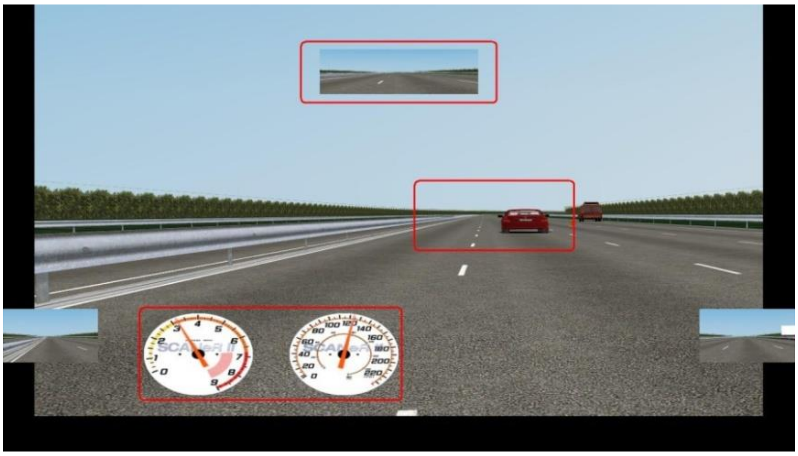


Figure 3. Screenshot of a lane change manoeuvre illustrating the AOI under investigation.

The distraction classifiers of users are being developed by using ML techniques (such as ANN, SVM, etc.). When the DCDM will be fully developed, its output will be included as an input for the LCA system: this way, different HMI strategies can be chosen depending on the distraction state of the driver.

Cooperative HMI

With the purpose of collecting valuable feedbacks for the development of the HMI, 10 participants (mean age = 34.3, SD = 6.7) were involved in a set of pilot tests. As preliminary task, the users were asked to fill out some questionnaires to evaluate Sensation Seeking (Arnett, 1994) and Locus of Control (Montag & Comrey, 1987), and also a socio-anagraphic form to assess users' driving experience and confidence with IVIS (such as radio or GPS). Then the interviewer briefly explained the operation and purpose of the LCA system under investigation, also by showing the screenshots of the visual interface; five minutes of free driving were then used as warm-up before the beginning of the test.

The scenario represented a 3 lanes highway and the users were instructed to drive on the central lane, following a lead vehicle at a distance that they considered safe. After a time interval of 5 to 40 s (randomized between the manoeuvres) of car following, the lead vehicle slowed down: other cars were in transit around the Ego vehicle and the user had to choose between changing lane or slowing down in turn. This manoeuvre was repeated 9 times for each of the two arranged driving sessions, the first of which was carried out without the LCA system, serving as baseline. Finally, participants were asked to fill out a questionnaire to evaluate their confidence and acceptance of the system.

Pilots evaluation

The main purpose of the pilots was to have a halfway evaluation from the users on the HMI, especially on the suitability of the graphics and of the acoustic warning. Besides, pilots also provided useful suggestions on the rationale and on the operating parameters of the system, which were taken into account for the following cycle of development. Participants' response to usability questionnaires regarding the Visual Interface (VI) and the Acoustic Interface (AI) is showed in figure 4, on a scale from 1 to 7.

Users were asked to evaluate the appropriateness (whether or not the interface fit its purpose) and the clarity (whether or not the intended message was clearly received) of the VI (fig. 4a) and the AI (fig. 4b). In particular, the evaluation of the AI suggested a replacement of the warning sound with a more alerting one, which will be implemented in the next version of the system.

Participants also gave some free-comments regarding their perceived confidence in the system and in particular the proactive behaviour previously described was appreciated by the users. As final question, 8 out of 10 participants answered that they would prefer to drive using this C-LCA system with respect to the baseline situation.

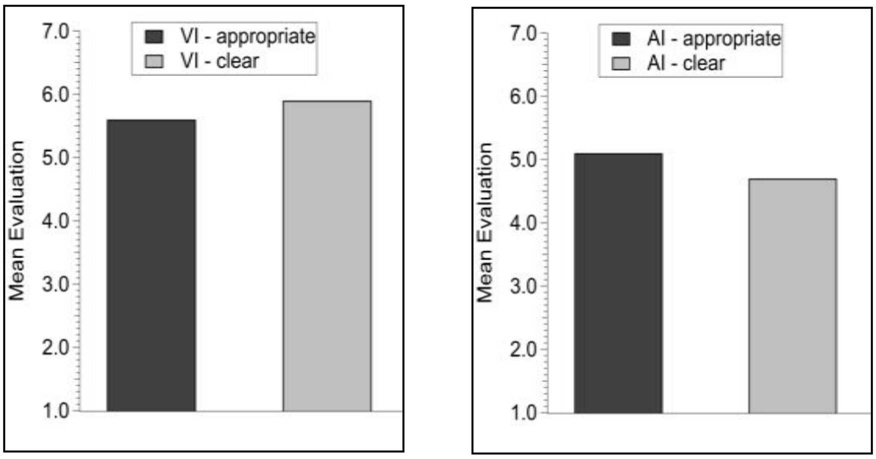


Figure 4. Mean evaluation: left: Visual Interface; right: Acoustic Interface, on a scale from 1 to 7.

Conclusions

This paper presents the concept of an innovative C-LCA HMI and outlines an experimental strategy having the purpose of supporting and validating the development process. Vehicle and eye-tracker data are being used to feed the distraction classifiers by means of different ML techniques in order to develop a DCDM: its output will then be included as an additional input to the HMI logic, so that different warning or recommending strategies may be chosen depending on driver's state. Pilot tests indicated that the users appreciate this kind of assistant systems: they are aware of the risks of highway traffic and they would rather have a LCA on their car. Comments by the pilots' participants will be taken into account to tune some parameters of the system, such as some threshold values, and their evaluation of the HMI indicated that the warning sound should be replaced with a more alerting one. Further experiments will be carried out as soon as the revised version of the C-LCA is fully developed: its effectiveness in improving the driving performance will be the main subject of the next investigations within the D3CoS project, as well as the user acceptance and trust.

Acknowledgement

This research has been funded by the ARTEMIS Joint Undertaking in the interest of the project D3CoS (http://www.d3cos.eu) – Grant Agreement 269336.

References

Arnett, J. (1994). Sensation seeking: a new conceptualization and a new scale. *Personality and Individual Differences, 16*, 289-296.

Baldi, P. & Brunak, S. (2001). *Bioinformatics: The Machine Learning Approach, 2nd edition*. Cambridge, MA: The MIT Press.

Endsley, M.R. (1996). Automation and situation Awareness. In R. Parasuramann and M. Mouloua (Eds.) *Automation and Human Performance: Theory and Applications* (pp. 163-181). Mahwah, NJ: Lawrence Earlbaum.

Fitch, G.M. & Hankey, J.M. (2012). Investigating improper lane changes: driver performance contributing to lane change near-crashes. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2012* (pp. 2231-2235).

Flemisch, F., Kelsch, J., Löper, C., Schieben, A., Schindler, J. & Heesen, M. (2008). Cooperative control and active interfaces for vehicle assistance and automation. *ATZ – Automobiltechnische Zeitschrift*, 9.

Heesen, M., Baumann, M., Kelsch, J., Nause, D. & Friedrich, M. (2012), Investigation of Cooperative Driving Behaviour during Lane Change in a Multi-Driver Simulation Environment. In D. de Waard, K. Brookhuis, F. Dehais, C. Weikert, S. Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier (Eds.) (2012), *Human Factors: a view from an integrative perspective. Proceedings HFES Europe Chapter Conference Toulouse 2012* (pp. 305-317).

Lee, J.D., Caven, B., Haake, S. & Brown, T.L. (2001). Speech-based interaction with in-vehicle computers: the effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors, 43*, 631–640.

Liang, Y., Reyes, M.L. & Lee, J.D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems, 8*, 340–350.

Mattes, S. (2003). The lane change task as a tool for driver distraction evaluation. *IHRA-ITS Workshop on Driving Simulator Scenarios*, Dearborn, MI.

Montag, I. & Comrey A.L. (1987). Internality and externality as correlates of involvement in fatal driving accidents. *Journal of Applied Psychology,72*, 339-343.

Roelofsen, M., Bie, J., Jin, L. & van Arem, B. (2010). Assessment of safety levels and an innovative design for the Lane Change Assistant. In *Intelligent Vehicles Symposium (IV), 2010 IEEE* (pp. 83-88).

Strayer, D.L. & Johnston, W.A. (2001). Driven to Distraction: Dual-Task Studies of Simulated Driving and Conversing on a Cellular Telephone. *Psychological Science, 12*, 462-466.

Tan, P.-N., Steinbach, M. & Kumar, V. (2005). *Introduction to Data Mining*. Boston, USA: Pearson Addison Wesley.

Tango, F. & Botta, M. (2009). Evaluation of Distraction in a Driver-Vehicle-Environment Framework: an application of different Data-mining techniques. In P. Perner (Ed.), *Advances in Data Mining. Applications and Theoretical Aspects* (pp. 176-190). Berlin: Springer Berlin Heidelberg.

Visvikis, C., Smith, T.L., Pitcher, M. & Smith, R. (2008). *Study on lane departure warning and lane change assistant systems* (PPR 374). Wokingham, UK : Transport Research Laboratory.

Vokrinek, J., Janovsky, P., Faigl, J., Benda, P., Tango, F. & Pinotti, D. (2013). A cooperative driver model for traffic simulations. In *Industrial Informatics (INDIN), 2013 11th IEEE International Conference* (pp. 756-761).

Young, K., Regan, M. & Hammer, M. (2003). *Driver Distraction: a review of the Literature* (Technical Report No 206). Monash, AUS: Monash Injury Research Institute.

Zhang, Y., Owechko, Y. & Zhang, J. (2004). Driver cognitive workload estimation: a data-driven perspective. In *Intelligent Transportation Systems, 2004. Proceedings. The 7th International IEEE Conference* (pp. 642-647).

Debiasing time saving judgements by manipulation of speed display

Gabriella Eriksson[1,2]*, Ola Svenson*[1,3]*, & Lars Eriksson*[2,4]
[1]*Department of Psychology, Stockholm University, Sweden*
[2]*Swedish National Road and Transport Research Institute, Linköping, Sweden*
[3]*Decision Research, Eugene, Oregon, USA*
[4]*Department of Psychology, Karlstad University, Sweden*

Abstract

The time saving bias predicts that when increasing speeding from a high speed (e.g. 100 kph) the time saved is overestimated and underestimated when increasing speed from a slow speed (e.g. 30 kph). An alternative meter indicating the inverted speed (min/km) was used to debias time saving judgements in an active driving task. The simulated driving task was to first drive a distance at a given speed and then drive the same distance again at the speed the driver judged was required in order to gain exactly three minutes in travel time compared to the first drive. A control group performed the same task with a speedometer and saved less than the targeted three minutes when increasing speed from a high speed and more than three minutes when increasing from a low speed, as predicted by the time saving bias. Participants in the alternative meter group were closer to the target. The study shows that biased intuitive judgements can be affected by changing information format.

Introduction

Drivers often need to arrive at their destination at a specific time. If a delay occurs due to unforeseen circumstances such as a traffic jam, the driver needs to make judgements of how fast he or she has to drive in order to arrive on time. Previous research has shown a time saving bias in judgements of time saved when increasing speed ((Eriksson, Svenson & Eriksson, 2013; Fuller et al., 2009; Peer, 2010a,b, 2011; Peer & Rosenbloom, 2013; Peer & Gamliel, 2012, 2013; Peer & Solomon, 2012; Svenson, 1970, 1973, 2008, 2009). The time saving bias implies that the time saved when increasing speed from an already high speed is overestimated. Contrarily, the time saved by a speed increase from a low speed is underestimated.

The actual time saved can be obtained by the following formula:

$$\text{Time gain} = cD\,(1/v_1 - 1/v_2) \qquad (1)$$

c is a constant used for conversion of the units of the distance measure, D is the distance travelled, v_1 is the original speed and v_2 the higher speed. According to the formula, an increase from a lower speed has a greater impact on the time saved than the same increase from a higher speed. Svenson (1970) investigated intuitive

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

judgments of the difference in travel time between two speeds, a high and a low speed, over the same distance. The study corroborated a time saving bias and the time saving judgements could be described with the following formula:

$$\text{Time gain} = cD^{e}\,(v_2 - v_1)/\,v_2 \qquad (2)$$

in which c and e are constants describing the perceived distance as a function of the objective distance D; v_1 is the original speed and v_2 the higher speed. A possible interpretation is that participants arrive at their time saving judgements by estimating the speed increase as a ratio or percentage of the faster speed which is a simplification of the physical curvilinear relationship between speed and time. An alternative description of the heuristic has been given by Peer and Gamliel (2012) who found that replacing v_2 in the denominator with v_1 provided a better fit to their data.

The miles per gallon (MGP) illusion is a bias that resembles the time saving bias. In a study by Larrick and Soll (2008), participants made judgements about the fuel consumption of different cars in the U.S. In the U.S., fuel efficiency is measured in miles per gallon, instead of litres per kilometres. The relationship between the amount of fuel consumed and a car's fuel efficiency, when expressed in miles per gallon, is curvilinear. Larrick and Soll (2008) found that participants judged the relationship as linear and therefore made inaccurate judgements of fuel consumed over a given distance when comparing two cars. They also found that judgements were more accurate when the information about fuel efficiency was formulated in terms of amount of gas consumed per given distance.

Peer and Gamliel (2013) conducted a questionnaire study to investigate if the time saving judgements would be more accurate if information about the inverted speed (pace) was given, similarly to the MGP illusion example. In their study, participants were given information about the inverted speed (pace) in minutes per 10 miles and speed in miles per hour. The study showed that judgements were more accurate when pace data were presented alongside the speed. In their study, participants were presented to the information and the problems in an online survey when they were not driving.

In the present study, we wanted to investigate if time saving judgements can be debiased in an active driving task by using a technique external to the decision maker, such as an information display proposed by Larrick (2004). In this study, we use a pace display informing participants instantaneously of the inverted speed in minutes per kilometre. The present study differs from the online survey by Peer and Gamliel (2013) in that in their study participants were given speed information as well as pace data, but in the present contribution participants had to solely rely on pace information so that any effects can be attributed to the pace information alone.

In the present contribution, we tested the hypothesis that the drivers' speed judgements would be debiased by presenting a pace meter indicating the inverted speed.

Method

Participants

There were twelve participants in the study. They were recruited from the participant pool of the Swedish National Road and Transport Research Institute in Linköping, Sweden. Participants were randomly assigned to three age groups with two men and two women in each group; 25-34 (M = 29.75 years, SD=3.40), 35-44 (M = 38.00 years, SD=2.16) and 45-54 (M = 49.50 years, SD=4.12). The majority of the participants had an average annual mileage of between 1500 and 2000 kilometres. All participants had a driver's license, all of them had had their license for at least five years. The participants recieved 500 SEK (about € 50) for their participation.

Apparatus

Simulator
An interactive fixed-base driving simulator at the Swedish National Road and Transport Research Institute in Linköping, Sweden was used in the study. The visual system consists of three 40 inch NEC M40-AV TFT displays with 1 920 x 1 200 pixel resolution, providing a field of view of 180 degrees. The seat, steering wheel, pedals and dashboard are from a real car.

Pace meter
A pace meter that provided instantaneous information of the inverted speed in minutes per kilometre was used (Figure 1). As the speed increases, the pointer of the meter moves from left to right, but the numbers indicated on the meter decreases from left to right (as speed increases it takes less time to travel one kilometre). The numbers indicated on the meter ranged from three minutes per kilometre up to 0.5 minutes per kilometre and were presented with .5 intervals. The movement of the pointer decreases as the speed increases, illustrating that the time saved diminishes with higher speeds.

Figure 1. Pace meter indicating the inverted speed in minutes per kilometre.

Scenario

The simulated road environment presented to the participants was a two-lane rural road. There was no traffic in the participant's lane and light traffic in the opposite direction. Two driving distances were simulated; one was 8.5 and the other 28.3 kilometres. The shorter driving distance was the first 8.5 kilometres of the longer distance, so both distances were parts of the same road.

Experimental design

A within participant design was used, each participant drove each of the two distances twice. Time was held constant so the driving distances were chosen so that the time spent on each of the two distances was the same. Half of the participants drove the shorter distance at the lower speed first and half drove the longer distance at the higher speed first. Age group and sex were balanced in both conditions (shorter distance/low speed first and longer distance/high speed first). A control group who performed the same task and drove the same distances in the same simulator with a conventional speedometer indicating speed in kph (Eriksson, Svenson, Eriksson, 2013) was used to allow for between groups comparisons.

Procedure

Participants first answered questions about background information such as age and years with a driver's license. Then, they were given an instruction of the experiment in writing. Before the simulator task began, participants engaged in a practice run to adapt to driving in the simulator. The practice run lasted for 15 minutes and was driven on the same road as the one used in the driving task and at any speed chosen by the participant.

After the practice run, participants were first told to drive either the shorter or the longer distance at a given speed (30 kph for the shorter distance and 100 kph for the longer distance). Participants were then asked to drive the same distance again at the speed they thought necessary to gain exactly three minutes in travel time. Participants were then asked to answer a set of questions regarding the simulator drive such as how much time they thought they had saved and their average mean speed (in kph). The simulator task was then repeated with the second distance and speed and after the task was finished participants answered the same set of questions regarding time saved and mean speed.

After the simulator task, participants were given a questionnaire with further time saving problems. In order to compare the perceptual motor task of driving to paper-and-pencil studies of time savings, two of the problems corresponded to the two simulator driving tasks. Participants were given the distance (8.5 and 28.3 respectively) and initial speed (30 and 100 kph) and were asked to estimate the mean speed required to gain three minutes in travel time.

Results

Participants estimated how much time they thought that they had saved in both conditions. Their judged time saving and the actual time that they saved are shown in Table 1 (pace meter group) and Table 2 (speedometer group). The judged time saving when increasing speed from 30 kph was 3.38 minutes compared to the actual time saved, 4.42 minutes in the pace meter conditon. When increasing speed from 100 kph, the judged time saved was 3.34 minutes compared to the actual time saved 3.47 minutes. In both conditions, participants judged their time saving as close to the targeted three minutes.

Table 1. Judged, actual and targeted time savings for the pace meter group.

Original speed (km/h)	Judged	Actual	Target
30	3.38 (.80)	4.42 (2.39)	3
100	3.34 (.94)	3.47 (1.44)	3

Note. All entries are expressed in minutes. Standard deviations in parentheses.

Table 2. Judged, actual and targeted time savings for the speedometer group (Eriksson, Svenson & Eriksson, 2013).

Original speed (km/h)	Judged	Actual	Target
30	3.56***(.90)	6.14 (1.86)	3
100	3.25*(.72)	2.21 (.85)	3

Note. All entries are expressed in minutes. Standard deviations in parentheses. Significant deviation of mean (judged) from actual time savings *p<.05, **p<.01 and ***p<.001.

The average actual time saved in the low speed condition was 4.42 minutes and not significantly different from the targeted three minutes, t_{11}=2.058. p = 0.064, two-tailed. In the high speed condition, the average actual time saving was 3.47 minutes which was not significantly different from the targeted three minutes, , t_{11}=1.133 p = 0.281, two-tailed. In the control study (Eriksson, Svenson & Eriksson, 2013), participants completed the same driving task with an ordinary kph speedometer and the actual time savings were significantly greater than three minutes when increasing from 30 kph (t_{11}=5.853, p = 0.0001, one-tailed) and significantly lower when increasing from 100 kph (t_{11}=3.228, p=0.004, one-tailed), which corroborates a time saving bias.

Figure 2 shows a comparison between the actual time saved in the control (speedometer) study and the present (pace meter) study. A Mann-Whitney U test was used to test the difference in actual time saved between the speedometer and pace meter group. In the high speed condition, the speedometer group saved significantly less time than the pace meter group (U = 34, Z = -2.194, p = 0.014, one-tailed). When increasing from the low speed, the speedometer group saved significantly more time than the pace meter group (U = 41, Z = -1.790, p = 0.039, one-tailed).

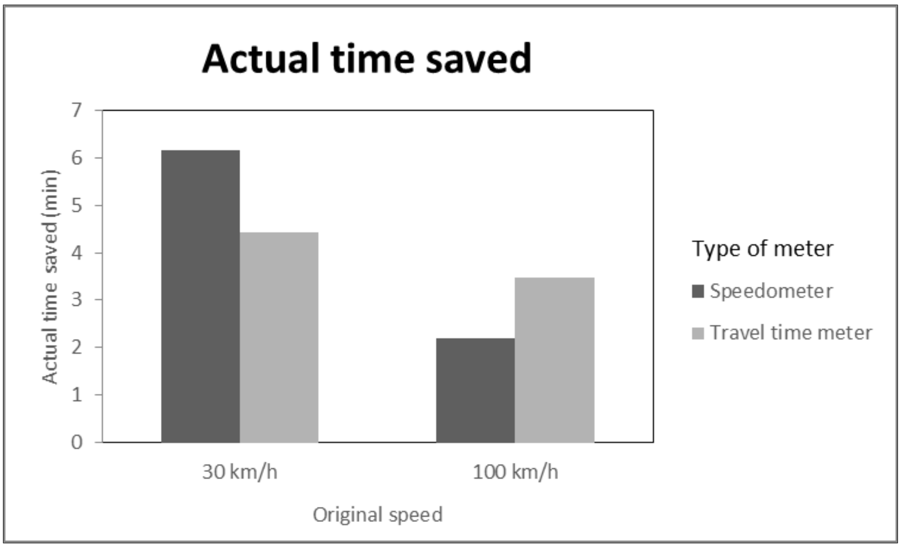


Figure 2. Acutal time saved (min) in both low and high speed conditions for the speedometer meter and the pace group.

Although the actual time savings differed depending on whether the speed was increased from a low or a high speed, the judged time savings did not. Figure 3 shows the judged time saved for both groups and conditions.

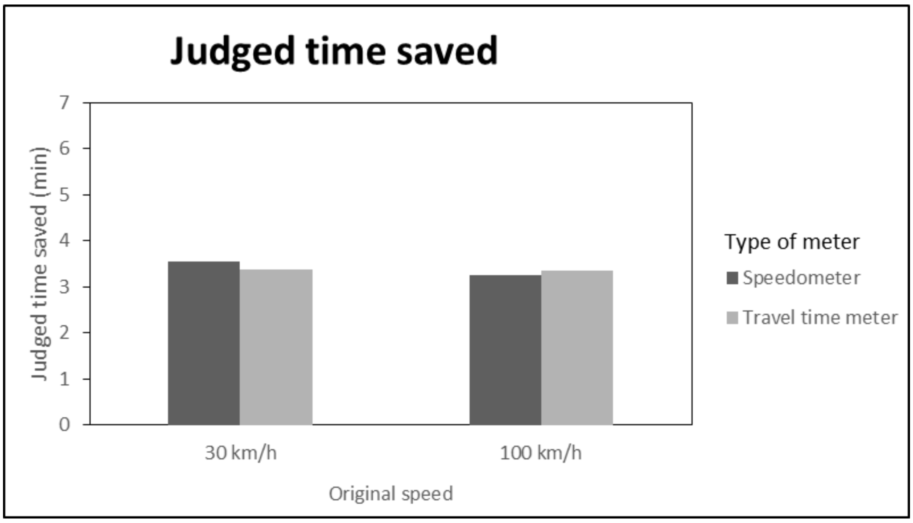


Figure 3. Judged time saved (min) in both low and high speed conditions for the speedometer meter and the pace group.

The mean speed estimates given in the questionnaire and the actual mean speeds chosen in the simulator tasks for both the speedometer and pace meter group are shown in Table 3. Mean speed estimates made in the questionnaire were similar to the mean speeds in the simulator task, except at the higher speed condition where participants in the pace meter group drove closer to the targeted mean speed in the simulator task.

Table 3. Participants' questionnaire estimates of mean speeds, actual simulator driving mean speeds and correct mean speeds for the two distances.

Type of meter	Speedometer	Pace meter
8.5 km at original speed 30 kph		
Questionnaire estimate	49.46	40.33
Chosen speed in simulator task	45.51	42.26
Correct speed	36.43	36.43
28.3 km at original speed 100 kph		
Questionnaire estimate	113.63	115.08
Chosen speed in simulator task	113.52	123.95
Correct speed	121.46	121.46

Discussion

A previous study by Eriksson, Svenson and Eriksson (2013) showed that participants make biased judgements of the time saved by increasing speed from a high and low speed in an active driving task. The present study showed that these intuitive biased judgements can be debiased by presenting a pace meter which shows the inverted speed in minutes per kilometre instantaneously. Although intuitive judgements of time savings have shown a robust bias (Eriksson, Svenson & Eriksson, 2013: Peer, 2010a, 2010b, 2011; Peer & Gamliel, 2012, 2013; Peer & Rosenbloom, 2013; Peer & Solomon, 2012; Svenson, 1970, 1973, 2008, 2009), the present study indicates that driving behaviour can be altered by providing in-car information adapted to the drivers' shortcomings.

A driver who overestimates the time saved of a speed increase from a high speed may experience time pressure when he or she realises that time is running short. The driver may then be tempted to increase speed even further to reach the destination on time. Also, drivers who realise how little time that actually can be saved at higher speeds may be less motivated to choose a higher speed in the first place. Hence, drivers need to be aware of the effect of speed on time saved in order for them to make optimal choices about their mean speed over a journey and debiasing such judgements may lead to a decreased mean speed which has a positive effect on road safety.

References

Eriksson, G., Svenson, O., & Eriksson, L. (2013). The time saving bias: Judgements, cognition and perception. *Judgment and Decision Making, 8*, 492-497.

Fuller, R., Gormley, M., Stradling, S., Broughton, P., Kinnear, N., O'Dolan, C., & Hannigan, B. (2009). Impact of speed change on estimated journey time: Failure of drivers to appreciate relevance of initial speed. *Accident Analysis and Prevention, 41*,10-14.

Larrick, R. P. (2004). Debiasing. In D.J. Koehler and N. Harvey (Eds.), *Blackwell Handbook of Judgment and Decision Making* (pp. 316-337). Oxford: UK.

Peer, E. (2010a). Speeding and the time-saving bias: How drivers' estimations of time saved when increasing speed affects their choice of speed. *Accident Analysis and Prevention, 42*, 1978–1982.

Peer, E. (2010b). Exploring the time-saving bias: How drivers misestimate time saved when increasing speed *Judgment and Decision Making, 5*, 477–488.

Peer, E. (2011). The time-saving bias, speed choices and driving behavior. *Transportation Research Part F: Traffic Psychology and Behaviour, 14*, 543–554.

Peer, E., & Gamliel, E. (2012). Estimating time savings: The use of the proportion and percentage heuristics and the role of need for cognition. *Acta Psychologica, 141*, 352-359.

Peer, E. & Gamliel, E. (2013) Pace yourself: Improving time-saving judgment when increasing activity speed. *Judgment and Decision Making, 8*, 106-115.

Peer, E., & Rosenbloom, T. (2013). When two motivations race: The effects of time-saving bias and sensation-seeking on driving speed choices. *Accident Analysis and Prevention, 50*,1135-1139.

Peer, E., & Solomon, L. (2012). Professionally biased: Evidence for misestimations of driving speed, journey time and time-savings among taxi drivers. *Judgment and Decision Making, 7*, 165–172.

Svenson, O. (1970). A functional measurement approach to intuitive estimation as exemplified by estimated time savings. *Journal of Experimental Psychology, 86(2)*, 204–210.

Svenson, O. (1973). Change of mean speed in order to obtain a prescribed increase or decrease in travel time. *Ergonomics 16*, 777–782.

Svenson, O. (2008). Decisions among time saving options: When intuition is strong and wrong. *Acta Psychologica, 127,* 501-509.

Svenson, O. (2009) Driving speed changes and subjective estimates of time savings, accident risks and braking. *Applied Cognitive Psychology, 23,* 543-560.

Use of a multidimensional approach in the automotive product development: Quality of turn indicator sounds

Verena Wagner & K. Wolfgang Kallus
Karl-Franzens-University Graz, Austria

Abstract

In the automotive industries, it becomes more and more important to connect customer requirements and technical specifications to develop sounds with high quality. The turn indicator sound can be heard very often during drives and gives the customer important feedback on correct function performance. Thus, this is one of the sounds, which play a role in the customer's perception of vehicle quality and which is safety relevant as well. In an experimental laboratory study the question was investigated, how a turn indicator sound must be designed to be perceived as a high-quality and pleasant. A multidimensional approach was chosen to combine subjective assessments, physiological measures of the cardiovascular and the electrodermal activity and physical parameters of the sounds. In total, 15 different turn indicator sounds were assessed by 48 participants. The results show that high-quality turn indicator sounds are characterized by the fact that they are rather gentle, soft, reserved and not too rough and sharp. This can be confirmed by the physiological reactions of the participants. The study shows how the connection of subjective and objective parameters can support product development. Also, it shows a possibility to involve the human factor in a highly technical environment.

Introduction

Recent developments show that the subjective perception of a product and its sound cannot only be described by technical parameters (Genuit et al., 2006). Therefore it becomes more and more important to connect customer requirements and technical specifications (Resch & Mast, 2006). Especially in the automotive industry, where product sound can be seen as a differentiating factor between different brands these connections play an important role (Fastl, 2005; Nor et al., 2008; Otto & Wakefield, 1993; Schifferstein, 2006). Previous research has shown that customers believe instinctively that a high-class product also produces high-quality sounds (Miśkiewicz & Letoweski, 1999; Schulte-Fortkamp et al., 2007). On the other hand, the sound quality of a product can influence the perceived quality of a product (Genuit, 2008).

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Turn indicator sound

The turn indicator sound can be described as an operational vehicle sound with signalling character (c.f. Cerrato, 2009; Mühlstedt et al., 2007; Zeller, 2009). The acoustic feedback shows the driver that the indicator is set even if he does not see the visual information. Because of that the driver does not have to avert his gaze from the street if the traffic situation would not allow it. The turn indicator sound is also one of these sounds that a driver will hear very often during a drive. If it is not created well, it can become aversive for the driver. So it seems to be important to pay great attention on the development of this operational sound.

From a customer's point of view, the turn indicator sound can also be seen as one perceptible aspect that gives customers some information about vehicle quality. Vehicle manufacturers can design a brand specific sound and use it as distinguishing aspect (Bronner, 2007; Kilian, 2007; Krugmann & Langeslag, 2007; Zeller, 2009). Empirical data indicate that the turn indicator sound can be seen as one important representative sound for vehicle interior sounds. This sound also plays an important role for perceived customer's vehicle quality and was pointed out as a sound that polarizes with reference to customer's assessment (cf. Beitz et al., 2010). Based on the results of studies, improvement opportunities in three dimensions of vehicle sound perception "timbre", "loudness" and "roughness/sharpness" are found (cf. Wagner et al., 2009).

Multidimensional approach

The multidimensional impressions, which a product and its sounds may activate in a customer, call for a multidimensional approach to assess the sound quality of the turn indicator. The subjective assessment of the impression of driving, analyses of sounds as well as of noise is described as multidimensional in the literature (Alt & Jochum, 2003; Bodden, 1997; Genuit, 2002; Genuit & Burkhard, 1995; Schulte-Fortkamp, 2010). Genuit and Burkhard (1995) as well as Bodden (1997) claim that different parameters should be taken in consideration for sound evaluation: subjective parameters (psychological part) and objective parameters (physical and psychoacoustic part). A third approach includes psychophysiological measures to assess activation and emotional reactions in addition to subjective customer's assessments. This way of product testing has been used successfully for products which address other senses than the sense of hearing (Boucsein, 2007; Boucsein et al., 2002).

Physiological parameters

Human behaviour as well as cognitive, emotional and social phenomena are accompanied by physiological processes in different physiological systems. These processes can be made visible and measurable with the help of psychophysiological parameters. For emotional tone and activation/arousal parameters of the autonomic nervous system (ANS) are widely used (Boucsein & Backs, 2009). Cardiovascular and electrodermal activity was recorded in the present study. Psychophysiological measurements are not under active control of humans during normal conditions (Boucsein & Backs, 2009) and they are relatively easy to measure using noninvasive measurement techniques (Boucsein, 2006). It is well-known that noise produces different physiological responses which are quite similar to stress-responses, (Babisch, 2002, 2005; Griefahn & di Nisi, 1992; Ising & Kruppa, 2001; Ising et al., 1990) but these aspects should not be part of the present study. The interest lies in physiological response differences which correspond to subjectively assess sound qualities and the impression of pleasantness. An investigation of Bradley and Lang (2000) with different sounds showed that the subjective classification of the sounds in the two dimensions “pleasure” and “arousal” is accompanied by different physiological responses of the participants. Similar results are also shown with the help of investigations in the area of music (Iwanaga & Moroki, 1999; Sammler et al., 2007), traffic sounds (Raggam et al., 2007) as well as with everyday sounds (Gomez und Danuser, 2004). Investigations with other products than vehicles and their sounds showed that it is possible to make emotional experience with a product objectively measurable with the help of physiological measures (Boucsein, 2006; Mandryk & Atkins, 2007). The results of the different research groups agree that more than one single physiological measure is necessary to reflect the different psychophysiological processes of activation, emotions and attention in (sound) perception wherefore a combination of different measures is recommended (Boucsein & Ottmann, 1996; Whang, 2008).

Acoustic and psychoacoustic parameters

Different acoustic and psychoacoustic parameters should be adopted as objective measures in the present study. To cope with the circumstances that a sound as well as its subjective impression cannot be described well by using a single parameter, different parameters were considered in this study. Sounds with strong characteristics in the psychoacoustic parameter sharpness are often assessed as unpleasant, annoying and aggressive (Fastl, 2005; Fastl & Zwicker, 2007; Genuit, 2008; Maschke & Jakob, 2010). At the same time a product sound gets a strong and powerful character if the parameter sharpness is well-balanced (Fastl, 2005). In addition, the parameter loudness is able to influence the perceived quality and the perceived pleasantness of a product sound (Fastl, 1997, 2005; Fastl & Zwicker, 2007; Griefahn & di Nisi, 1992). Strong correlations between the perceived quality of engine sounds and the psychoacoustic parameter impulsiveness are also reported (Hashimoto, 2000).

Aim of the study

The study addresses characteristics which customer assign to a high quality turn indicator sound. Also, the importance of different dimensions of sound perception and the influence of differences due to age and gender are addressed in the present study. A multidimensional approach which combines subjective assessments of the participants as well as psychophysiological measures and physical characteristics of the turn indicator sounds were systematically included in the study.

Method

A laboratory experiment with repeated measurement was chosen for this study. During the experiment the participants were sitting in a sound-isolated acoustical cabin. The sounds were presented via headphones.

Sample

In total, data from 48 participants with normal hearing ability were incorporated in the study. Age and gender of the participants were balanced (male/female, < 35 years old/35 years old and older). The 48 participants were between 21 and 60 years old with an average age of 36.2 years (SD = 11.56). All participants owned a valid driver’s licence and drove regularly.

Materials

In total, 15 different turn indicator sounds were assessed by the participants. Two different lengths of sounds were used in the laboratory experiment. A preliminary investigation with city drives and drives on motorways shows that in average the short sound has to have a length of three to five seconds. Because of this a sound length of four seconds was used. To assess the different sounds, a second lengths of sounds (30 seconds with an on/off-pattern of four seconds) was used: these second lengths of sounds consider that drivers normally hear the turn indicator sound in different situations: for example, it can be heard for a short time during an overtaking manoeuvre or if the driver changes lane, or for a longer time, when the driver wants to turn (e.g. crossroads, traffic light). The preliminary investigation results show as well that during a turning manoeuvre the turn indicator sound was heard by a minimum of 20 seconds, mostly even much longer. Based on these results it was decided to use trials with the parts: short sound period (4 sec.) – pause (10 sec.) – long sound period (30 sec.) for each sound in the second part of the experiment where the participants had to assess each sound. The sequences of the different sounds were randomly assigned.

The questionnaire comprised a broad range of items (7-point rating scales and items based on semantic differentials) addressing different aspects of participant’s evaluations of the turn indicator sounds, items to assess these sounds including items of vehicle sound perception dimensions (cf. Wagner et al., 2009) and also items to specify the emotions and somatic feelings of the participants during the experiment.

Procedure

At the beginning of the experiment, each participant had to render an audiometry to check the hearing abilities of the sample. Only participants with normal hearing abilities were included into the final sample. After a short questionnaire (socio-demographic data and general attitude towards vehicle sounds), the electrodes for the physiological measurement were fixed. After baseline-recordings, all 15 turn indicator sounds (4 s.) were played to show the participants the evaluation framework of the experiment. After a 90 second psychophysiological recording, the turn indicator sounds were presented (4 s. – 10 s. pause – 30 s.) in groups of five sounds, each group followed by a psychophysiological recording (90 s.). When the participant finished the assessment of all 15 sounds a follow-up survey had to be filled out. The psychophysiological recordings took place during rest intervals and during the sound presentations.

Physiological measurement

The physiological recordings were conducted with the varioport system from Becker Meditec. The participants wear the portable recorder in the acoustical cabin during the experiment to record loggings of the trigger as well as cardiovascular (ECG) and electrodermal activity (EDA). ECG was recorded using a thorax lead. EDA was recorded as skin conductance from the inner palm (thenar and hypothenar) of the non dominant hand of the participants (two Ag/Ag-Cl-electrodes filled with 0.5% NaCl paste, diameter = 22 mm; recordings: 0.5 V constant voltages, resolution = 0.002 μS).

The software variograf was used to convert the data. For the elimination of artifacts and the computation of the different measures, different software packages from Boucsein's laboratory were used (Schaefer, 1999, 2000, 2002, 2005). The measures which were used in the statistical analyses were for ECG: heart rate (HR) in beat per minute (bpm) and heart rate variability (HRV) calculated as mean square of successive differences (MQSD) and for EDA: Level EDL, non-specific skin conductance response (NS.SCR), sum-amplitude and mean sum-amplitude. A baseline correction has been done for all measured physiological values. For every physiological measure, two mean values were calculated from the measurements during the trials (short sound period – pause – long sound period) to analyze the physiological effects of each of the turn indicator sounds: one mean value for the 10 seconds of the pause after hearing the short sound and one mean value for the 30 seconds while the participants are listening to the so called long sound.

Sound analysis

The software ArtemiS was used to analyze all 15 turn indicator sounds. All analyzed sounds were binaural recordings of the turn indicator sounds, recorded with an artificial head from HEAD acoustics in the original experimental setting. In this study, different acoustic parameters were used: A-rated sound pressure level [dB(A)], specific loudness [soneGD] (DIN 45631), sharpness [acum] (DIN 45692) and specific impulsiveness [iu]. All parameters represented an average value over each turn indicator sound signal (two-tone-unit).

Statistical analysis

Analyses of variance and regression analyses were performed for calculating the results. A significance level of 5 % was adopted for the results. Due to a descriptive approach no α-correction was conducted (Abt, 1987). The statistical analyses of the data were conducted using the software SPSS 17.0 for Windows.

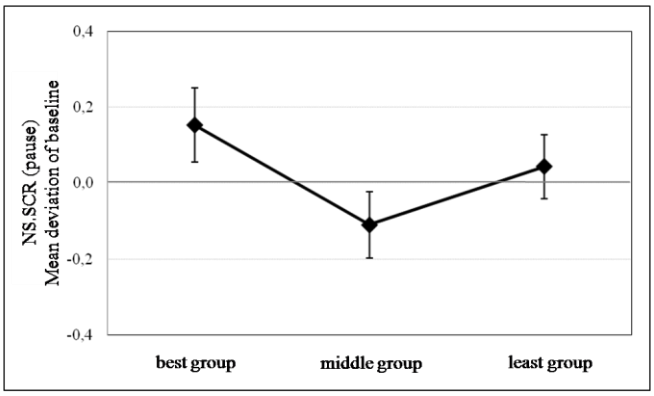
Results

Perceived quality differences of the turn indicator sounds

Based on the results of an analysis of variance with repeated measures three sound groups of significantly differing quality ratings (quality, comfortable/pleasantness) could be formed: best rated group (5 sounds), middle rated group (5 sounds) and least rated group (5 sounds), $F(1.73, 81.22) = 101.64$, $p < .0001$.

Differential effects
No significant differences for the independent variables age (2 groups), $F(1, 44) < 1$, *ns*, and gender (2 groups), $F(1, 44) < 1$, *ns*, as well as for their interaction age x gender, $F(1, 44) = 1.49$, *ns*, can be shown for the factor quality sound groups (3 groups: best, middle, and least assessed).

Psychophysiological measures
Results of the analyses of the physiological parameters and the factor quality sound groups only show significant differences for the electrodermal parameter NS.SCR during the pause between hearing the short and the long sound, $F(2, 43) = 7.03$, $p = .002$ (see figure 1).



Note. Error bars reflect Std. Error of Mean; 0.0 = Baseline level

Figure 1. Differences in electrodermal response for the three quality sound groups

Post-Hoc analyses show significant differences between the electrodermal response during best and middle assessed group (Sidak, $p = .002$), a tendency between middle and least assessed group (Sidak, $p = .053$) and no statistical significant differences between best and least assessed turn indicator group.

The adding of the two independent parameters age and gender into the analyses show a significant main effect of age groups with the cardiovascular parameter HRV (MQSD, heart rate variability) during the pause period, $F(1, 44) = 4.91$, $p = .032$. Older participants react less aroused/distracted (higher HRV) on the different turn indicator sounds than younger participants. Additionally a significant interaction between age and gender, $F(1, 44) = 3.95$, $p = .053$, can be shown with the cardiovascular parameter HRV (MQSD) during the sound period. Younger women react stronger and less relaxed on the different turn indicator sounds than older women and men. Overall, the strongest reactions can be shown for sounds of the least rated sound group. Significant effects with other psychophysiological measures cannot be shown.

Dimensions of sound perception

The mean values for each of the three dimensions “timbre”, “loudness” and “roughness/sharpness” were calculated as our former research shows that both requirements and assessments of vehicle sounds can be represented using these vehicle sound perception dimensions (Wagner et al., 2009). A regression analysis with these three dimensions as regressors and the quality rating score as dependent variable shows that for turn indicator sounds the dimension “timbre” is not significantly related to the quality rating score. In addition, the dimension “loudness” is negatively related and the dimension “roughness/sharpness” is positively related to the quality rating score. So, a high-quality turn indicator sound is characterized by the fact that it is rather gentle, soft and reserved as well as not too rough and sharp. The values of the regression analysis can be seen in table 1.

Table 1. Linear regression analysis for the dimensions predicting the quality rating score

Variable	B	SE B	β
Timbre	-.069	.039	-.054 *ns*
Loudness	-.758	.040	-.636**
Roughness/sharpness	.085	.036	.074*

Note. Adj. $R^2 = .495$. $^{**}p < .0001$, $^{*}p < .05$.

To enable a closer look into the three dimensions of vehicle sound perception and their varieties for the analyzed turn indicator sounds, analyses of variance with repeated measures were calculated based on the subjective assessments in each dimension. These analyses show that the 15 different turn indicator sounds can be assorted into different groups on the basis of their different subjective assessments in each dimension. Multivariate analyses of variance validate the significant differences between the formed groups. The turn indicator sounds were grouped into three significant different assessed groups in the dimension “loudness”, $F(8.37, 393.24) = 22.73$, $p < .0001$, into three significant different groups because of their assessment in the dimension “roughness/sharpness”, $F(9.52, 447.27) = 10.24$, $p < .0001$, and have to be arranged into four significant different assessed groups with reference to the dimension “timbre”, $F(8.98, 422.08) = 37.97$, $p < .0001$. Although the regression analyses shows no significant relevance ($p = .076$) of the dimension “timbre” with regard to the quality rating score, it was included in the following analyses to clear coherences with other measures and this dimension.

Differential effects

Neither for the three different assessed groups in the dimensions “loudness” and “roughness/sharpness”, nor for the four different groups of the dimension “timbre”, significant differences because of age or gender can be shown.

Psychophysiological measures

Significant effects for the different groups of vehicle sound perception on physiological parameters could be show in a repeated measurement MANOVA. Results are shown in table 2.

Table 2. Significant psychophysiological reactions on subjective assessments in the different dimensions of vehicle sound perception

Dimension	Physiological Parameter	*F*	df	*p*
Loudness (3 groups)	Cardiovascular activity:			
	HRV (MQSD) pause period	3.49	2/43	.039
	HRV (MQSD) during sound	6.19	1.75/76.82	.005
Roughness/	Electrodermal activity:			
Sharpness (3 groups)	NS.SCR during sound	3.31	1.74/76.48	.048
Timbre (4 groups)	Cardiovascular activity:			
	HRV (MQSD) pause period	7.87	1.87/82.44	.001
	HRV (MQSD) during sound	5.07	1.53/67.30	.015

For “loudness” and “timbre”, significant cardiovascular reactions can be found while “roughness/sharpness” affects electrodermal activity selectively. Post-hoc analyses show that the subjectively sensed richest, darkest and low pitched sounds (dimension “timbre”) bind least attention and result in fewer arousal/distraction and at the same time are more pleasant than other turn indicator sounds. If a sounds perception is not reserved, gentle and soft (dimension “loudness”), the level of relaxation of the participants decrease. Furthermore, very sharp and rough perceived turn indicator sounds call for more attention, is more distracting and result in larger emotional arousal than other sounds.

Acoustic parameters and subjective sound perception

To clarify connections between the acoustic parameters and the perceived quality of the analyzed turn indicator sounds, a regression analysis with all acoustic parameters as regressors and the quality rating score as dependent variable has been carried out. Because of the relatively strong correlation (r = .752, p = .001) between the two volume related parameters specific loudness and A-rated sound pressure level [dB(A)], two separate regression analyses – one with specific loudness, sharpness and specific impulsiveness and one with A-rated sound pressure level [dB(A)], sharpness and specific impulsiveness were carried out. The results of these two analyses show, that specific loudness results in a better prediction. Specific loudness and sharpness are negatively related to the quality rating score. This means that a turn indicator sound which produces a perception of high quality should be physically characterized by lower values in specific loudness and sharpness. The values of the analyses can be seen in table 3.

Table 3. Regression analyses for acoustic parameters predicting the quality rating score

Variable	B	SE B	β
Regression analysis 1			
A-rated sound pressure level [dB(A)]	-.198	.068	-.630*
Sharpness	-.266	1.083	-.051 *ns*
Specific impulsiveness	-.078	.075	-.238 *ns*
Regression analysis 2			
Specific loudness	-.248	.063	-.856**
Sharpness	-2.431	1.028	-.471*
Specific Impulsiveness	-.069	.063	-.212 *ns*

Note. Regression analysis 1: Adj. $R^2 = .450$. $^{**}p < .01$, $^{*}p < .05$
Regression analysis 2: Adj. $R^2 = .602$. $^{**}p < .01$, $^{*}p < .05$

Discussion

The concentration on a specific source of sound allows a detailed investigation of the customer's reactions on differently sounding turn indicators using a multidimensional approach. The absence of significant perceived quality differences due to age or gender supports the idea of the implementation of a brand sound of the turn signal against a differential approach which considers varying target groups of different vehicle segments. Because of the frequent usage of the turn indicator during normal drive situations, the application of one corporate turn indicator sound into all vehicles of different segments of a company can also be used as additional and easy recognizable item to strengthen brand identity. As well the integration of the results of the physiological measurements supports this recommendation. The results show that due to differences in the perceived quality of the evaluated 15 turn indicator sounds three different groups can be formed (best, middle, least rated group), which can be characterized by different electrodermal and cardiovascular responses. The turn indicator sounds of the middle rated group do not activate emotions as strong as turn indicator sounds which belong to the best or least quality group. Even older participants react less aroused on the different sounds. From a physiological point of view turn indicator sounds of the best rated perceived quality group should be implemented to reach positive emotions of the driver and at the same time prevent mental workload. Sounds which belong to the middle rated group can be used if the intention is to implement a sound which remains in the background. The results advise against the implementation of turn indicator sounds of the least rated group not only because of their poorer values in the perceived quality, but also because these potentially trigger arousal in the direction of negative emotions.

The regression analyses with the acoustic parameters supports the suggestion that a turn indicator sound which produces a perception of high quality can physically be characterized by lower values in the volume related parameters A-rated sound pressure level [dB(A)] and specific loudness. In addition, especially in the combination with specific loudness, the parameter sharpness should be less pronounced. These results correspond with results from Fastl (1997, 2005) and Fastl and Zwicker (2007) who assume that loudness is able to influence the perceived quality of a product sound. Fastl (2005) proposes furthermore a well-balance of sharpness. On the other hand, the importance of specific impulsiveness for perceived sound quality which Hashimoto (2000) reports for engine sounds could not be verified for turn indicator sounds in the present study.

The inclusion of the dimensions of vehicle sound perception (c.f. Wagner et al., 2009) allows deriving recommendations for future sound developments. The regression analyses with the three dimensions of vehicle sound perception show that a high-quality turn indicator sound is rather gentle, soft and reserved as well as not too rough and sharp. This result is supported by the physiological reactions of the participants that show that if the perception of a turn indicator sound is not reserved, gentle and soft, the level of relaxation of the participants decrease. Too sharp and rough perceived turn indicator sounds call for more attention of the listener and result in stronger emotional arousal. The physiological measures support, that subjectively sensed rich, dark and low pitched turn indicator sounds are most pleasant, bind least attention and result in fewer arousal/distraction than other sounds.

To sum up, the results of this study show that a multidimensional approach which brings together different subjective and objective aspects of a sound and its perception, can be seen as an important step to determine further optimization options for the development of high-quality sounds.

References

Abt, K. (1987). Descriptive data analysis: A concept between confirmatory and exploratory data analysis. *Methods of Information in Medicine*, *26*, 77-88.

Alt, N.W., & Jochum, S. (2003). Sound Design Under the Aspects of Musical Harmonic Theory. *SAE International 2003-01-1508*, 1-8.

Babisch, W. (2002). The noise/stress concept, risk assessment and research needs. *Noise and Health, 4*, 1-11.

Babisch, W. (2005). Noise and Health.*Environmental Health Perspectives*, *113*, 14-15.

Beitz, T., Wagner, V., & Enigk, H. (2010). Importance of operational sounds for vehicle quality. In S. Pischinger, J.-W. Biermann, K. Genuit, and M. Vorländer (Eds.), *Proceedings Aachener Acoustics Colloquium* (pp. 33-36). Aachen.

Bodden, M. (1997). Instrumentation for Sound Quality Evaluation. *ACUSTICA – acta acustica*, *83*, 775-783.

Boucsein, W. (2006). Psychophysiologische Methoden in der Ingenieurspsychologie. In B.Zimolong, and U.Konradt (Eds.), *Sonderdruck aus Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete: Serie III Wirtschafts-, Organisations- und Arbeitspsychologie. Band 2: Ingeneurspsychologie* (pp. 317-358). Göttingen: Hogrefe.

Boucsein, W. (2007). Psychophysiologie in der Ergonomie. In K. Landau (Ed.), *Lexikon der Arbeitsgestaltung* (pp. 1045-1047). Stuttgart: GentnerVerlag.

Boucsein, W., & Backs, R.W. (2009). The Psychophysiology of Emotion, Arousal, and Personality: Methods and Models. In V.G. Duffy (Ed.), *Handbook of Digital Human Modeling.Research for Applied Ergonomics and Human Factors Engineering* (pp. 35-1 – 35-18). Boca Raton: CRC Press.

Boucsein, W., & Ottmann, W. (1996). Psyhophysiological stress effects from the combination of night-shift work and noise. *Biological Psychology, 42*, 301-322.

Boucsein, W., Schaefer, F., Kefel, M., Busch, P., & Eisfeld, W. (2002). Objective emotional assessment of tactile hair properties and their modulation by different product worlds. *International Journal of Cosmetic Science, 24*, 135-150.

Bradley, M.M., & Lang, P.J. (2000). Affective reactions to acoustic stimuli. *Psychophysiology, 37*, 204-215.

Bronner, K. (2007). Schöner die Marken nie klingen … Jingle all the Way? Grundlagen des Audio-Branding. In K. Bronner, and R. Hirt (Eds.), *Audio-Branding. Entwicklung, Anwendung, Wirkung akustischer Identitäten in Werbung, Medien und Gesellschaft* (pp. 82-96). Praxisforum Medienmanagement, Band 5,2. Nachdruck, München: Verlag Reinhard Fischer.

Cerrato, G. (2009). Automotive Sound Quality – Accessories, BSR, and Brakes. *Sound and Vibration, Sept.*, 10-15.

DIN(1991). *DIN 45631:1991-03, Berechnung des Lautstärkepegels und der Lautheit aus dem Geräuschspektrum; Verfahren nach E. Zwicker*. *Procedure for calculating loudness level and loudness*. Berlin: Beuth.

DIN (2009). *DIN 45692:2009-08, Messtechnische Simulation der Hörempfindung Schärfe. Measurement technique for the simulation of the auditory sensation of sharpness*. Berlin: Beuth.

Fastl, H. (1997). The Psychoacoustics of Sound-Quality Evaluation. *ACUSTICA – acta acoustica, 83*, 754-764.

Fastl, H. (2005). Psycho-Acoustic and Sound Quality. In J. Blauert (Ed.), *Communication Acoustics* (pp. 139-162). Berlin Heidelberg: Springer Verlag.

Fastl, H., & Zwicker, E. (2007). *Psychoacoustics. Facts and Models*. 3rd edn., Heidelberg: Springer.

Genuit, K. (2002). Geräuschwahrnehmung. Parameter bei der Beurteilung von Fahreindrücken: Hören – Fühlen – Sehen – Wissen. In K. Becker (Ed.), *Subjektive Fahreindrücke sichtbar machen II. Korrelation zwischen objektiver Messung und subjektiver Beurteilung von Versuchsfahrzeugen und –komponenten* (pp. 143-162). Haus der Technik Fachbuch, Band 12, Renningen, Germany: expert verlag.

Genuit, K. (2008). Product Sound Quality of Vehicle Noise – A Permanent Challenge for NVH Measurement Technologies. *SAE International 2008-36-0517*, 1-17.

Genuit, K., & Burkhard, M. (1995). Subjective Measurement of Noise and Vibration Using Objective Techniques. *Sound and Vibration, May*, 28-34.

Genuit, K., Schulte-Fortkamp, B., & Fiebig, A. (2006). Neue Verfahren zum Benchmarking von Fahrzeuginnengeräuschen. In K. Becker (Ed.), *Subjektive Fahreindrücke sichtbar machen III. Korrelation zwischen objektiver Messung und subjektiver Beurteilung von Versuchsfahrzeugen und Komponenten* (pp. 127-145). Haus der Technik Fachbuch, Band 56, Essen: expert verlag.

Gomez, P., & Danuser, B. (2004). Affective and physiological responses to environmental noises and music. *International Journal of Psychophysiology, 53*, 91-103.

Griefahn, B., & di Nisi, J. (1992). Mood and cardiovascular functions during noise, related to sensitivity, type of noise and sound pressure level. *Journal of Sound and Vibration, 155*, 111-123.

Hashimoto, T. (2000). Sound quality approach on vehicle interior and exterior noise – Quantification of frequency related attributes and impulsiveness. *Journal of the Acoustical Society of Japan (E), 21*(6), 337-340.

Ising, H., & Kruppa, B. (2001). Zum gegenwärtigen Erkenntnisstand der Lärmwirkungsforschung: Notwendigkeit eines Paradigmenwechsels. *Umweltmedizin in Forschung und Praxis, 6*(4), 181-189.

Ising, H., Rebentisch, E., Babisch, W., Curio, I., Sharp, D., & Baumgärtner, H. (1990). Medically relevant effects of noise from military low-attitude flights – results of an interdisciplinary pilot study. *Environmental International, 16*, 411-423.

Iwanaga, M., & Moroki, Y. (1999). Subjective and Physiological Responses to Music Stimuli Controlled Over Activity and Preference. *Journal of Music Therapy, 36*(1), 26-38.

Kilian, K. (2007). Von der Markenidentität zum Markenklang als Markenelement. InK. Bronner, and R. Hirt (Eds.), *Audio-Branding. Entwicklung, Anwendung, Wirkung akustischer Identitäten in Werbung, Medien und Gesellschaft* (pp. 54-69). Praxisforum Medienmanagement, Band 5, 2. Nachdruck, München: Verlag Reinhard Fischer.

Krugmann, D., & Langeslag, P. (2007). Akustische Markenführung im Rahmen eines identitätsbasierten Markenmanagements. In K. Bronner, and R. Hirt (Eds.), *Audio-Branding. Entwicklung, Anwendung, Wirkung akustischer Identitäten in Werbung, Medien und Gesellschaft* (pp. 70-79). Praxisforum Medienmanagement, Band 5, 2. Nachdruck, München: Verlag Reinhard Fischer.

Mandryk, R. L., & Atkins, M. S.(2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies, 65*, 329-347.

Maschke, C., & Jakob, A. (2010). Psychoakustische Messtechnik. In M. Möser (Ed.), *Messtechnik der Akustik* (pp. 599-642). Heidelberg: Springer.

Miśkiewicz, A., & Letoweski, T. (1999). Psychoacoustics in the Automotive Industry. *ACUSTICA. acta acustica, 85*, 646-649.

Mühlstedt, J., Unger, H., & Spanner-Ulmer, B.(2007). Akustische Informations- und Warnsignale: Analyse, Gestaltungsmethodik und Evaluierung. In M. Rötting, G. Wozny, A. Klostermann, and J. Huss (Eds.), *Fortschritt Berichte VDI. Prospektive Gestaltung von Mensch-Technick-Interaktion* (pp. 203–208). Reihe 22, Nr. 25, Düsseldorf: VDI Verlag.

Nor, M.J.M., Fouladi, M.H., Nahvi, H., & Ariffin, A.K. (2008). Index for vehicle acoustical comfort inside a passenger car. *Applied Acoustics, 69*, 343-353.

Otto. N.C., & Wakefield, G.H. (1993). A Subjective Evaluation and Analysis of Automotive Starter Sounds. *Noise Control Engineering Journal, 41*, 377-382.

Raggam, R.B., Cik, M., Höldrich, R.R., Fallast, K., Gallasch, E., Fend, M., Lackner, A., & Marth, E. (2007). Personal noise ranking of road traffic: Subjective estimation versus physiological parameters under laboratory conditions. *International Journal of Hygiene and Environmental Health, 210*, 97-105.

Resch, S., & Mast, P. (2006). Engineered Emotion. In K. Becker (Ed.), *Subjektive Fahreindrücke sichtbar machen III. Korrelation zwischen objektiver Messung und subjektiver Beurteilung von Versuchsfahrzeugen und Komponenten* (pp. 118-126). Haus der Technik Fachbuch Band 56, Essen: expert verlag.

Sammler, D., Grigutsch, M., Fritz, T., & Koelsch, S. (2007). Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music. *Psychophysiology, 44*, 293-304.

Schifferstein, H.N.J. (2006). The perceived importance of sensory modalities in product usage: A study of self-reports. *Acta Psychologica, 121*, 41-64.

Schulte-Fortkamp, B. (2010). Bewertung von Fahrzeuggeräuschen: Ökologische Validität und subjektive Evaluation von Geräuschen zur Bestimmung der Qualität von Fahrzeuggeräuschen. In K. Genuit (Ed.), *Sound-Engineering im Automobilbereich. Methoden zur Messung und Auswertung von Geräuschen und Schwingungen* (pp. 121–132). Heidelberg: Springer-Verlag.

Schulte-Fortkamp, B., Genuit, K., & Fiebig, A. (2007). A New Approach for Developing Vehicle Target Sounds. *Sound and Vibration, Oct*, 12-15.

Wagner, V., Föhl, U., & Kallus, K.W. (2009). Vehicle Sound quality and customer satisfaction. Geräuschqualität im Fahrzeug und Kundenzufriedenheit. In *Proceedings zum Aachener Acoustics Colloquium* (pp. 25-34). Aachen.

Whang, M. (2008). The Emotional Computer Adaptive to Human Emotion. In J.H.D.M. Westerink, M.Ouwerkerk, T.J.M. Overbeek, W.F. Pasveer, and B. de Ruyter (Eds.), *Probing Experience.From Assessment of User Emotions and Behaviour to Development of Products* (pp. 209-219). Dordrecht: Springer.

Zeller, P. (2009). *Handbuch Fahrzeugakustik. Grundlagen, Auslegung, Berechnunung, Versuch*. ATZ/MTZ-Fachbuch, Wiesbaden: Vieweg + Teubner.

Keep calm and don’t move a muscle: Motor restlessness as an indicator of mental workload

Francesco Di Nocera[1], Serena Proietti Colonna[1], Claudio Capobianco[2], Federica Dessì[2], Simon Mastrangelo[2], Axel Steinhage[3]
[1]Sapienza U. of Rome, Italy; [2]Ergoproject S.r.l., Italy; [3]Future-Shape GmbH, Germany

Abstract

Subtle motor activity of the gluteal muscles and legs of the operator sitting in a chair can be obtained by using capacity sensors applied to the seat itself, thus making viable to easily record variations in motor restlessness during the execution of a task. The present study was aimed at studying the relation between variations in mental workload and variations in motor activity that are unrelated to the task to perform. Preliminary results from a laboratory study employing behavioural, ocular, subjective and motor measures show that motor restlessness can be considered as a promising indicator that can be usefully integrated in any data collection procedure for assessing the operator functional state.

Introduction

A leitmotiv of Human Factors research is that multiple approaches are needed for measuring mental workload, being the construct itself determined by many different factors related to both the nature of the task and the characteristics of the operator. Another reason for using multiple measures is the intrinsic imperfection of any of them in terms of sensitivity, diagnosticity, intrusiveness.

A large number of sensitive subjective, behavioural, and psychophysiological measures are reported in the literature. Many of them are still commonly used in research and practice, many others have failed to meet the requisites of diagnosticity or have temporarily failed because of the limits imposed by technology (Di Nocera, 2003). The use of eye movements, for example, has been revitalized after eye-tracking systems have become cheaper, less intrusive, and actually implementable in real world domains. Single trial ERPs such as the P300, instead, are -in our opinion- still unusable. The limitations due to the difficulties in extracting a clean signal from the on-going EEG avoiding artifacts, as well as the unreliable predictive power of a change in amplitude or latency, connote ERPs as a measure hard to export outside the laboratory setting.

The quest for a fairly good measure of mental workload will probably have no end, considering that 1) the real objective is that of finding a measure that can work in real operating environments and –most importantly- in the real-time, 2) the

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

pervasiveness and miniaturization of portable technology is tremendously growing and already delivers novel and unpredictable solutions 3) different measures are needed to suit different operating settings, also because they show sensitiveness within different bandwidths (de Waard, 1996). The latter consideration is not trivial. For example, infrared based solutions for eye-tracking may not be the best approach when the operational settings are characterized by abrupt and frequent changes in lighting. Moreover, sophisticated techniques for measuring mental load may simply be unnecessary. Consider, for example, the difference between the need to measure mental workload of baggage screening vs. customer care operators. In both cases, information on the operator functional state is useful for mitigating the workload through alarming systems, automation support and the like. However, the activities carried out in the first case need a better time resolution than the second. Indeed, sudden changes in mental workload can strongly affect performance in the first case, whereas in the second case exact timing is not an issue. This difference applies to many settings in which the mental load assessment does not serve a timely change in the operative procedures. In those cases, cheap, smooth, not at all intrusive, and even rough measurement strategies may be better suited.

In this paper we will take into consideration the *incidental* motor activity that is unrelated to the primary task imposing the load. We started from the consideration that many operative settings require individuals to spend much time seated over a chair. Over a prolonged period of time, sitting becomes uncomfortable and we believe that the spontaneous motor activity of the operator constrained in that position can be considered as a candidate measure of his/her functional state.

Motor restlessness

Motor activity has been usually neglected in mental workload assessment with the exception of the NASA-TLX (Hart & Staveland, 1988), as the physical demand was included among the factors affecting the overall mental workload. In that case, however, the focus was on the amount and intensity of physical activity required to complete the task. What is of our interest here is, instead, is the physical activity that is not related to the primary task.

As far as we know, the only experimental study in the HF/E domain specifically involving motor restlessness as a variable is that carried out by Galinsky and colleagues (1993). In that case, restlessness was employed as and indicator of stress in a vigilance task and results showed that it increased during the task along with fatigue. After that contribution, the idea of including restlessness within the set of measures usually recorded during workload/vigilance studies has not been taken into consideration anymore. Still, movement is a pervasive human activity, not only when it relates to the operating task: the incidental motor activity may be interesting as well.

Although the type of movements we are interested in here are not pathological in nature, the urge to move one's body to stop uncomfortable or odd sensations is typical of a neurological disorder known as Restless Legs Syndrome (RLS). Sir Thomas Willis had noted the symptoms as far back as 1672 (Molnár, 2004): "Wherefore to some, when being abed they betake themselves to sleep, presently in the arms and legs, leapings and contractions on the tendons, and so great a restlessness and tossings of other members ensue, that the diseased are no more able to sleep, than if they were in a place of the greatest torture" (from "De Anima Brutorum"). The RLS is a controversial disorder, as many physicians consider its incidence exaggerated by manufacturers of drugs used to treat it (Woloshin & Schwartz, 2006). However, whether a diagnosis of RLS is genuine or not, the pervasive habit to engage in incidental motor activity while performing another task by many people it is an undeniable fact.

In this paper we will thus take into consideration the *incidental* motor activity that is unrelated to the experimental task employed for imposing the taskload. Such subtle motor activity may be related to the unpleasant feeling of being uncomfortable, or -more generically- to the impatience of the individual. For example, tapping with one's foot impatiently it is a common experience in relation to nervousness, agitation, boredom or a combination of these. Such motor restlessness is among the activities commonly grouped under the term "fidgeting" (although, fidgeting includes many other behaviours, such as playing with one's hair or clothes, that are not of our interest here). Fidgeting is commonly considered a bad habit and also symptom of mental tension. Nevertheless, fidgeting has also been related to better concentration (Pine et al., 2013).

Study

Participants

Twelve individuals (7 females, mean age = 29.6 st. dev. = 5.82) were involved in this study. All participants had normal or corrected to normal vision.

Experimental setting

A custom version of the Tetris game, a commonly known tile-matching puzzle videogame, was coded using Matlab. The version used in this study acted as a common version of the Tetris game with the exception that in this case the game started from a blank screen (starting condition) each time the stack of Tetriminos (game pieces) reached the top of the playing field and no new Tetriminos were able to enter. This condition commonly denotes the end of the game, whereas in this very situation was scored as a loss. Number of losses was used as dependent variable for assessing participants' performance. Three difficulty conditions were implemented by manipulating the falling speed of the Tetriminos (Easy, Intermediate, Hard: 125, 100, and 80 ms per row respectively). Each condition lasted 10 minutes and the condition order was randomized across participants.

Ocular data collection

The X2-30 wide eye tracking system (Tobii, Sweden) was used for recording ocular activity. The X2-30 Eye Tracker is a standalone eye tracker that can be used in various setup by attaching it to monitors, laptops or for performing eye-tracking on physical objects with a sampling rate of 30 Hz. Its large head movement box allows the subjects to move during recording while maintaining accuracy and precision.

The X2-30 uses infrared diodes to generate reflection patterns on the corneas of the subject's eyes. These reflection patterns, together with other visual data about the subject, are collected by image sensors. Image processing algorithms identify relevant features, including the eyes and the corneal reflection patterns. Complex mathematics is used to calculate the 3D position of each eyeball, and finally the gaze point on the screen.

Fixations data collected in the three gaming conditions were used as input for the software package ASTEF (Camilli et al., 2008) in order to compute the Nearest Neighbor Index (NNI), an index of spatial dispersion that has been repeatedly found to be correlated to mental workload (Camilli et al., 2007; 2008; Di Nocera et al. 2006; 2007).

Motor data collection

In order to measure the subjects' movements on the seat, the PressureTex system (Future-Shape GmbH, Germany) was used. It consists of two layers of conductive textile with a non-conductive compressible layer in-between implementing a textile capacitor. The lower conductive layer is structured in typically 32 independent sensor fields per square meter the capacitance of which is measured by means of integrated flexible electronics modules. When force is applied, the textile is compressed leading to a rise in capacitance at that location. Figure 1 shows the typical dependency of capacitance on the pressure applied.

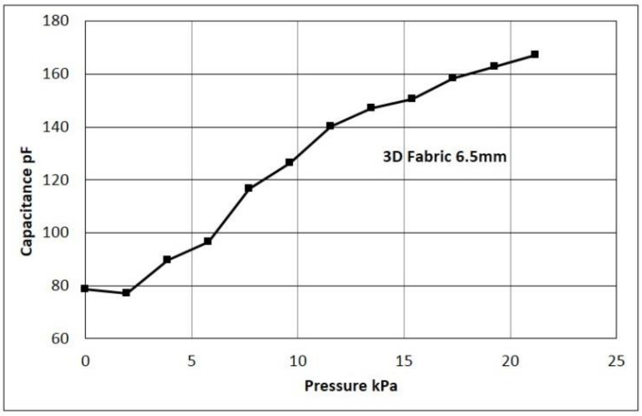


Figure 1. PressureTex typical dependency of capacitance on the pressure applied.

The capacitance data is transmitted wirelessly from the electronics modules to a radio receiver plugged into a PC's USB port. Whereas the spatial and temporal resolution of the sensors are typically 1/32 m² and 1/100 s, the data for the experiments described in this paper were obtained by low-pass filtering the maximum pressure variation over the whole seat and over 1 minute.

Procedure

Participants were seated in front of a 17” display, at a distance of approximately 60 cm, over a chair mounting the capacitance sensors described above. The room was dimly illuminated only by the display. After calibration of the eye-tracking system, they underwent a practice run of the Tetris game. Practice was run at the Intermediate level of difficulty for all subjects, who were informed on the nature of the level of difficulty for avoiding context effects in the subjective assessment (see Colle & Reid, 1998) they carried out by compiling the NASA-TLX at the end of each gaming session.

Data analysis and results

Performance, subjective, ocular, and motor data were used as dependent variables in repeated measures ANOVA designs using Taskload as repeated factor (Easy vs. Intermediate vs. Hard). One subject was excluded from the analysis on motor data because he/she did not produce any recordable change over the sensors during an entire session. Results showed a significant effect of Taskload on the number of losses ($F_{2,20}$=159.16, $p<.001$). As expected, performance deteriorates as difficulty increases. Duncan post-hoc testing showed that all differences were statistically significant.

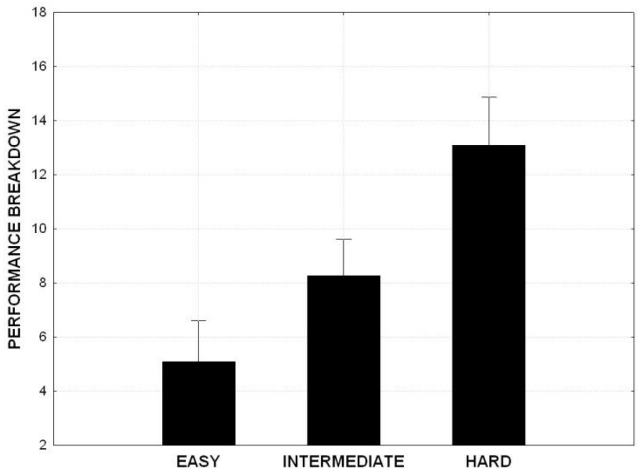


Figure 2. Performance breakdown in the three taskload conditions. Error bars denote 95% confidence intervals.

Results showed a significant effect of Taskload on the NASA-TLX weighted scores ($F_{2,20}$=5.60, $p<.05$). Perceived workload was lowest in the Easy condition. Particularly, Duncan post-hoc testing showed that only the difference between the Easy condition and the other two was statistically significant.

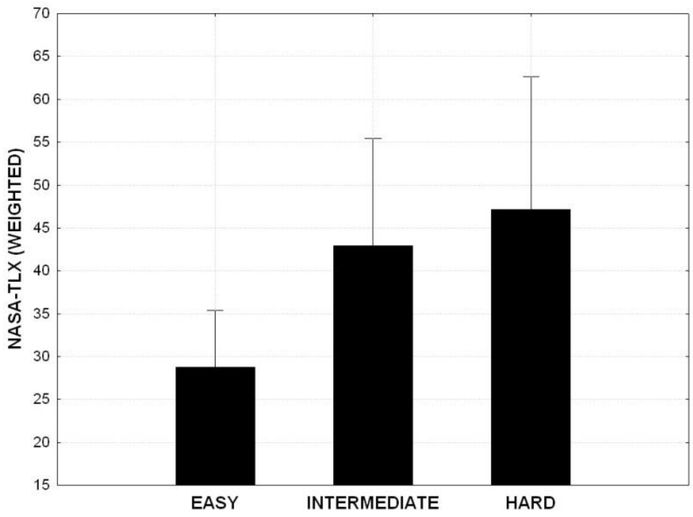


Figure 3. NASA-TLX weighted scores in the three taskload conditions. Error bars denote 95% confidence intervals.

Results also showed a significant effect of Taskload on the distribution of eye fixations ($F_{2,20}$=5.87, p=.01). The fixation pattern became more widespread as taskload increased. Particularly, Duncan post-hoc testing showed that the difference between the Easy and the Intermediate condition was significant, whereas the difference between the Intermediate and the Hard condition showed a tendency towards statistical significance (p=.07).

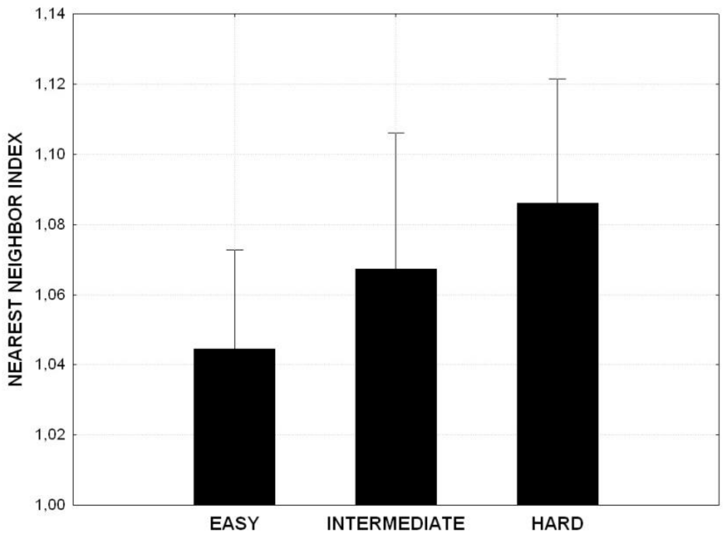


Figure 4. Nearest Neighbour Index in the three taskload conditions. Error bars denote 95% confidence intervals.

A significant effect of Taskload was also found on the average excursion of movements over the chair ($F_{2,18}$=12.07 p=.001). Greater restlessness effects were associated with the Easy condition. Particularly, Duncan post-hoc testing showed that only the difference between the Easy condition and the other two was statistically significant. However, the difference between the Intermediate and the Hard condition approached significance (p= .09).

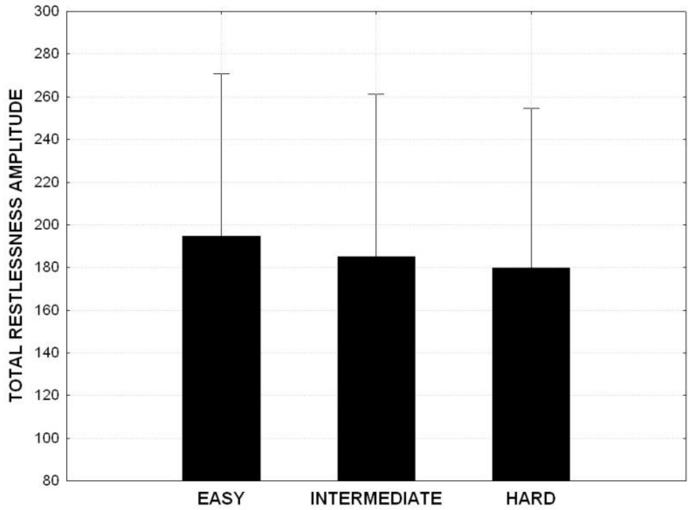


Figure 5. Restlessness amplitude (pF) in the three taskload conditions. Error bars denote 95% confidence intervals. Subtle motor activity of the gluteal muscles and legs of the operator sitting in a chair can

Discussion and Conclusions

When we started devising this study, we were exploring the opportunity to investigate the relation between subtle motor activity and mental load. We candidly admit there was no theoretical linkage between the two at that time, except the trivial consideration that expressions of human variability are somewhat linked one to another, and that it was worth investigating another parameter possibly linked to the operator functional state. The effect found (namely, the lower is the mental workload the greater is the restlessness) it is suggestive of a relation between these two constructs that is worth to be further investigated in successive studies.

People apparently make more incidental motor activity when taskload is low. This excludes that the type of motor activity we have measured here is related to apprehension for one's performance in the primary task. On the contrary, we could expect boredom in this case. It is worth noting that our participants underwent a training session at the intermediate level of difficulty, thus the first time they approached the easiest level of difficulty they may have experienced the tediousness of the slow-falling Tetriminos. However, interacting with a slow-paced task may reflect in drifting attention to a faster-paced activity (e.g. food-tapping). This concept is not new in the functional analysis of behaviour and Susan Schneider has convincingly described it in her book “The Science of Consequences”. According to Schneider (2012), not only variety is a reinforcer (“even three-month-old babies will turn their heads so that they can see complex patterns but not simple ones”), also activity itself is intrinsically reinforcing. A slow-paced task may be a poorly rewarding activity from which people are willing to escape. With that in mind, “boredom” would indicate a lack of reinforcers, not a lack of stimulation.

What we have so far described as “restlessness” may indeed be an active (albeit mindless) strategy for focusing on the primary task. Fidgeting might indeed provide a form of a “mental break”. Mehrabian and Friedman (1986), for example, suggested that fidgeting is an activity overflow and that it is more probable when the organism’s physical activity is constrained by the central or focal act. As a matter of fact, in the context of a sustained attention task, task switches can decrease the vigilance decrement effect (Ariga & Lleras, 2011).

Of course these are mere speculations, and a specific research programme should address these very hypotheses. Nevertheless, this study provided additional evidence for the distribution of eye fixations as an indicator of mental load and the results on the relation between taskload and incidental motor activity are very encouraging. Finding a linkage between a spontaneous activity and the operator functional state is always appealing in the HF/E domain. Also, as reported above, in many operational settings the mental load assessment does not serve a timely change in the operative procedures. Motor restlessness (or fidgeting) could be the cheap, smooth, non-intrusive metric to employ in those cases. Future studies will address the many issues that could not be solved in one single experiment (e.g. the large variability affecting the movement data) and, as a first step, we are considering the opportunity to address the sensitivity issue by enhancing the capacitance sensors needed for gathering the subtle activity we are looking for.

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

Mehrabian, A., Friedman, S.L. (1986). An analysis of fidgeting and associated individual differences. *Journal of Personality, 54*, 406–429.

Camilli, M., Nacchia, R., Terenzi, M., & Di Nocera, F. (2008). ASTEF: A Simple Tool for Examining Fixations. *Behavior Research Methods, 40*, 373-382.

Camilli, M., Terenzi, M., & Di Nocera, F. (2007). Concurrent validity of an ocular measure of mental workload. In D. de Waard, B. Hockey, P. Nickel, and K. Brookhuis (Eds.), *Human Factors Issues in Complex System Performance* (pp. 117-129). Maastricht, The Netherlands: Shaker Publishing.

Camilli, M., Terenzi, M., & Di Nocera, F. (2008). Effects of Temporal and Spatial Demands on the Distribution of Eye Fixations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52, 1248-1251.

Colle, H.A., & Reid, G.B. (1998). Context effects in subjective mental workload ratings. *Human Factors*, *40*, 591-600.

De Waard, D. (1996). *The measurement of drivers' mental workload* (Doctoral Dissertation). Available at the University of Groningen Dissertation database (http://dissertations.ub.rug.nl).

Di Nocera, F. (2003). On Reliability and Stability of Psychophysiological Indicators for Assessing Operator Functional States. In G.R.J. Hockey, A.W.K. Gaillard, & O. Burov (Eds.), *Operator Functional State: The Assessment and Prediction of Human Performance Degradation in Complex Tasks* (pp. 162-173). Amsterdam: IOS Press.

Di Nocera, F., Camilli, M., & Terenzi, M. (2007). A random glance at the flight deck: pilot's scanning strategies and the real-time assessment of mental workload. *Journal of Cognitive Engineering and Decision Making, 1*, 271-285.

Di Nocera, F., Terenzi, M., & Camilli, M. (2006). Another look at scanpath: distance to nearest neighbour as a measure of mental workload. In D. de Waard, K.A. Brookhuis, and A. Toffetti (Eds.), *Developments in Human Factors in Transportation, Design, and Evaluation* (pp. 295-303). Maastricht, the Netherlands: Shaker Publishing.

Galinsky, T.L., Rosa, R.R., Warm, J.S., & Dember, W.N. (1993). Psychophysical determinants of stress in sustained attention. *Human Factors, 35* , 603-614.

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.

Molnár, Z. (2004). Timeline: Thomas Willis (1621–1675), the founder of clinical neuroscience. *Nature Reviews Neuroscience, 5*, 329–335.

Pine, K.J., Reeves, L., Howlett, N., Fletcher, B.C. (2013). Giving cognition a helping hand: the effect of congruent gestures on object name retrieval. *British Journal of Psychology, 104,* 57-68.

Schneider M.S., (2012). *The Science of Consequence: How They Affect Genes, Change the Brain, and Impact Our World*. Amherst (NY): Prometheus Books.

Woloshin, S. & Schwartz, L.M. (2006). Giving Legs to Restless Legs: A Case Study of How the Media Helps Make People Sick. *PLoS Medicine, 3*(4), e170.

Timing of in-vehicle advisory warnings based on cooperative perception

Frederik Naujoks & Alexandra Neukum
Centre for Traffic Sciences (IZVW), University of Wuerzburg
Germany

Abstract

Future cooperative perception technologies provide the possibility of assisting drivers by so-called advisory warnings in potentially dangerous driving situations. The effectiveness of such advisory warnings will possibly depend on (1) warning timing and (2) situation-specific expectations of the driver. Using a fixed-base driving simulator, n=20 participants encountered three different conflict situations, which varied in the possibility to anticipate critical situations: (1) turning vehicle taking the driver´s right of way (anticipation level: low), (2) pedestrian entering the road between parked cars (anticipation level: medium) and (3) cyclist passing a signalized intersection (anticipation level: high). The situations were completed with and without assistance by a prototypical advisory warning system. Advisory warnings were provided via head-up-display and accompanied by an unobtrusive acoustic signal. Warning timing was varied in five steps (last possible warning moment up to four seconds prior to the last possible warning moment in 1s-steps). Advisory warnings strongly reduced objective and subjective situation criticality if they were provided at least one second before the last possible warning moment, whereas drivers subjectively preferred advisory warnings between two and three seconds before. The advisory warnings were of greatest effect when conflict situations were hard to anticipate.

Introduction

The Ko-PER research project aims at achieving improvement in preventive traffic safety. Ko-PER is a German national research project, which is founded by the Federal Ministry of Economics and Technology. Its objective is to use distributed sensory networks in order to provide road users with a complete picture of the local traffic environment. Here, wireless vehicle-vehicle and vehicle-infrastructure communication is used to provide information about the current traffic situation and to consolidate it with vehicle-localized environmental perception (so-called cooperative perception). Thus, in comparison to sensor systems that are only vehicle-localized, conflict situations can be detected more completely (e.g., by eliminating sight obstructions) and much earlier, which enables on-time driver support. Functional and HMI development run parallel during the Ko-PER research project. The objective of the HMI design is to determine suitable information

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

strategies in order to provide early driver assistance in impending and time-critical conflict situations.

This paper presents the results of a simulator study during which advisory warnings were provided before critical traffic situations in order to mitigate conflict. Based on research results pertaining to imminent crash warnings, it is expected that the effectiveness of such advisory warnings on one hand depends on the time of transmission of such data to the driver, and on the other hand is contingent upon the situation-specific expectation by the driver (Schmidt & Krüger, 2011); that is to say, the driver having the option to react in anticipation of conflict situations.

Background

Motivation

The basis for the output of an information or a warning signal is so-called ‘situation analysis’ (Weidl & Breuel, 2012). Here, based on the consolidated data of the cooperative perception, current traffic situations are captured holistically and a probabilistic situation prognosis, including a forecasting situation criticality, are generated. The assessment of the prognosticating situation criticality (e.g., via a time-based measure such as Time-to-collision) is utilized in order to prioritize the type of driver assistance (e.g., advisory warning at (still) low-level situation criticality or imminent crash warning at high prognosticated situation criticality). The challenge when designing early advisory warnings is the decision regarding the transfer of potentially unreliable information to the driver. Although, from a technical point of view, it is possible to collect these data, based on its probabilistic nature, it might be unreliable, and therefore may have a negative effect on the efficiency and acceptance of the driver support (Bliss & Acton, 2003; Sorkin, 1988). Based on findings regarding urgent warnings (Brown, Lee & McGehee, 2001; Lee, McGehee, Brown & Reyes, 2002; McGehee, Brown, Lee & Wilson, 2002) the *time frame for advisory warnings* must be localized. In order to ensure maximum reliability and, at the same time, ascertain a minimum degree of distraction, advisory warnings should be provided as late as possible, but as early as needed in order to initiate an adequate driver’s response. However, currently, information about this time frame, which represents an important foundation for the design of the HMI, is only available in isolated cases (e.g., Lenné & Triggs, 2009; Totzke, Naujoks, Mühlbacher & Krüger, 2012) but was prepared within the scope of the study at hand.

Conceptual framework

Established concepts to provide driver assistance in impending conflict situations are largely based on vehicle-localized sensor systems and are considered to render urgent driver warning signals in order to prevent collisions or, if necessary, autonomously intervene with the operation of the vehicle (e.g., Brown et al., 2001; Rhede, Wäller & Oel, 2011; Winner, 2009). Pertinent review articles consider intervals of 700 to 1,500 ms prior to an imminent accident as a suitable time frame for such urgent warnings (referred to as ‘imminent crash warnings’, e.g., Lenné & Triggs, 2009; Spence & Ho, 2008). When considering cooperative perception, the time frame prior to the imminent crash warnings (> 1,500 ms), makes the presentation of driver information about impending conflict situations accessible (Neukum, 2011). These messages referred to as ‘advisory warnings’ (Lenné & Triggs, 2009) or ‘risk information’ (Rhede et al., 2011) should alert the driver to a potential conflict and render him/her ready to respond. On the contrary, imminent crash warnings aim at the immediate reaction of the driver in order to avoid a collision. During the Ko-PER conception phase of the driver assistance system, it was assumed that the established chronological warning structures, that have been described or investigated many times, should be maintained in order to ensure a consistent message transfer to the driver. Therefore, the chronological structure of the warning concepts that are based on vehicle-localized sensor systems will be supplemented by adding an early information component (i.e., advisory warnings) by means of a cooperative sensor system without changing previous chronological structures of established warning concepts (i.e., imminent crash warnings, see Figure 1).

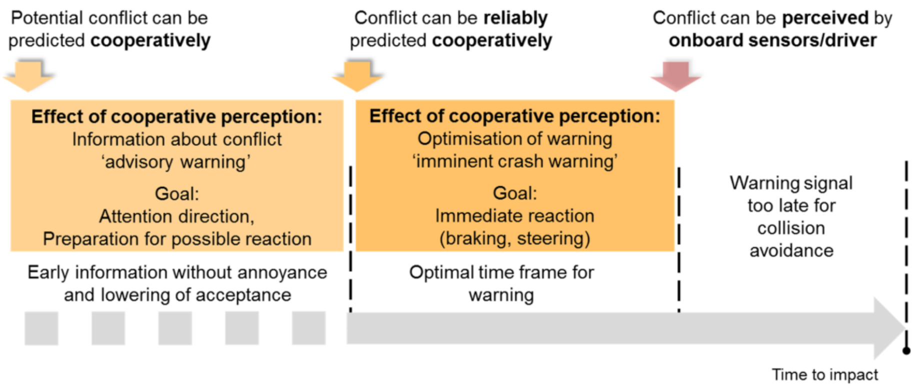


Figure 1. Schematic of a time frame for driver assistance during the last seconds prior to a collision (Neukum, 2011).

Configuration of advisory warnings for collision avoidance

In addition to the *point in time* when to provide effective advisory warnings, which is the focus of this study, a decision must be made concerning the *modality* and *content* of the information provided. In part, this may be based on available research papers. From the literature pertaining to the configuration of advisory warnings, the following recommendation for the *modality* design can be derived (e.g. Campbell, Carney & Kantowitz, 1997; COMSIS Corporation, 1996; Dingus, Jahns, Horowitz & Knipling, 1998; Green, Levison, Paelke & Serafin, 1993; International Harmonized Research Activities (IHRA) working group on Intelligent Transport Systems (ITS), 2008; Rhede et al., 2011):

- Application of visual displays rather than voice messages or exigent sounds.
- Acoustic signals (warning sounds or audio response) are to be avoided and should be reserved only for urgent warnings. This applies in particular to frequent activation that may cause false alarms (which is to be expected for

early advisory warnings). These urgent auditory warnings may have an adverse effect in this case.

- An announcement that uses a non-intrusive tone can increase the effectiveness of the advisory warning.

Furthermore, additional requirements for the *content of the advisory warning* can be derived from the above-mentioned configuration recommendations. It is recommended to further display the type of the respective conflict (referred to as *conflict specificity*) and the location at which the conflict is imminent (referred to as *directional specificity*). It is further conceivable to provide information about the distance remaining to the imminent conflict (referred to as *location specificity*) or providing data about the predicted probability of the conflict (referred to as *risk specificity*). Whether such specific data stimulate a favorable effect on driving behaviour in imminent conflict situations cannot be conclusively explained based on the available literature. There are studies that provide information about a faster driver response when provided with directional-specific signals (Ho & Spence, 2005; Spence & Ho, 2008; 2009) as well as studies that do not provide evidence for the benefits of such signals (Lee, Gore & Campbell, 1999; Bliss & Acton, 2003; Cummings, Kilgore, Wang & Kochhar, 2007). With regard to the effectiveness of conflict (Cummings, Kilgore, Wang & Kochhar, 2007; Thoma et al., 2009), local (Totzke et al., 2012), and risk-specific (Gupta, Bisantz & Singh, 2002; Lee, Hoffman & Hayes, 2004) warnings, there are very few studies with mixed results. Therefore, positive statements about the benefits of such warnings cannot be supported conclusively.

Method

Experimental design

The sample consisted of N = 20 participants, half of them male and half of them female of various age groups (20-39 years of age: N=10, 50-71 years of age: N=10). At the time of the study, the driving experience of the participants varied between two and 53 years (M = 28.10 years, SD = 17.29 years). The participants were selected from an existing WIVW (Wuerzburg Institute for Traffic Sciences) test driver panel. All participants had previously participated in extensive simulator training.

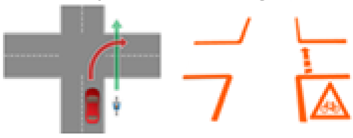
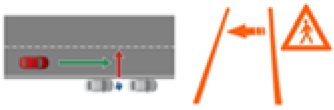
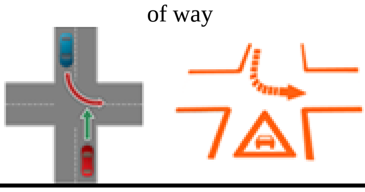
The study was conducted in the WIVW static driving simulator using a within-subject design with the experimental factors (1) warning *timing* and (2) situation-specific *expectations* of the driver. Supplied with or without driver assistance, the participants manoeuvred through a simulator course that included different conflict situations. Advisory warnings about upcoming conflicts were provided up to four seconds prior to the last possible warning moment and in five increments. The second focus of this study was a selection of three conflict situations (see Table 1). Here, the degree of predictability of conflicts varied. In an *anticipated situation*, a cyclist crosses the road while the vehicle attempts a right-hand turn (high anticipation). In an *unexpected scenario*, a pedestrian appears between two parked vehicles, enters the road and crosses it (medium anticipation). In a *surprising situation*, a turning vehicle ignores a stop sign, disregarding the right of way of the participant (low anticipation). In order to control sequence effects, six different random scenario orders were created and assigned to the drivers at random. Furthermore, the simulator course included several non-critical scenarios in which no driver intervention was needed.

Human machine interface

The drivers received visual-auditory advisory warnings, which contained the respective display of the opposing road user and the direction of the conflict. In combination with an unobtrusive acoustic signal (500 Hz sinus), it was displayed in the simulated HUD (see Table 1). The point in time when the advisory warning was provided was varied in five stages, whereby one second was added to the last possible warning moment t_0 (see Table 1). Under consideration of the initial speed, it was assumed that after receiving the information, the driver activates the brakes after a reaction time (T_r) of one second and continues deceleration at 8 m/s² until the vehicle comes to a complete halt: $t_0 = T_r + (v/2a)$ whereby: $T_r = 1s$, $a = -8\ m/s^2$.

Triggering the advisory warning was a function of the driver's behaviour. The remaining time to reach the respective conflict point[1] was taken into consideration. If during this phase the respective threshold time (t_0, t_{0+1s} etc.) fell short, the advisory warning was triggered. If the driver initiated the deceleration when approaching the point of conflict, no advisory warning was provided.

Table 1. Scenarios, HUD displays, and time of warning presentation as a function of the respective driving speed.

Scenario	*v[km/h]*	*Timing of advisory warning*				
		t_0	t_{0+1s}	t_{0+2s}	t_{0+3s}	t_{0+4s}
Anticipated: Right turn – crossing cyclist at traffic light 	25	1.43	2.43	3.43	4.43	5.43
Unexpected: Pedestrian enters the road between parked cars 	50	1.87	2.87	3.87	4.87	5.87
Surprising: Turning vehicle takes right of way 	50	1.87	2.87	3.87	4.87	5.87

[1] Time to the point of conflict [s] = Distance to the point of conflict [m] / speed [m/s]

Dependent variables

In addition to the recording of objective data, e.g., approaching and deceleration behaviour of the driver during conflict situations, subjective assessments of the experienced criticality (Neukum et al., 2008; Figure 2 left) were raised. Furthermore, the participants rated the extent to which they experienced the situation as surprising, how timely the information was presented and how helpful the information was (Figure 2).

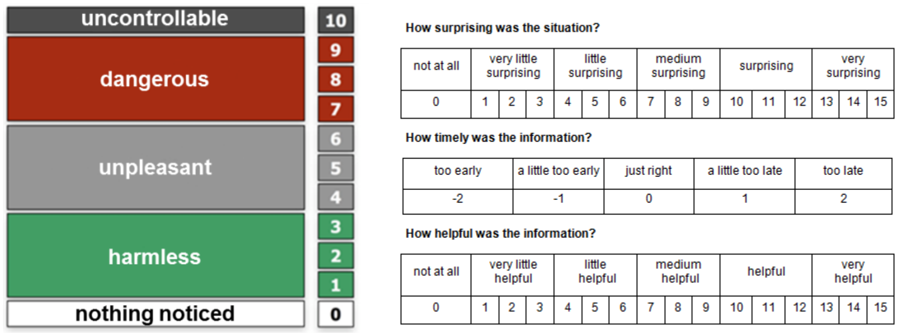


Figure 2. Subjective measures.

Statistical procedures

Given the experimental setup (within-subject design), global effects of the independent variables (scenario: low vs. medium vs. high anticipation of critical driving scenario, level/timing of driver assistance: no assistance vs. advisory warning at t_0 vs. … advisory warning at t_{0+4s}) on driving behaviour and subjective assessments were investigated using linear mixed models (fixed effects: scenario, timing and interaction scenario x timing; random effect: participants) with an alpha level of 5%. Differences in subjective and objective situation criticality to unassisted driving were of special interest in this study. Here, t-tests for dependent samples were applied in order to test for significant differences to baseline driving.

Results

As is evident in Table 2, most participants fell below the trigger threshold for advisory warnings pertaining to the surprising and unexpected situation. During the anticipated situation, the participants reduced their speed by themselves and, therefore, triggered a late advisory warning (t_0 und t_{0+1s}) less frequently.

Table 2. Frequency of triggering advisory warnings.

Timing/Scenario	t_0	t_{0+1s}	t_{0+2s}	t_{0+3s}	t_{0+4s}
Anticipated: Right turn – crossing cyclist at traffic light	3	7	16	19	20
Unexpected: Pedestrian enters the road between parked cars	12	20	20	20	20
Surprising: Turning vehicle takes right of way	20	19	20	20	20

Table 3. T -test for dependent sample statistics, difference to baseline in participants' assessment of surprise.

Timing/Scenario		t_0	t_{0+1s}	t_{0+2s}	t_{0+3s}	t_{0+4s}
Anticipated: Right turn – crossing	*t*		1.89	-0.25	0.57	0.44
cyclist at traffic light	*df*		6	15	18	19
	p		.108	.804	.574	.662
Unexpected: Pedestrian enters the	*t*	-0.65	1.56	3.08	3.99	3.43
road between parked cars	*df*	11	19	19	19	19
	p	.527	.135	.006	.001	.003
Surprising: Turning vehicle takes	*t*	-0.94	5.02	6.25	6.48	7.95
right of way	*df*	19	18	19	19	19
	p	.357	<.001	<.001	<.001	<.001

Consequently, the intended manipulation of the predictability of the situations was being fulfilled. This was also substantiated in the subjective assessment of level of surprise in the conflict situations (see table 3) during baseline drives without driver assistance. On average, the surprising situation was rated as 'surprising', the unexpected situation as 'medium surprising' and the anticipated situation as 'little surprising'. The impact of advisory warnings depended on the warning timing and the respective situation (scenario: $F = 14.13$, $p < .001$; timing: $F = 11.78$, $p < .001$; interaction: $F = 2.84$, $p = .009$). When considering the surprising and the unexpected situation, it becomes clear that, compared to non-assisted driving, the situations were consistently assessed as less surprising from t_{0+2s} (in the unexpected scenario) or t_{0+1s} (in the surprising scenario, see Table 3). On the contrary, the anticipated situation was little surprising without and with driver assistance.

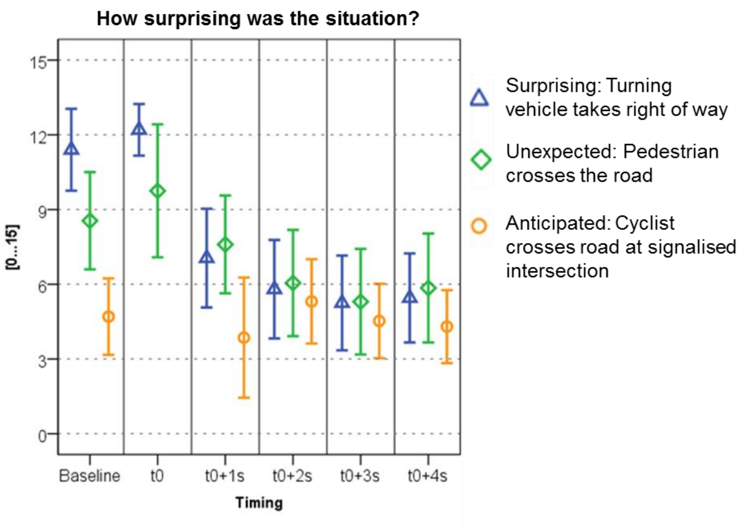


Figure 3. Reported surprise during the encountered situations. Illustrated are mean values[2]

If advisory warnings were provided, the participants' assessments regarding the situation criticality confirmed, as expected, the predictability of the respective conflict (see Figure 3). While the majority of participants rated the anticipated situation as 'harmless', the unexpected situation was mainly considered as 'unpleasant', and the surprising situation was assessed as 'dangerous'. This was in accordance with the minimum Time-to-arrival (TTA_{min}) to the respective conflicting road user (see Figure 3), whereby the surprising situation resulted in the shortest TTA_{min}, followed by the unexpected and anticipated situations.

Table 4. T-test for dependent samples statistics, difference to baseline in participants' assessment of criticality.

Timing/Scenario		t_0	t_{0+1s}	t_{0+2s}	t_{0+3s}	t_{0+4s}
Anticipated: Right turn – crossing cyclist at traffic light	*t*		2.05	-0.36	0.00	0.33
	df		6	15	18	19
	p		.086	.728	1.00	.748
Unexpected: Pedestrian enters the road between parked cars	*t*	-1.70	0.35	2.26	3.33	2.46
	df	11	19	19	19	19
	p	.118	.728	.036	.004	.024
Surprising: Turning vehicle takes right of way	*t*	-0.68	5.08	6.43	6.43	7.78
	df	19	18	19	19	19
	p	.504	<.001	<.001	<.001	<.001

Depending on the predictability of the conflict, the presentation of advisory warnings affected subjective and objective situation criticality (see Figure 3; subjective assessment: scenario: $F = 35.46$, $p < .001$; timing: $F = 16.17$, $p < .001$;

[2] *(Based on the limited number of only 3 drivers that actually received driver support (see Table 2), the results for the anticipated situation of time t_0 were not incorporated into the assessment.)*

interaction: F = 5.16, p <.001; TTA_{min}: scenario: F = 6.16, p = .003; timing: F = 46.19, p <.001; interaction: F = 10.45, p <.001). In the anticipated situation, there were neither subjective nor objective changes when using driver assistance. Whether advisory warnings were provided or not, only harmless situations or an average TTA_{min} of ≥ 2 seconds were the result. In the unexpected scenario, a decrease of subjective criticality occured starting with t_{0+2s}, and a decrease in objective criticality with t_{0+1s} (see Table 4 and 5).

Table 5. T -test for dependent samples statistics, difference to baseline in minimum Time-to-arrival.

Timing/Scenario		t_0	t_{0+1s}	t_{0+2s}	t_{0+3s}	t_{0+4s}
Anticipated: Right turn – crossing cyclist at traffic light	*t*		0.15	-0.60	2.18	1.68
	df		6	15	16	19
	p		.888	.556	.044	.110
Unexpected: Pedestrian enters the road between parked cars	*t*	1.82	-2.18	-4.38	-3.06	-4.04
	df	11	18	19	18	18
	p	.096	.043	.000	.007	.001
Surprising: Turning vehicle takes right of way	*t*	2.22	-1.99	-5.89	-4.77	-6.09
	df	18	10	13	17	15
	p	.039	.074	<.001	<.001	<.001

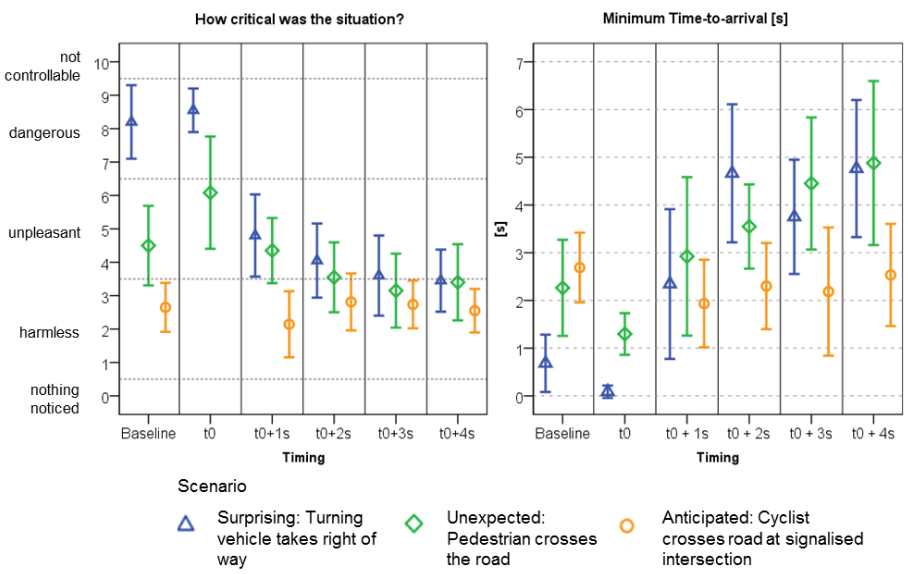


Figure 3. Reported situation criticality (left) and TTA_{min} to the conflicting road user (right); illustrated are mean values and 95% confidence intervals. If drivers decelerated below a threshold of 5km/h before the conflicting road user entered the conflict area, the resulting TTA_{min}-values were not reported because they produced large outliers in the data set.

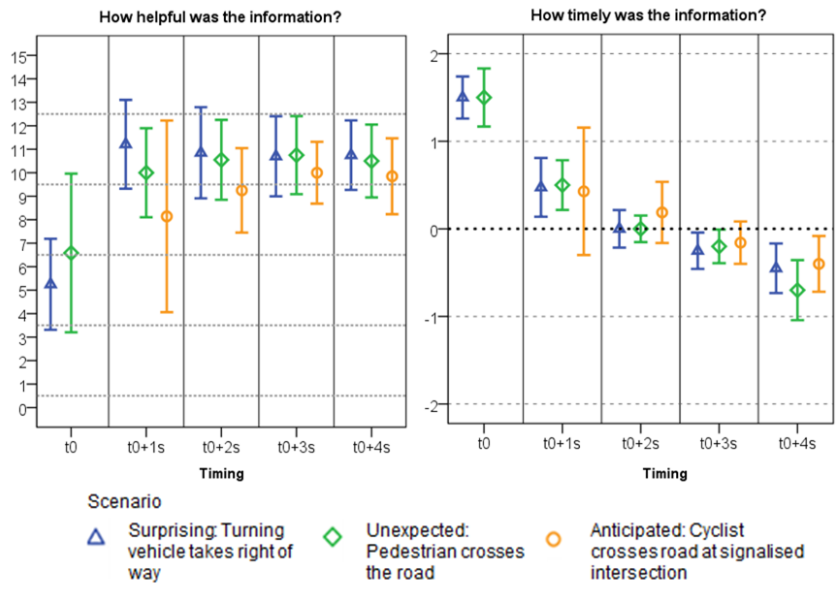


Figure 4. Reported usefulness (left) and timeliness (right) of the advisory warnings; illustrated are mean values and 95% confidence intervals[i].

Compared to non-assisted driving during surprising scenarios, the decrease in subjective criticality occured starting from t_{0+1s}, and with objective criticality at t_{0+2s} (see also Table 4 and 5). At this point, it must be stated that, when compared to non-assisted driving, drivers reported a higher criticality if the advisory warnings were provided at t_0. Similarly, objective data during these drives indicated lower TTA_{min}-values than those drives without assistance. However, this applied to the unexpected as well as the surprising situation.

If advisory warnings were provided only at t_0, those messages were evaluated on average as being of little help or falling within the medium range of being helpful (see Figure 4). If advisory warnings were provided earlier, the assessment of these warnings' usefulness also increased independently from the driving scenario (scenario: F = 2.51, p = .094 ; timing: F = 6.60, p <.001; interaction: F = 0.35, p = .929). This reflects the reported mitigation of situations if information about upcoming conflicts was received prior to the last possible warning moment; however, this did not hold true if data were received at the last possible moment. With respect to the timeliness with which the advisory warnings were provided, all situations indicate that advisory warnings presented two to three seconds prior to the last possible warning moment (t_{0+2s} or t_{0+3s}), received the best assessment (see Figure 4; scenario: F = 0.48, p = .623; timing: F = 69.95, p <.001; interaction: F = 0.31, p = .947).

Discussion

The *effectiveness* of early advisory warnings based on cooperative perception in different conflict situations was investigated. The objective of the study was to localize the time frame for such early driver assistance under consideration of the effects related to situational driver expectations. In particular, a positive effect of the driver information could be illustrated in surprising and unexpected situations. In anticipated situations, the driver directs his/her attention autonomously on the conflict situation and, accordingly, does not benefit from such advisory warnings. When advisory warnings were provided, the strongest effect was found when surprising situations occurred. Furthermore, the study results provide data for an optimized advisory warning time frame. Advisory warnings must not be provided as early as possible. On the contrary, information about upcoming conflicts must be provided at least one second prior to the last possible warning moment, better still, two seconds beforehand, in order to have a positive effect on driving behaviour and the associated criticality of the situation. Data supplied even earlier to the driver do not contribute to the mitigation of the situation. Participants consistently preferred, independent from the investigated scenarios, to receive the advisory warnings two to three seconds prior to the last possible warning moment, rather than receiving information on imminent conflict situations. However, at this point it must be emphasized that the determined time frame for advisory warnings applies to the prototype implemented driver-vehicle interface. That is to say, it refers to discrete visual-auditory advisory warnings provided as a visual display element in the simulated HUD accompanied by an acoustic signal. It can be assumed that a more intensive signal, as usually applied during imminent crash warnings, will have an effect at a later point in time.

With regard to the *acceptance* of advisory warnings, it can be stated that such data are generally perceived by drivers as useful, provided the data are not presented at t_0. The usefulness of early advisory warnings (prior to the last possible warning moment) is consistently assessed as medium to high. Furthermore, the information received during scenarios that were still safely managed by the driver without assistance is considered helpful on average. The results emphasize that the information system for the avoidance of traffic conflicts, as it was implemented prototypically in this study, are generally perceived as useful by the drivers. However, the subjective assessment of the usefulness does not reflect in all cases an objective benefit in driving behaviour.

The increase in situation criticality for advisory warnings provided at t_0 during unexpected and surprising scenarios must be seen as a potentially problematic issue. Mathematically, from the moment of receiving the information, the driver has one second to start the activation of the brakes. Nonetheless, compared to non-assisted driving, lower TTA_{min}-values can be found in these scenarios. The following offers possible explanations: (1) adaptation of the brake reaction to the information system (in most assisted scenarios more reaction time is available), (2) an increase in speed during assisted driving compared to non-assisted driving, or (3) a distraction effect caused by reading the visual display element and thus requiring more response time.

A clarification of this effect was not possible within the scope of this study; however, subsequent studies are anticipated.

Acknowledgment

This work was funded by the Federal Ministry of Economics and Technology of the Federal Republic of Germany under grant no. 19 S 9022 and carried out in the joint Ko-PER research project within the scope of the Ko-FAS research initiative (www.kofas.de).

References

Bliss, J.P. & Acton, S.A. (2003). Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied Ergonomics, 34*, 499-509.

Brown, T.L., Lee, J.D. & McGehee, D.V. (2001). Human performance models and rear-end collision avoidance algorithms. *Human Factors, 43*(3), 462-482.

Campbell, J.L., Carney, C., & Kantowitz, B.H. (1997). *Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO)* (Technical Report FHWA-RD-98-057). Washington, D.C.: U.S. Department of Transportation, Federal Highway Administration.

COMSIS Corporation (1996). *Preliminary human factors guidelines for crash avoidance warning devices* (NHTSA Project No. DTNH22-91-07004). Silver Spring, MD: COMSIS.

Cummings, M.L., Kilgore, R.M., Wang, E., & Kochhar, D.S. (2007). Effects of single versus multiple warnings on driver performance. *Human Factors, 49*, 1097–1106.

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: Erlbaum.

Green, P., Levison, W., Paelke, G., & Serafin, C. (1993). *Preliminary human factors guidelines for driver information systems* (Technical Report UMTRI-93-21). Ann Arbor, MI: The University of Michigan Transportation Research Institute

Gupta, N., Bisantz, A.M., & Singh, T. (2002). The effects of adverse condition warning system characteristics on driver performance: An investigation of alarm signal type and threshold level. *Behavior and Information Technology, 21*(4), 235-248.

Ho, C. & Spence, C. (2005). Assessing the effectiveness of various auditory cues in capturing a driver's visual attention. *Journal of Experimental Psychology: Applied, 11*, 157–174.

International Harmonized Research Activities (IHRA) working group on Intelligent Transport Systems (ITS) (2008). *Statement of principles on the design of high-priority warning signals for in-vehicle intelligent transport systems – draft. Informal Document Number ITS-16-03*.

Lee, J.D., Gore, B.F., & Campbell, J.L. (1999). Display alternatives for invehicle warning and sign information: message style, location and modality. *Transportation Human Factors 1*, 347–375.

Lee, J.D., Hoffman, J.D., & Hayes, E. (2004). Collision warning design to mitigate driver distraction. In J. Herbsleb & G. Olson (Eds.), *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 65-72). New York: ACM.

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*.

Lenné, M.G. & Triggs, T.J. (2009). Warning drivers of approaching hazards: The importance of location cues and multi-sensory cues. In D. De Waard, J. Godthelp, F. L. Kooi & K. Brookhuis (Eds.), *Human Factors, Security and Safety* (pp. 203-211). Maastricht: Shaker Publishing.

McGehee, D.V., Brown, T.L., Lee, J.L., & Wilson, T.B. (2002). Effect of warning timing on collision avoidance behavior in a stationary lead vehicle scenario. *Transportation Research Record: Journal of the Transportation Research Board, 1803*, 1-6.

Neukum, A. (2011, 28.September). *Wenn das Fahrzeug mehr sieht als der Fahrer – Konsequenzen für die Gestaltung der Fahrer-Fahrzeug Schnittstelle.* Paper presented at the Ko-FAS Zwischenpräsentation, Schloss Aschaffenburg.

Neukum, A., Lübbeke, T., Krüger, H.-P., Mayser, C., & Steinle, J. (2008). ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen. In M. Maurer & C. Stiller (Eds.), *5. Workshop Fahrerassistenzsysteme - FAS 2008* (pp. 141-150). Karlsruhe: fmrt.

Rhede, J.G., Wäller, C., & Oel, P. (2011). Der FAS Warnbaukasten – Strategie für die systematische Entwicklung und Ausgabe von HMI-Warnungen. In VDI Wissensforum (Ed.), *6. VDI-Tagung - Der Fahrer im 21. Jahrhundert Fahrer, Fahrerunterstützung und Bedienbarkeit (VDI-Berichte 2134)* (pp. 63-75). Düsseldorf: VDI Verlag GmbH.

Schmidt, G. & Krüger, H.-P. (2011). Violating the driver's threat anticipation and the effect of different modalities for Forward Collision Warnings (FCW) In D. de Waard, N. Gérard, L. Onnasch, R. Wiczorek, and D. Manzey (Eds.) *(2011). Human Centred Automation* (pp. 85 - 98). Maastricht, the Netherlands: Shaker Publishing.

Sorkin, R.D. (1988). Why are people turning off our alarms? *Journal of the Acoustic Society of America, 84*, 1107-1108.

Spence, C. & Ho, C. (2008). Multisensory warning signals for event perception and safe driving. *Theoretical Issues in Ergonomics Science, 9*, 523 - 554.

Thoma, S., Lindberg, T., & Klinker, G. (2009). Evaluation of a generic warning for multiple intersection assistance systems. In D. de Waard, H. Godthelp, F. Kooi and K. Brookhuis (Eds.), *Human Factors, Security and Safety* (pp. 173-188). Maastricht: Shaker Publishing.

Totzke, I., Naujoks, F., Mühlbacher, D., & Krüger, H.-P. (2012). *Precision of congestion warnings: Do drivers really need warnings with precise information about the congestion tails position?* In D. de Waard, N. Merat, A.H. Jamson, Y. Barnard, and O.M.J. Carsten (Eds.) (2012). Human Factors of Systems and Technology (pp. 235-248). Maastricht, the Netherlands: Shaker Publishing.

Weidl, G. & Breuel, G. (2012). Overall Probabilistic Framework for Modeling and Analysis of Intersection Situations, *16th Int. Forum on Advanced Microsystems for Automotive Applications (AMAA 2012)*. Berlin.

Winner, H. (2009). Frontalkollisionsschutzsysteme. In H. Winner, S. Hakuli and G. Wolf (Eds.), *Handbuch Fahrerassistenzsysteme* (pp. 522-542). Wiesbaden: Vieweg & Teubner.

Drivers' hand positions on the steering wheel while using Adaptive Cruise Control (ACC) and driving without the system

G.F. Bianchi Piccinini[1], G. Prati[2], L. Pietrantoni[2], C. Manzini[2], C.M. Rodrigues[3], & M. Leitão[4]
[1]ICOOR-Interuniversity Consortium for Optimization and Operations Research, Italy
[2]University of Bologna, Italy
[3]Universidade do Porto, Portugal
[4]Instituto Politécnico do Porto, Portugal

Abstract

Adaptive Cruise Control (ACC) is an Advanced Driver Assistance System that maintains a predefined speed and headway to the vehicle in front. Despite the expected benefits, the effects of ACC usage on drivers' behaviour are not completely identified yet. In this context, the present study was performed to investigate drivers' perceived risk and mental workload associated with the usage of ACC, through the measurement of drivers' hand positions (high, medium, and low) on the steering wheel. In addition, the study also aimed to understand the effect of sensation seeking on the drivers' hand position. A study was conducted in a driving simulator, with 26 participants who drove twice, once with ACC and once manually, along the same route. Using Wilcoxon exact test, the results showed that the percentage of high and low hand position on the steering wheel did not change while driving with ACC or manually. Besides, no differences were found between the ACC driving condition and the manual driving condition on self-reported experienced mental workload and perceived risk. However, the score on the sensation seeking questionnaire was related to the hand positions selected by the participants.

Introduction

Adaptive Cruise Control (ACC) is an Advanced Driver Assistance System (ADAS) that partly automates the longitudinal driving task by maintaining specific speed and headway to the vehicle in front, according to the settings defined by the user. The device automatically maintains the distance to a vehicle ahead by reducing fuel flow and/or actively braking the vehicle. Through the partial automation of the driving task, ACC is expected to bring a reduction of drivers' mental workload, as already demonstrated in previous studies (Stanton et al, 1997; Hoedemaeker & Brookhuis, 1998). The usage of the system might also affect the users' perceived risk during

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

driving, considering that part of the task has been automated by the system (Rajaonah et al, 2008). Both variables (mental workload and perceived risk) are relevant for road safety research because they are frequently considered in studies related to road accidents (Oppenheim et al, 2010) and they also affect reliance on automation, as underlined by Riley (1996).

Mental workload can be defined as the capacity allocated to the performance of a task (De Waard, 1996) and, in the context of driving, it can be affected by several variables, including traffic density and the usage of automation such as ADAS/IVIS. On the other hand, risk perception refers to the subjective experience of risk in potential traffic hazards (Deery, 1999) and is determined by two inputs: information regarding potential hazards in the traffic environment, and information on the ability of the driver (and the capabilities of the vehicle) to prevent those potential hazards from being transformed into actual accidents (Brown & Groeger, 1988). Interestingly, previous research found that drivers' mental workload strongly correlates with drivers' perceived risk, in experimental conditions. In particular, in Fuller et al. (2008), it was found that drivers' ratings of risk were highly correlated with drivers' ratings of task difficulty (which is linked to mental workload). Similarly, Lewis-Evans and Rothengatter (2009) reported in their study that drivers' experience of task difficulty while driving in a simulator and feeling of risk during the driving task were correlated. In recent years, the assessment of hand positions on the steering wheel has been gaining validity for the measurement of drivers' perceived risk and mental workload. Compared to other measurements (travelling speed, headway to the vehicle in front and lane position), the position of the hands on the steering wheel can be easily assessed both in driving simulator and naturalistic driving studies, and, therefore, it deserves the consideration of road safety researchers.

In a first study conducted on the topic, Walton and Thomas (2005) assessed drivers' hand positions on the steering wheel of vehicles passing at different road sites. The hand positions were classified according to the number of hands on the top half of the steering wheel (that is above the dotted line in Fig. 1): two hands, one hand and no hands were the 3 categories identified. The results showed that 50% of drivers were observed with only one hand on the top half of the steering wheel. In addition, it was found that higher objective risk (related to traveling speed) led to more control hand positions.

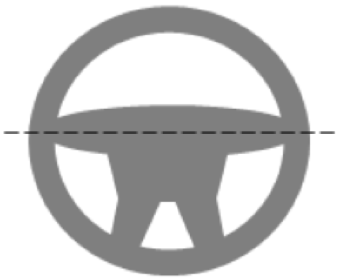


Figure 1. Hand positions measurement according to Walton and Thomas (2005)

In a subsequent study, Thomas and Walton (2007) adopted the same measurement to determine differences in hand positions on the steering wheel between SUV and car drivers. Overall, SUV drivers were observed to exert less control on the steering wheel. Given that SUV are deemed safer than other vehicles by SUV owners and by the general public (Davis & Truett, 2000), less control on the steering wheel shown by SUV drivers was related to the reduced perception of risk compared to car drivers.

Later, De Waard et al. (2010) studied the relationship between mental workload and perceived risk and hand positions on the steering wheel in a driving simulator study. Contrary to what had been done previously, the researchers adopted a different measurement to take into account also the bottom part of the steering wheel. The scoring of hand position was divided into high, medium and low control. High control meant that the left and right hands were located respectively in the left and in the right blue areas in Figure 2. Medium control meant that two hands were placed on the steering wheel, but only one (left or right) was located in the respective left or right blue areas. Finally, low control meant that the two hands were placed in the pink area in or that one hand or no hands were on the steering wheel. During the experiment, drivers had to perform a lane merging task in different traffic situations, merging lane length and number of cars in the acceleration lane. The results showed that, during the lane merging manoeuvre, a change in hand positions on the steering wheel seemed to reflect changes in drivers' mental workload and reported risk. Notably, drivers assumed a lower control position after merging into the highway, once they reached the left lane (higher speed lane). Besides, age differences were found within the panel concerning hand positions: older divers drove more often with a high control position, compared to younger drivers.

Figure 2. Hand position measurement according to De Waard et al. (2010)

In a following study, Fourie et al. (2011) collected observations of drivers' hand positions from the roadside and attempted to link them to travelling speed and headway to the vehicle in front. Concerning speeds, drivers who placed two hands on the top half of the steering wheel drove slower than drivers who placed one hand on the top half of the steering wheel. Similarly, drivers who placed two hands on the top half of the steering wheel had lower reciprocal headways than drivers who placed one hand on the top half of the steering wheel. Based on these findings, hand positions on the steering wheel were linked to drivers' characteristics (e.g., sensation seeking). In addition, the authors found gender differences in relation to hand positions.

On the whole, from the studies performed on the topic, risk perception and mental workload level seem to influence hand positions on the steering wheel. In this context, considering that Adaptive Cruise Control is supposed to influence driver's workload and perceived risk, then, the hand positions on the steering wheel should also be affected by the employment of ACC. In order to know more on the topic, in the present study, the attention was turned to how the use of Adaptive Cruise Control (ACC) might affect drivers' hand positions on the steering wheel. Based on ACC functioning, the following hypotheses were proposed:

1) The use of ACC will lead participants to adopt more frequently lower control hand positions, compared to driving without ACC.
2) The use of ACC will lead participants to adopt medium control hand positions as much as during driving without ACC
3) Driving without ACC will lead participants to adopt more frequently higher control hand positions, compared to driving with ACC.

The rationale behind these hypotheses is that the hands' positions on the steering wheel are related to the drivers' risk perception and to drivers' mental workload: during the use of ACC, drivers have lower perception of risk due to the partial automation of the driving task and, therefore, they might adopt a lower control hand position on the steering wheel. Furthermore, when using ACC, drivers have lower perceived workload and, therefore, as with perception of risk, they might adopt the position of lower control on the steering wheel. Based on this assumption, the following hypothesis was also proposed:

4) The use of ACC will be related to lower self-reported risk and self-reported mental effort, compared to driving without ACC.

Therefore, the first aim of this study is to investigate the effect of the use of ACC on drivers' workload and perceived risk, through the observation of drivers' hand positions on the steering wheel and through self-reported measures. A second aim is to investigate the relationship between drivers' hand positions on the steering wheel and the personal characteristic of 'Sensation Seeking', defined as the inclination to "seek varied, novel, complex and intense sensations and experiences and the willingness to take physical, social, legal and financial risk for the sake of such experience" (Zuckerman, 1994, p. 27). As previously reported, in a research work (Fourie et al., 2011), it was supposed that drivers' hand positions on the steering wheel are affected by personal characteristics and, in the specific case, by sensation seeking (which causes drivers to adopt higher speeds and lower headways). In order to confirm this assumption, the following hypotheses were also made:

5) Sensation seeking will be positively related to low hand positioning.
6) Sensation seeking will not be related to medium hand positioning.
7) Sensation seeking will be negatively related to high hand positioning.

Method and materials

Overall, 26 participants took part in the experiment, 24 males and 2 females with an average age of 25.3 years (*SD* = 9.0). They were selected among a group of ACC users and among staff of the Faculty of Engineering of the University of Porto.

The study was performed at the driving simulator of the Faculty of Engineering of the University of Porto. The simulator was composed of a Volvo 440 Turbo provided with the steering wheel, the pedals (accelerator, brake and clutch), the gear stick and the instrument panel. The simulated Adaptive Cruise Control allowed for the selection of the desired speed and headway to the vehicle in front through the controls on the steering wheel.

The participants drove twice in the simulated environment, once with ACC and once manually (without the system). The designed test route was a 46-km stretch of a real motorway in Portugal, divided into six sections, each one presenting different characteristics with regard to traffic conditions, length, speed limits and weather conditions. The simulated environment included other vehicles moving at 90 km/h on the right lane and at 120 km/h on the left lane. The routes were counterbalanced between the participants in order to avoid carry-over effects.

The participants were filmed during the routes by 4 cameras, recording respectively the simulated road, the driver, the steering wheel and the feet on the pedals. The hands' positions on the steering wheel were coded one of the author, according to the classification proposed by De Waard et al. (2010) and, therefore, divided into low, medium and high control. The time during which a participant kept his or her hands in one of the three positions (low, medium, high) was reported in seconds and, later, turned into percentages. Hence, each participant was characterized by three percentages (Low, Medium, High) for both driving conditions (ACC and manually). The coding of hand positions was performed only in Section 2 and Section 4 of the route because, in those stretches, the traffic was constant and no critical situations were planned.

In order to assess drivers' attitude towards sensation seeking, the Portuguese version of the Arnett Inventory of Sensation Seeking (AISS) questionnaire (Arnett, 1994) was administered. Before the first route, the pariticipants had to indicate, for each one of the 40 items, which defintion applied best for them on a four-point scale from '1 = describes me very well' to '4 = does not describe me at all'. The final score was made up of the sum of the two subscales (Novelty and Intensity), and the higher the score, the higher the driver's sensation seeking attitude. The internal consistency of the Arnett Sensation Seeking questionnaire was satisfactory, with a Cronbach's alpha value of .75.

In addition, after each route (with ACC and manually), the drivers were asked to fill in two questionnaires, one measuring perceived risk and the other one measuring mental workload. For the former, the participants had to choose what was the risk experienced during the route, on a scale ranging from '1 = no risk at all' to '10 = very high risk'. Likewise, for the mental workload, the drivers had to report the workload perceived during the task, on a scale ranging from '1 = no effort' to '10 = very high effort'.

For the purposes of this paper, the independent variables were the driving condition (drive along the same route with ACC and manually) and the score on the sensation seeking questionnaire. The dependent variables were the drivers' hand positions on the steering wheel and the ratings on the two questionnaires measuring perceived risk and mental workload. Notably, the drivers' hand position on the steering wheel was evaluated by the variation of the 2 independent variables, whereas the ratings on the two questionnaires were assessed exclusively by the variation of the driving condition.

Results

The Wilcoxon signed rank test was used to determine whether the percentages of respectively high, medium, and low drivers' hand position varied between the ACC and manual conditions. These statistical analyses were conducted with the SAS statistical package v.9.1. The results showed that the percentages of high hand positions did not significantly vary between the ACC driving condition (*Mdn* = 43.15) and the manual driving condition (*Mdn* = 46.75), $S = -47.0$, $p = .212$. The same happened with percentages of low hand positions in the non-ACC usage condition (*Mdn* = 18.15) compared to the manual condition (*Mdn* = 17.15), $S = 35.5$, $p = .378$. On the other hand, the percentages of medium hand positions were significantly lower in the manual condition (*Mdn* = 14.50) than in the ACC usage condition (*Mdn* = 30.20), $S = 72.5$, $p = .049$ (despite being this result at the limit of significance).

Concerning the two questionnaires applied after each route, no differences were found between the ACC driving condition and the manual condition on self-reported perceived risk ($S = -11.5$, $p = .570$) and mental workload ($S = -12.5$, $p = .562$).

To address the second aim of the present study, partial least squares (PLS) analysis was conducted with the SmartPLS Version 2.0 (M3) software (Ringle et al., 2005), because it does not require strong assumptions about the distributions of the data (Cassel et al., 1999; Chin, 1998) and it achieves high levels of statistical power even if the sample size is small (Chin & Newsted, 1999). To determine the inner weights, the centroid scheme was used and a uniform value of 1 was chosen as an initial value for each of the outer weights (Henseler, et al., 2009; Henseler, 2010). To evaluate the significance of the parameter estimates, a bootstrapping procedure was used, and implemented in SmartPLS. Specifically, 5,000 bootstrap samples with replacement were requested and the individual sign changes option was used (Henseler et al., 2009).

Figure 1 shows the relationships between sensation seeking and hand positioning. High and medium hand positioning were not significantly related to sensation seeking. However, consistent with our hypothesis, sensation Seeking was positively related to low hand positioning. Since SmartPLS does not compute significance tests for the variance explained in the dependent latent variables, the effect sizes of the R^2 values was calculated using Cohen's f^2. According to Cohen (1988), effect sizes of

.02, .15, and .35 are considered small, medium, and large, respectively. The effect size was moderate for low hand positioning (f^2 = .11), small for medium hand positioning (f^2 = .07), and very small for high hand positioning (f^2 = .01).

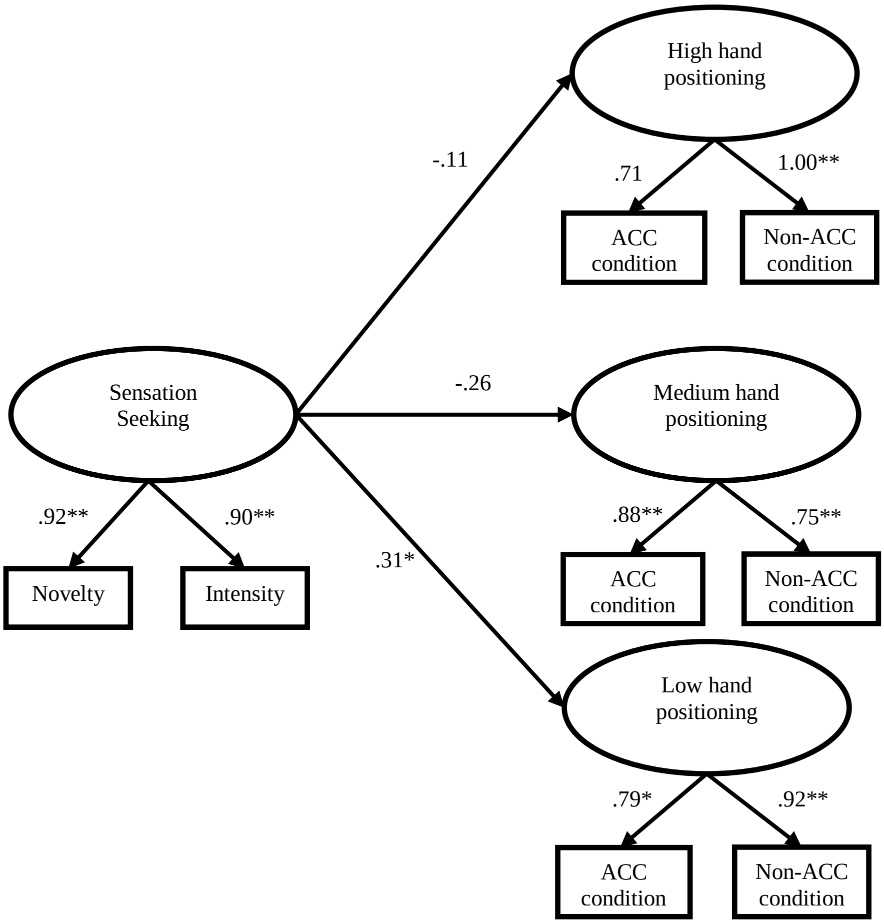


*Figure 3. Relationships between Sensation Seeking and Hand Positioning (Note. * p < .01; ** p < .001)*

Discussion

The present research was mainly conducted to detemine whether the use of ACC affects the perceived risk and the mental workload of drivers. In order to achieve this, two measurements were used: the observation of drivers' hand position on the steering wheel, and two questionnaires expressly designed for the study. The research also aimed to investigate the influence of sensation seeking on mental workload and on perceived risk, again looking at changes in drivers' hand positions on the steering wheel.

Regarding the first aim, the core assumption was that drivers tend to adopt lower control hand positions while driving with ACC, compared to driving manually. This assumption is not confirmed by the results, since low and high drivers' hand positions did not significantly vary between the ACC and the manual driving condition. Therefore, it can be assumed that use of the device does not lead drivers to perceive lower mental workload and lower risk while driving. This result is confirmed by the fact that no differences were found between the ACC driving condition and the manual driving condition on self-reported experienced mental workload and perceived risk. Those conclusions are partly in contrast with previous research (Stanton et al, 1997; Hoedemaeker & Brookhuis, 1998; Young & Stanton, 2004) that reported lower driver mental workload levels while driving with ACC. However, in contrast to this research, previous studies have provided drivers with highly demanding tasks and/or risky situations where mental workload and perceived risk are at a high level. In such conditions, the ACC could be useful in reducing these levels, through the assistance provided by the partial automation of the driving task. On the other hand, when the task is monotonous (as it was in the case of the present study), it is possible that the system cannot bring any reduction to the level of mental workload and perceived risk. In this discussion, it is interesting to note however that the percentage of medium hand positions is significantly lower in the manual driving condition than in the ACC usage condition. This result was not expected and it should receive closer attention in further research.

Concerning the second aim of this research, the results partly confirmed the hypotheses formulated. Indeed, sensation seeking was positively related to low hand positioning, to indicate that drivers with a high score on the sensation seeking questionnaire adopted lower hand positions on the steering wheel. This outcome might be seen as confirmation of the fact that sensation seeking is strongly correlated to risky driving (Jonah, 1997) and also show coherence with the results reported in Fourie et al. (2011).

Overall, two main results can be drawn from this study:

1) The use of ACC does not influence the driver's mental workload and perceived risk;
2) Sensation seeking significantly affects hand positioning, as an indicator of perceived risk.

Some limitations should be mentioned about the work performed. First of all, the small sample size (26 drivers) might have prevented the finding of relationships between the variables (usage of ACC and drivers' hand positions), and could have led to a reduced capacity for detecting their effects. Besides, the methodology used could be regarded as a second limitation, due to the uncertain level of fidelity encapsulated within the simulated environment, which can be both physical and functional (Stanton, 1996). Finally, another limitation is related to the coding of drivers' hand positions that has been done by a single researcher, and that might have caused an interpretation bias. However, it should be noted that this research was mainly designed as an exploratory study, and, therefore, the limitations mentioned above are acceptable. Besides, since this work addresses a quite new and unexplored research field, it can contribute to influencing the future research directions mentioned below.

First of all, the effects on driver behaviour caused by the use of Adaptive Cruise Control appear to be still rather unclear, and further clarification is required as to whether the device produces an increase or a decrease in drivers' mental workload and perceived risk. With respect to this, further investigations are also needed into whether these two constructs can actually be considered to be linked to each other, and therefore evaluated together. Besides, future studies should also confirm whether drivers' hand positions can be considered to be a reliable indicator of mental workload and perceived risk. Finally, in previous research it has been found that drivers' hand position is affected by several variables (e.g. age and gender). Due to the small number of studies on the topic, it would be interesting to once again test and validate those relationships in other contexts and situations, and to evaluate other factors (e.g. level of traffic, type of road and weather conditions), that could affect drivers' mental workload and perceived risk.

References

Arnett, J. (1994). Sensation seeking: A new conceptualization and a new scale. *Personality and Individual Differences*, *16*, 289-296.

Brown, I.D., & Groeger, J.A. (1988). Risk perception and decision making during the transition between novice and experienced driver status. *Ergonomics, 31*, 585–597.

Cassel, C., Hackl, P., & Westlund, A.H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics*, *26*, 435-446.

Chin, W.W. (1998). The partial least squares approach for structural equation modeling. In G.A. Marcoulides (Ed.), *Modern methods for business research. Methodology for business and management* (pp. 295-336). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Chin, W.W., & Newsted, P.R. (1999). Structural equation modeling analysis with small samples using partial least squares. In: R.H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 307–42). Thousand Oaks, CA: Sage.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (second ed.). Hillsdale, NJ: Erlbaum.

Davis, S.C., & Truett, L.F. (2000). An Analysis of the Impact of Sport Utility Vehicles in the United States. Available from internet: <http://cta.ornl.gov/cta/Publications/Reports/ORNL_TM_2000_147.pdf>.

Deery H.A. (1999). Hazard and Risk Perception among Young Novice Drivers. *Journal of Safety Research, 30*, 225–236.

De Waard, D. (1996). The measurement of drivers' mental workload. Ph.D.thesis, University of Groningen. The Netherlands: Traffic Research Centre, Haren.

De Waard, D., Van Den Bold, T.G.M.P.R., & Lewis-Evans, B. (2010). Driver hand position on the steering wheel merging into motorway traffic. *Transportation Research Part F: Traffic Psychology and Behaviour*, *13*, 129-140.

Fourie, M., Walton, D., & Thomas, J. A. (2011). Naturalistic observation of drivers' hands, speed and headway. *Transportation Research Part F*, *14*, 413-421.

Fuller R., McHugh C. & Pender S. (2008). Task difficulty and risk in the determination of driver behaviour. *Revue européenne de psychologie appliquée 58*, 13–21.

Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics, 25*, 107-120.

Henseler, J., Ringle, C., & Sinkovics, R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing (AIM)*, *20*, 277-320.

Hoedemaeker, M. & Brookhuis, K.A. (1998). Behavioural adaptation to driving with an adaptive cruise control (ACC). *Transportation Research Part F, 1*, 95-106.

Jonah, B.A. (1997). Sensation seeking and risky driving: A review and synthesis of the literature. *Accident Analysis and Prevention*, *29*, 651-665.

Lewis-Evans, B. & Rothengatter, T. (2009). Task difficulty, risk, effort and comfort in a simulated driving task—Implications for Risk Allostasis Theory. *Accident Analysis and Prevention*, *41*, 1053-1063..

Oppenheim, I., Shinar, D., Carsten, O., Barnard, Y., Lai, F., Vanderhaegen, F. & Kecklund, L. (2010). *Critical review of models and parameters for driver models in different surface transport systems and in different safety critical situations*. Deliverable No. D1.1, EU FP7 ITERATE Project. Retrieved August 31, 2013, from http://www.iterate-project.eu/.

Rajaonah B., Tricot N., Anceaux F. & Millot P. (2008). The role of intervening variables in driver–ACC cooperation. *International Journal of Human-Computer Studies 66*, 185–197.

Riley, V. (1996). Operator reliance on automation: Theory and data. In R. Parasuraman and M. Mouloua (Eds.), *Automation and Human Performance. Theory and Applications* (pp. 19-35). Mahwah, NJ: Lawrence Erlbaum Associates.

Ringle, C. M., Wende, S., & Will, A. (2005). SmartPLS (Version 2.0 M3). Hamburg, Germany: SmartPLS. Retrieved from http://www.smartpls.de.

Stanton, N.A. (1996) Simulators: research and practice. In N.A. Stanton (Ed.) *Human Factors in Nuclear Safety*. London: Taylor and Francis.

Stanton, N.A., Young, M., & McCaulder, B., 1997. Drive-By Wire: The case of driver workload and reclaiming control with Adaptive Cruise Control. *Safety Science, 27*, 149-159.

Thomas J.A., & Walton, D. (2007). Measuring perceived risk: Self-reported and actual hand positions of SUV and car drivers. *Transportation Research Part F, 10*, 201-207.

Walton, D., & Thomas, J.A. (2005). Naturalistic observations of driver hand positions. *Transportation Research Part F, 8*, 229-238.

Young, M. S., & Stanton, N. A. (2004). Taking the load off: investigations of how adaptive cruise control affects mental workload. *Ergonomics*, *47*, 1014-1035.

Zuckerman, M. (1994). *Behavioural Expressions and Biosocial Bases of Sensation-Seeking*. Cambridge: Cambridge University Press.

Ship voyage plan coordination in the MONALISA project: user tests of a prototype ship traffic management system

Thomas Porathe, Linda de Vries, & Johannes Prison
Maritime Human Factors, Chalmers University of Technology, Sweden

Abstract

EU has promised to reduce emissions by 80 % by 2050. For the shipping industry "slow steaming" for just-in-time arrival promises reductions of emissions. But a rapid increase in the construction of offshore wind farms planned in the North Sea may lead to ships facing a very complex and safety critical traffic environment in the future. Both of these issues bring ship traffic management to attention. In the Baltic Sea, the EU project MONALISA (Motorways & Electronic Navigation by Intelligence at Sea) has been looking at a voyage plan coordination system where a Ship Traffic Coordination Centre (STCC) handles a specific area, e.g. the Baltic Sea. A prototype system was developed and tested in a full mission bridge simulator environment for some simple scenarios. Qualitative data were collected; the main aim was to test mariners' acceptance of such a system. The participants were in general positive to the tested system; younger somewhat more than older. Some concern was expressed over risks of de-skilling and a common concern was the importance of the final control of the vessel resting with the captain on-board.

Introduction

This paper will present work on ship voyage plan coordination done in the TEN-T EU project MONALISA. Ship voyage plan coordination will be introduced from three perspectives: an environmental, an efficiency and a safety perspective.

The environmental perspective

EU has promised to reduce greenhouse gas emissions by 80 % by 2050 (European Commission, 2012). That is a substantial promise. The shipping industry is responsible for about 4-5 % of all greenhouse gas emissions globally (Harrould-Kolieb, 2008). That may not be much. But if the collective shipping industry was a country, it would be the 6th largest producer of greenhouse gas emissions. While land transportation can drive on electricity produced by wind, water and fossil-free fuels, these options are very limited for shipping. One large opportunity lies in what is called *slow steaming*. By reducing the speed of a typical container vessel by 30 %, a 50 % reduction in fuel consumption, and thereby also fuel costs and emissions, can

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

be achieved (Cariou, 2011). This is, however, a complex issue, as slower speed also means that more ships have to be engaged in order to maintain the transport capacity. The bottom line is that slow steaming has a potential to reduce emissions from the shipping industry. A study by Pierre Cariou (2011) shows that slow steaming has the potential of reducing emissions by around 11%, looking at data from 2 years back. This is close to the target of a 15% reduction by 2018 that was proposed by the International Maritime Organisation’s Marine Environment Protection Committee, 2009.

The efficiency perspective

Today, ports have a very close horizon. The harbour masters of Scandinavia’s biggest port, Gothenburg, and Humber Ports, one of U.K.’s largest ports, reveal that they do not know more than 2-3 hours beforehand whether a ship will arrive on time or not (personal communication 2012 and 2013). This leads to ports operating on a “first come, first served” basis which in turn leads to ships normally going full speed ahead and then anchoring up, issuing their notice of readiness. If slow steaming is to become a reality, a just-in-time, ship traffic management system with time slots has to be put in place that allows ship arrivals and departures to become predictable. This would in turn be beneficial to the entire transportation chain, on into the hinterland. Another factor is the increased complexity of ship traffic. The North Sea is an extremely busy shipping area and the English Channel the busiest strait in the world (133,444 ship passages in 2012).

A study in another EU research project, ACCSEAS, predicts that this number will increase to 200,000 by 2025 (ACCSEAS, 2013). To further complicate the picture, there is a rapid exploitation of sea areas for wind energy. In the German Bight, for instance, 10,000 wind turbines are currently being planned to replace German nuclear power. And to achieve the EU goal of 80 % reduction mentioned above, sea-based wind farms must keep expanding. In Figure 1 (left) the projected state of the North Sea is shown, with the amount of ship traffic passing the red lines (and the 2012 figure in brackets). The dark blue polygons are areas where today there is open sea, and tomorrow there will be wind mill farms. In Figure 1 (right) there is a close up of the “windmill city” of the German Bight, where shipping risks being conducted under more or less “street-like” conditions.

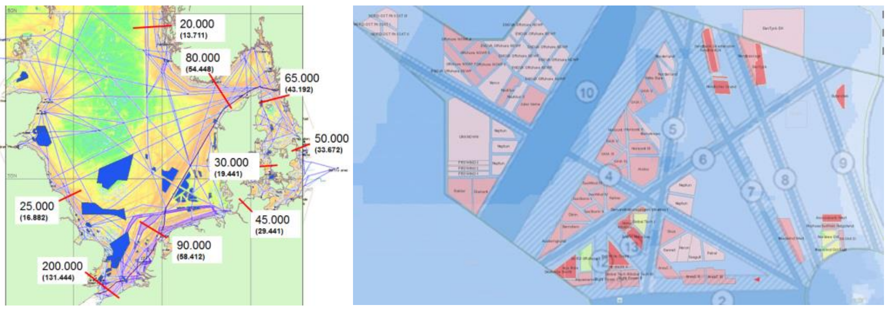


Figure 1. To the left, the predicted number of ships in the North Sea passing each red line in 2025 (the figures in brackets are the actual numbers in 2012). The blue areas are the planned

wind farms. To the right is a close up of the German Bight area, with the planned wind farms and shipping lanes (data from the ACCSEAS project, 2013).

However, if slow steaming is to be standard practice, a still larger increase in the number of ships is needed to not only keep up, but increase transportation capacity. Add to that the expected reduction in available sea space. Both these factors also bring ship traffic management to attention.

The safety perspective

On the 31 May 2003, the 225 metre long Chinese bulk carrier *M/V Fu Shan Hai* collided with the 100 metre long Cyprus registered container ship *M/V Gdynia*, about 3 nautical miles north of the island of Bornholm, in the southern Baltic Sea. The collision was the result of a failed avoidance manoeuvre by *Gdynia* as the give-way vessel. The collision ripped open a hole in the forward part of *Fu Shan Hai* which subsequently sank. The crew of 27 were all rescued (DMA, 2003).

Fu Shan Hai had departed from Ventspils, in Latvia, on 30 May 2003 at 16:20 local time on a voyage to China. *Gdynia* had departed from the port of Gdynia, in Poland, the same evening at 23:25 on a voyage to Hull, England. The *Fu Shan Hai* and *Gdynia* collision serves as an example of the 10-15 vessels every year that are totally lost after a collision at sea (IUMI, 2012). Although there is a lot to be said about the actions of the ships in the final stages before the collision, this paper will focus on the fact that both ships happened to be at the same place at the same time, the question being: could this have been avoided?

All larger, so-called SOLAS vessels are required to make a “berth-to-berth voyage plan” before leaving port, according to Reg. 34 of SOLAS Chapter V and IMO Resolution A.893(21) (IMO, 2010). When *Fu Shan Hai* left Ventspils, she had such a plan for her voyage across the Baltic Sea, up through the Belt and on all the way to China. Taking into account her service speed, the crew could calculate an estimated time of arrival in China. The same would be true for any point on the voyage. Also, *Gdynia* had such a voyage plan.

If these two plans could have been superimposed on top of each other in a central coordination system, it would, knowing the service speed of both vessels, have been possible to predict that they would be at some point at the same time. Then only a very slight change of speed would lead to the avoidance of the close quarters situation which occurred. However, sharing one’s voyage plan with a coordination centre is a controversial issue, challenging the captain’s ultimate control of the vessel. The presented study has been seeking to answer how the suggested MONALISA voyage plan coordination system is perceived by Swedish mariners.

MONALISA and the concept of route exchange

MONALISA is a TEN-T EU project, started in 2010 and which will be finished in 2013. The project has been coordinated by the Swedish Maritime Administration. Other partners in the project are Chalmers University of Technology, Sweden; Danish Maritime Authority; Finnish Transport Agency; Gate House, Denmark; SAAB Transponder Tech, Sweden; and SSPA Sweden AB. The MONALISA project deals with different maritime questions, but in this paper only the particular issue of route exchange will be considered.

In order to avoid yet another system on the bridge, the route exchange system needs to be integrated into the Electronic Chart and Display Information System (ECDIS) that already is central on a ship's bridge. ECDIS is a computerised information system that contains electronic nautical charts, and will become mandatory on deep sea vessels by 2018. The voyage plan is, for instance, made in the ECDIS and resides in the navigation computer, and thus could easily be shared if relevant infrastructure and standards were in place.

In order to test route exchange, a prototype "ECDIS-like" ship-based test platform was developed to mimic a standard ECDIS, but which was equipped with functionality to exchange routes. A corresponding shore-based station was also developed where routes could be received and compared. A first set of user tests, with simple cases of route exchange between only one ship and a shore centre, were carried out in a simulator centre at Chalmers University of Technology, in Sweden, with bridge officers and Vessel Traffic Service (VTS) operators.

The main research question was whether such a system would be accepted by the professional participants. Secondary questions had to do with workload, changes in procedures and usability improvements. Another major task was to look for unintended consequences of change, as new technologies always carry an inherent risk of new types of accidents.

Theory

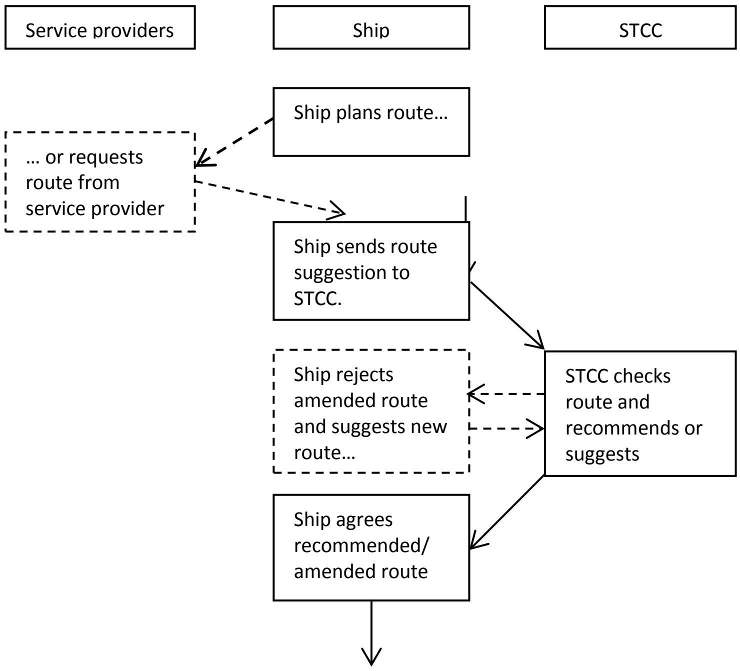
The theories behind this type of system are well-known to all human factors researchers. Donald Norman (1988) introduced the notion of *knowledge in the head and knowledge in the world*. A nautical chart is a representation of the physical world. It represents a crystallisation of knowledge from generations of hydrographers and geographers. By drawing a pencil line representing an intended voyage, knowledge is stored (in the form of checks for under keel clearance, risks etc.), and cognition is downloaded during the voyage. The projection of the chart is such that, by measuring the bearing of the intended path, the course to steer is immediately clear. The nautical chart is used, here, as a *cognitive tool* in a process of *distributed cognition*, investigated by Edwin Hutchins (1995). On a more general level, this is all part of the area of Joint Cognitive Systems (Hollnagel & Woods, 2005), and earlier, Cognitive Systems Engineering (Rasmussen et al., 1994), to mention just some important examples.

Method

In order to test ship traffic management in an ecologically valid way, route exchange needed to be tested on a practical level in a ship and shore simulator. The first step was to design a prototype application where captains and shore operators could experience, first hand, the effects of such a system. Although much focus was put on the conceptual level of acceptance within the maritime community, it was also necessary to ascertain that procedures on the bridge were not changed in such a way that would increase workload or require more manning.

Information flow in route exchange

The information flow can, from a user perspective, be described as follows. The voyage is planned on-board as usual (or maybe requested from a service provider, e.g. a weather routing service or a route library). Once the route has been validated (checked for under-keel clearance, and that it follows fairways and traffic separation schemes), and departure and arrival times have been added, the route is sent to the STCC. The route is now "pending" (signified by a yellow dashed pattern in the chart systems). The STCC checks the route, (checks it for under keel clearance, violation of NoGo areas and loss of separation, i.e. conflicts with other ships or traffic congestions, not tested in this study). The STCC can now either "recommend" the route as it is, or send the route back with suggested changes (signified in the charts by green dashed/"recommended" and red dashed/"not recommended" routes). If the STCC suggests changes, the ship can either accept or reject these changes, or suggest another modification. Routes can be sent back and forward until an agreement is reached and a final solid green/"agreed" route is reached (see Figure 2).



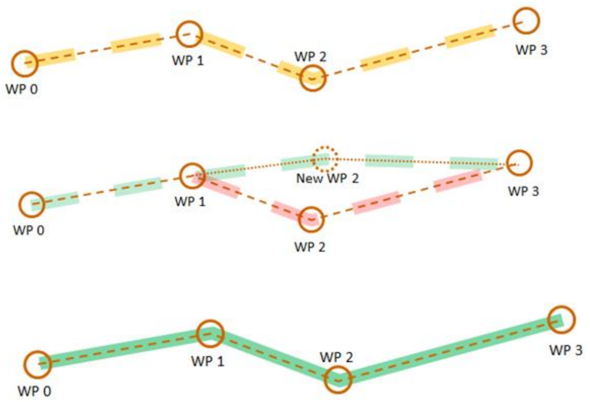


Figure 2. On top, is a simplified diagram of the information flow for the route exchange process. To the bottom, is the symbology used in the electronic chart systems to signify (top) a "pending" route sent from ship to shore, now awaiting optimisation, then a stage in a route negotiation (middle) where a deviation is "recommended" (green) instead of the original (red) segment. Finally (bottom), the "recommended" route has been "acknowledged" by the ship and becomes green/"agreed".

Two prototype systems were developed and tested during this study: one from SAAB Transponder Tech and one from the Danish Maritime Authority. Both systems had a ship and a shore display, and there were no significant differences between them. In this paper, we will make no distinction between the two systems.

The prototype systems were set up in the full mission bridge simulator environment at Chalmers University of Technology in Gothenburg, Sweden, using one bridge and one shore station. AIS targets and routes were sent between ship and shore using prototype protocols developed for the test. The technical details of these protocols will later be presented through the MONALISA 2 project.

The first user tests with simple cases of route exchange between one ship and a shore centre were carried out in two sessions, in May and September 2013, with 12 ship captains, 11 maritime academy 4th year cadets and 5 VTS operators. This gave a sample of subject matter experts: 25 male, 3 female; ages 27-66, mean age 42; professional experience from none (cadets) to 40 years, the mean being 15.7 years of professional experience (cadets excluded).

Each block of the study consisted of 4 scenarios where one captain and a cadet manned the ship bridge and a VTS operator the STCC. Each such block took about 3 hours to conduct, including familiarisation with the prototype systems, the 4 scenarios and a debriefing session.

Scenario 1: Route planning initiated by ship. The ship is anchored outside Hirtshals in the northern part of Denmark and the bridge crew plan a route from their present position to Helsingborg in southern Sweden along the standard "T-route" (the deep water route through Kattegat). Once planned, the route is sent to STCC. STCC checks the route and sends a “recommended” route back. The ship finally “acknowledges” the route (see Figure 3, left).

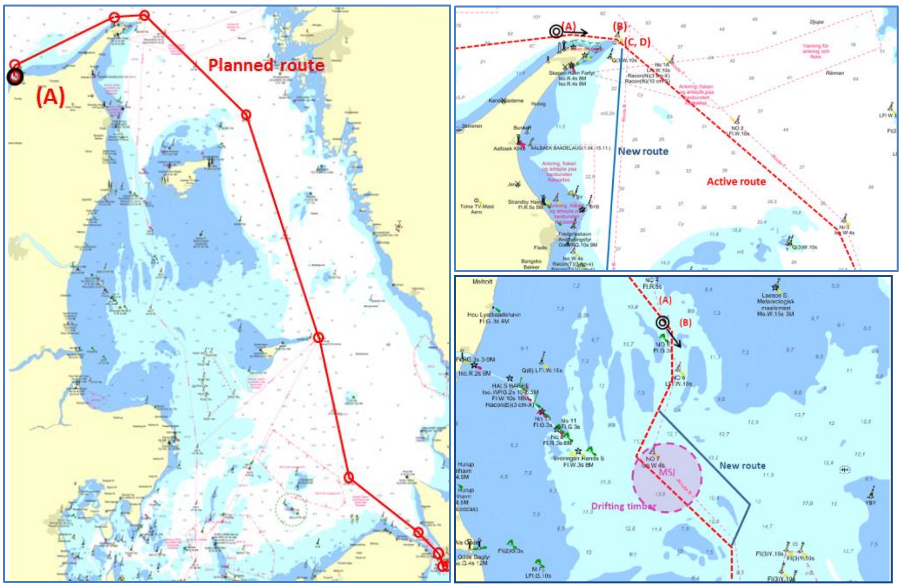


Figure 3. Three of the four test scenarios. Left scenario 1 with the route from Hirtshals to Helsingborg that the bridge crew was to plan. Top right, scenario 2, a route change done en route, initiated by the ship. Bottom right, scenario 4: drifting timber causing STCC to suggest a route deviation. (The images are just scenario descriptions, not ECDIS with MONALISA symbology).

Scenario 2: Route change initiated by ship. An hour later, the ship is en route to Helsingborg on its agreed route, just north of Skaw, on the northern tip of Denmark. The engine room calls and says they will need to do some repair in the engine room. They ask if the bridge can provide calmer conditions. The captain decides to change the route to the “B-route” (in the west part of Kattegat), in the sheltered waters along the eastern shores of Denmark. The new route is planned and sent to STCC (see Figure 3, top right).

Scenario 3: Route change initiated by STCC. There has been a delay in port operations at Helsingborg, and a request to slow down is sent from STCC to the ship. New ETAs (Estimated Times of Arrival) are suggested in the waypoints ahead. The change is negotiated and finally agreed.

Scenario 4: Route change initiated by STCC. A ship ahead has lost part of its deck cargo (timber), which is now floating in the water. An MSI (Maritime Safety Information) is sent out and displayed in the chart (see Figure 3, lower right). STCC suggests a route change around the dangerous area, which is negotiated and agreed (see Figure 4).

Figure 4. Watch officer and captain on the simulator bridge. The route is visible in the prototype screen (scenario 4, before the MSI with drifting timber has been received).

Data collection

Qualitative data were collected using video recordings during sessions, and a debriefing with both ship and shore participants after the four blocks. An observer was also present both in the ship simulator, and at the shore station, taking notes and prompting participants to think aloud. A questionnaire was also sent out by mail to the participants after a couple of weeks.

Results and discussion

Observations, comments and statements made during the debriefings were classified into four levels: conceptual, procedural, functional and HMI. Results on the detailed levels of functions and HMI will not be presented in this paper.

Conceptual level

The hypothesis was that ship-board participants would be negative towards the route exchange concept, but instead all participants were, in general, positive towards the concept of voyage plan coordination; younger somewhat more than older. But even if older participants were more concerned with issues like de-skilling, they still accepted the system. A pensioned captain with 40 years of experience said: “I don’t like this, but I see it coming, and I guess it is alright.”

The most discussed issue was that of control; if voyage plan coordination were to lead to control being shifted from the ship to the shore. Most bridge officers pointed out that it was important that the captain still had the last word, being on the scene and experiencing the situation first hand. Several participants saw a likelihood of conflicts between the STCC and vessel on the issue of control, and between the STCC and ship owners on the issue of costs.

From a shore perspective, the ability to check routes and see vessels' intentions was welcomed, but concerns were raised about workload when dealing with several vessels in a heavy traffic or emergency situation.

On the question of whether a route exchange system has a future, comments ranged from pointing out that it is inevitable, to stating that it may have a positive effect on the quality of navigation if captains can learn to trust it. Some comments were to the effect that it will never be accepted by captains and ship owners.

Procedural level

Participants in both vessel and STCC felt that new routines were involved in operating the system, but within a familiar environment, so that, once they understood what was expected of them, and how to do it, it was easy to get accustomed.

All differentiated between the planning and monitoring functions of the system, and the need for separate routines for each. In most cases, physical separation of the functions, either separate screens or a separate workplace, were suggested. It was pointed out that individual vessels have different routines for route planning, either first using paper charts, or using the ECDIS straight away, which, depending on the method currently employed, could lead to a change in routine.

With regard to vessel manning, some felt that an extra person may be needed on the bridge in heavy traffic situations, to leave the on-watch officer free to navigate and avoid collision. It was also felt by some that, depending on the degree of freedom allowed to deviate from the planned course, implementation of such a system would result in extra workload for the captain, who would be required to approve all changes made.

Routines in the STCC, as opposed to the existing VTS stations, would depend on the role of the STCC. Issues raised by the participants were workload, the capacity to deal with heavy traffic, the ability to plan routes around obstructions and the time constraints involved. Most participants felt that the role of the STCC in an MSI situation should be limited to entering the area on the chart and sending to affected vessels.

Survey

In a survey that was sent to the participants several weeks after they participated in the study, the question *What is your opinion about the tested route exchange concept?* was asked. Eighteen answers where received, out of 28, and 14 were "positive" or "very positive", and 4 "did not know". No one was negative. On the question *Do you think a similar route exchange concept will become reality in the future?* 17 answered "probably" or "most probably" and only 1 answered "probably not".

Conclusion and future studies

Most participants, both younger and older, were more or less positive towards the ship traffic management concept.

Having said that, there was discussion on the yet undecided scope of the proposed route exchange system and the role of the STCC; would it be monitoring, advisory, assistance or full control? Would it involve a change to the established principle that the captain is ultimately responsible for the vessel? Would the captain be required to relinquish some degree of control of the vessel to the shore centre? Where would responsibility and liability lie for delays, costs incurred, accidents etc.? Several participants mentioned the likelihood of conflict between the STCC and vessel on the issue of control, and between the STCC and ship-owners on the issue of costs. All participants agreed that the final decision needed to stay with the captain on-board.

One has to be aware of the limited validity of these findings, considering the small number of participants and the cultural homogeneity (all Swedish). Also, the tested scenarios only involved one ship, without the complexity of routing several ships. Future studies will target both these limitations. Preparations for test-beds in other parts of the world have started, and, in the MONALISA 2 project, complex simulation involving more ships will be conducted.

Acknowledgements

This research has been possible thanks to funding from the European Union's Seventh Framework Programme for infrastructure TEN-T, and also from the Region Västra Götaland.

References

ACCSEAS (2013). www.accseas.eu

Cariou P. (2011). Is slow steaming a sustainable means of reducing CO_2 emissions from container shipping? Elsevier, Transportation Research Part D: Transport and Environment, 16 (3), 260–264

DMA, Danish Maritime Authority (2003). CASUALTY REPORT: Collision between Chinese bulk carrier FU SHAN HAI and Cypriot container vessel GDYNIA. http://www.dmaib.dk/Sider/Ulykkesrapporter.aspx

European Commission (2012). A Roadmap for moving to a competitive low carbon economy in 2050. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52011DC0112:EN:NOT

Harrould-Kolieb, E. (2008). *Shipping impact on climate: a source with solutions*. Washington DC: Oceania.

Hollnagel, E., & Woods, D.D. (2005). *Joint cognitive systems. Foundations of cognitive systems engineering*, Boca Raton, FL; London: Taylor & Francis

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Graham, P. (2012). Casualty and World Fleet, Statistics as at 01.01.2012. London: IUMI (International Union of Marine insurance), Facts and figures committee.

IMO, International Maritime Organization. (1974). International Convention for the Safety of Life at Sea (SOLAS), 1974. Amended several times. http://www.imo.org/About/Conventions/ListOfConventions/Pages/International-Convention-for-the-Safety-of-Life-at-Sea-(SOLAS),-1974.aspx

Marine Environment Protection Committee. (2009). *Analysis on the Appropriate Values of the Reduction Rates of the Required EDDI*. 60/4/36 Submitted by Japan, 15 January 2010.

Norman D.A.(1988). *The psychology of everyday things*. New York: Basic Books.

Rasmussen, J., Pejtersen, A.M., & Goodstein, L.P. (1994). *Cognitive Systems Engineering*. New York: Wiley.

Prevention of physiological and psychological stress at computer-equipped workplaces

Piia Tint, Kalju Meigas, Viiu Tuulik, Viive Pille, Kristel Oha, Karin Reinhold, Deniss Karai, Varje-Riin Tuulik, Mariliis Lauri
Tallinn University of Technology
Estonia

Abstract

This investigation was carried out in the frames of Interreg 4A project "Workability and social inclusion" headed by the Arcada University of Applied Life. Tallinn University of Technology and Riga Stradins University were involved. A questionnaire based on the Nordic, WAI (Workability index), and Kiva questionnaires was compiled to study psychosocial and physical working conditions at computer-equipped workplaces for 295 workers. Occupational hazards were measured. To investigate the level of stress of workers the cortisol level in saliva was measured. After the first questioning of workers on the psychosocial factors at workplace, the Metal Age programme was implemented and after the intervention, the Kiva questionnaire and the cortisol level in saliva were measured again. The results showed that if the preventive measures for solving the problems at workplace are implemented, and employers and employees are trained and consulted using the appropriate programmes, the stress situations could be avoided. The workers were divided into two groups (under 40 years- A and over 40 years- B). These groups were found to differ in the perception of psychosocial risk factors at the workplace. Group B assessments for the psychosocial working conditions were better than those of group A. In group B employees appeared to be more afraid of losing their jobs. Therefore, the employees in group B were not so interested in the work atmosphere as in group A. It is not easy to find a new position after losing one at the age over 40 years.

Introduction

The work was carried out in the frames of Interreg 4A project" Workability and social inclusion" headed by the Arcada University of Applied Life of Finland. The number of occupational diseases is the specific indicator that influences hazards on the worker in the work environment. Occupational diseases in Estonia are usually diagnosed in the late stage when the worker is already disabled. In this late stage it is difficult to find proper rehabilitation methods for a total health recovery. The majority of occupational diseases in Estonia are connected with musculoskeletal disorders (MSDs). Intensive use of computers causes major health problems like tissue damages, imbalance in blood flow, formation of the carpal tunnel syndrome (Oha et al., 2010). Computer-workers are under pressure as increasing amounts of

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

work have to be done within limited time. Stress is not only a feeling. It changes functions in the body: release of a variety of hormones, increased breathing, quickened pulse, and the production of more stomach acid. Computer work causes social problems: it distracts an individual from normal social or family relations and this in turn may lead to depression (Eltayeb et al., 2007). The interaction between the body and the work environment is complicated and four important systems (central nervous, automatic nervous, endocrine and immune) are involved in this network (Raja et al., 1996).

There are numerous hazards in the work environment (low temperatures, draught, noise etc.) that affect office workers and can damage the peripheral and central nervous system (Tint et al., 2012a). The physical complaints are very closely connected with psychological disturbances of computer workers (Zakerian & Subramaniam, 2009). The first symptoms of stress to appear can be physical (lack of appetite, sleeplessness, headache, pain in chest) and psychological (difficulties to concentrate, imbalance, anxiety, irritability, difficulties to make a decision, burnout). Illnesses and social problems develop under permanent stress. In stressful situations, the emotions created increase adrenocorticotropic hormone activating adrenal cortex, which extract cortisol. Changes in the cortisol level is an indicator of stress level (Melamed et al., 1999; Hofman, 2001). It is necessary to develop solutions for decreasing unreasonable stress and anxiety in workers suffering from stress situations at work.

Stressors like time limits, bad relationships between co-workers or with the employer, too much work wanted by the employer are considered to be the factors that can cause fatigue in the upper extremities (Feuerstein et al., 2004; Kulin & Reaston, 2011; Panari et al., 2012).

Lahtinen et al. (2002) focus on the psychological and social dimensions of managing and solving indoor air problems in their interviews and questionnaires. Their interviews were very critical of the process of solving the indoor air problem. The study supported the hypothesis that psychosocial factors play a significant role in indoor air problems.

Brauer and Mikkelsen (2010) studied the psychosocial work environment - at the individual and workplace level. The moderate differences between the workplaces in the perception of the indoor environment, as well as large differences between the individuals in the same building indicate that some occupants do perceive problems in the indoor environment even in the absence of a general indoor problem in the workplace. The authors also mention that the complaints are usually “over-reported” or over-estimated.

Stuffy air, noise, temperature, lighting deficiency might be the supplementary risk factors for developing MSDs and psychosocial stress at workplaces (Tint et al., 2012a). The main physiological and psychological stress factor is a poorly designed workplace (Tint &Traumann, 2012b).

Social relationships are important for the physical health of workers (Eisenberg & Cole, 2012). Socially connected people live longer than socially isolated people and the first have increased resistance to a variety of somatic diseases ranging from heart disease to cancer (Miller et al., 2009).

Studies that consider the working conditions, physical overload and psychosocial risk factors are complex and have not been conducted in Estonia until now. Therefore, the results of the project "Work ability and social inclusion" are very important, offering ideas for further research to improve of the psychosocial work environment. The results could also be implemented in other post-socialist countries as the early stage in the work environment was almost the same for all of these states.

The research question: is it possible to reduce the physical and psychosocial risk at workplaces by speaking with people, training them and solving the problems regarding the issues of their complaints? Different hazardous factors (indoor climate, psychosocial factors, static posture etc.) are influencing the computer-worker (figure1). If the improvement methods in the working environment are implemented, the level of stress of workers has to be decreased.

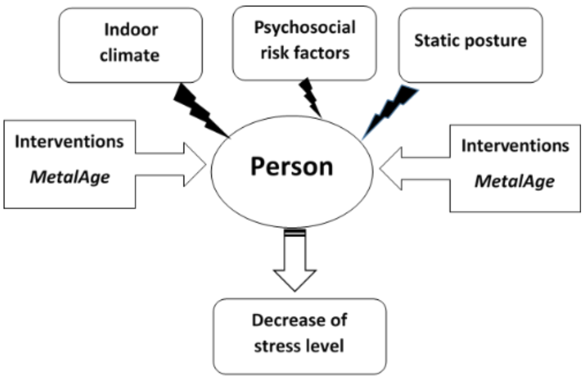


Figure 1. The structure of the study

Material and methods

The aim of the investigation was to increase the work ability, stress management, good leadership behaviour and thereby social inclusion of the workforce. Therefore, the Metal Age programme (Näsmän, 2011) is perfectly relevant for this purpose. A questionnaire based on the Nordic, Workability index and Kiva questions was compiled and forwarded to the workers by the Internet or as copy before the intervention of the Metal Age programme. 295 workers from different office-rooms from different institutions (including high school) were interviewed about the health risks and health disturbances connected with their work. The health disturbances were specified.

The Kiva questionnaire after the Metal Age programme implementation was conducted with 136 persons selected from four institutions (from those most interested in co-operation and improvements in the working environment, figures 2-5) and the cortisol level of computer-workers was measured before and after the Metal Age implementation for the same persons in 29 cases.

Figure 2. Office in institution 1

Figure 3. Office in enterprise 2

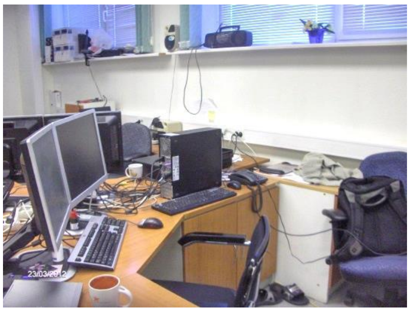


Figure 4. Office in enterprise 3

Figure 5. Office in high-school

Questionnaires

The Nordic (Lindström et al., 2000), WAI (Tuomi et al., 1998) and Kiva (Näsman, 2011) questionnaires were used to investigate the stress factors arising from the relationship between the employees and employers at the workplace.

Work Ability Index (WAI) is determined on the basis of the answers to a series of questions that take into consideration the demands before the interview with an occupational health professional who rates the responses according to the instructions.

Kiva questionnaire characterizes the wellbeing of workers at work. The ratings were given in an 8-point scale (1- not at all, 8- very much so, certain or well). The Kiva questionnaire was composed of seven questions:

1. Have you enjoyed coming to work in the last weeks?
2. I regard my job meaningful
3. I feel in control of my work
4. I get on with my fellow-workers
5. My immediate superior performs as superior
6. How certain are you that you will keep the job with this employer?
7. How much can you influence factors concerning your job?

Measurements of working conditions

The indoor air conditions were measured using the following standards and measuring equipment: EVS-EN-ISO 7726:2003 “Thermal environments- Instruments and methods for measuring physical quantities”; EVS-EN 15251:2007 “Indoor environmental input parameters for design and assessment of energy performance of buildings addressing indoor air quality, thermal environment, lighting and acoustics”, EVS-EN 12464-1:2011 “Light and lighting- Lighting of workplaces- part 1: Indoor work places”, EVS 891:2008 “Measurement and evaluation of electrical lighting in working places”. The measurement equipment used for microclimate: TESTO 435. TESTO 435 enables also the measurements of CO_2. Workplaces lighting and and screens were measured using the light-metre TES 1332 (ranges from 1-1500 lx). The lighting was measured on the worktable, on the screen and on the keyboard. Dust was measured with HazDust EPAM-5000.

Determination of fatigue in muscles

Myometer “MYOTON-3” was used to diagnose the functional state of the skeletal muscles of office workers. Myometer is a hand held state device developed at University of Tartu, Estonia (Vain & Kums, 2002). Myoton exerts a local impact on the biological tissue by means of a brief mechanical impulse. The impact force is small enough so that it causes no changes in the neurological reaction of the biological tissue. The tissue responds to the mechanical impact with damping or oscillation which is registered by an acceleration sensor located on the measuring tip of the device (figure 6 and 7).

Figure 6, 7. Myometer for muscles stiffness measurements

Determination of cortisol level

The high-performance liquid chromatography method (HPLC Water Alliance with UV detection) was used to determine cortisol in saliva (Kawasaki et al., 1979; Kalman et al., 2004). Saliva samples were collected three times during the day: in the mornning (8-9), at noon (12-13) and in the afternoon (16-17). Each participant was asked to hold special sampling tubes “Salivette” in their mouth for three minutes. The samples were analysed by the Laboratory of Hygiene and Occupational Diseases in Riga Stradins

University.

Results

For statistical data processing (except working environment hazards measurements), the computing programme SPSS.13 (Windows) was used. The arithmetic mean and standard deviation (SD) were calculated. To ascertain connections between the characteristics, the Spearman rank correlation (r- correlation coefficient) was applied, differences between the groups were tested with a Student's t-test. The difference $p<0.05$ was considered statistically significant.

Measurements in the work environment

Table 1 shows the results of measurements in the work environment. In winter the humidity of the air is too low. By the norms (EVS-EN 15251:2007), relative humidity of 40-60% is required for the worker to feel comfort. The level of carbon dioxide ~1000 ppm is felt by the workers as poor microclimate. The lighting of workplaces equipped with computers is usually good, in the frames of norms (300-500 lx), but sometimes infotechnologists prefer working in dark (without electrical lighting). However, this situation must be avoided.

Table 1. Results of measurements indoors in offices (2012-2013)

Room type	*T, ^0C Cold/warm season U=0.6^0C*	*R, % Cold/warm season U=2.0%*	*L, lx U=10.4%*	*CO_2, ppm U=10%*	*Dust, mg/m^3 U=10%*
Office 1, Figure 1	20-22/ 28-30	22-23/ 35-65	495-890	537-998	0.030
Office 2, Figure 2	20-22/ 24-28	15-25/ 35-75	200-250	500-750	0.020
Office 3, Figure 3	18-22/ 20-24	20-30/ 40-74	350-600	350-1200	0.015
Office 4, Figure 4	17-20/ 22-28	15-30/ 40-70	690-1209	478-1152	0.011

U- the uncertainty of measurements; T- temperature of the air; R-relative humidity; L-lighting; CO_2- concentration of carbon dioxide in the air; Dust- dust concentration in the air

Responses to the questionnaires

Table 2 presents the results of the survey involving 295 office workers working with computers (94 men and 197 women) about their health disturbances. The respondents were divided into two groups: A, age <40 (40 not included) - 137 persons and B, age >39(40 included) - 152 persons. The average age of group A was 30.97 years and group B - 54.5 years (*SD* 10.0). People who responded, had been working in the same occupation for 4.81 years (group A) and 17.38 years (group B, *SD* 9.0), respectively. Over 90% of the respondents were engaged in mental work in both groups (A and B). In group A, MSDs were observed by 53.6% of the respondents; the cardiovascular disturbances were observed by 20% of the respondents (A); visual disturbances occurred in 16% of persons (A). The problem of overweight in group A occurred in 20% of the respondents. Diabetes occurred in two people. In group B MSDs were observed by 50% of the respondents; cardiovascular disorders by 45% of the respondents; visual disturbances occurred in 23% of the respondents. The results from the questionnaire show that the computer workers assess their health status considerably high. They are optimistic in solving the problem that the monotonous work with computers will continue and believe that their health status in the future will stay on the same level using the steadily enhancing rehabilitation means.

Table 2. Health complaints according to the questionnaire (Nordic, WAI)

Group/ Disturbances	*A (persons <40- included; years of age), % of all investigated*	*B (persons >40 years of age), % of all investigated*
MSDs	53.6	50.1
Cardiovascular disturbances	20.0	45.0
Visual disturbances	16.1	23.2
The problem of overweight	20.2	25.4
The health status good	55.2	43.4

Considering the fact that people begin to work with computers at ever younger age, according to the current investigation, more MDSs are observed by young people (<40 years of age) than by older workers (>40 years of age). The positive result is that 43.4% of the older workers consider their health status good. Visual disturbances were diagnosed for group A and B almost at the same rate. It shows that eyes get tired and sight is worsening already at a young age.

Determination of fatigue in muscles

MSDs were examined in depth for 66 display unit workers. Nordic questionnaire was used beforehand to find out the persons who might have MSDs with the aim to diagnose the disease in the early stage. Myoton allows pollide abductors muscle tone and muscle stiffness to be measured. Pain intensity was evaluated on a 10-point scale. The body mass index was assessed, average being 25.0 (*SD* 5.0).

The study group consisted of 37 women and 29 men, at the mean age of 41.7 years and the average length of service of 9 years. Employees declared the average working time with displays at 7.14 hours each day.

Muscle and joint complaints were reported only for 13 workers (19.7%). The majority of the respondents declared the existence of two or more local pain points. Neck pain complaints occurred in 37 respondents (56%), and the severity of pain was assessed at an average of 4.18. Right shoulder pain occurred in 22 patients (33.3%) and left shoulder pain in 18 (27.3%) of the respondents. Pain in shoulders was assessed with 3.80 (right) and 2.80 (left) balls. Wrist pain in the right arm was declared in 13 cases (19.7% of all) and the left wrist pain only in three cases (4.5%); the severity of 4.57 (right) and 4.01 (left). Back pain was complained by 25 (37.8%) persons with severity of 4.28 balls (table 3).The presence of pain was generally of short-term duration, mostly for 1-7 days.

Table 3. The results of measurements of fatigue in muscles with Myoton

Pain region	*Number of workers (% of all investigated)*	*Severity of pain (0-10)*
Neck	37 (56%)	4.18
Shoulder, right	18 (27.3%)	3.80
Shoulder, left	17 (25.7%)	2.80
Elbow, right	4 (6%)	4.71
Elbow, left	4 (6%)	2.12
Wrist, right	13 (19.7%)	4.57
Wrist, left	3 (4.5%)	4.01
Back	25 (37.8%)	4.28

Myometric study revealed differences in the pain severity of complaints from employees with muscle pain and trapeze muscle pain in tonus and muscle stiffness.
As muscle strain is coming from the static posture, it can be influenced by the work psycho-emotional stress.

Thus, it might be concluded that static muscle tension causes increased muscle tone. Loaded in front of the screen is mostly the guiding hand. It is important to find the organizational measures that would ensure regular breaks and exercises, possibly combining them.

Responses to the Kiva questionnaire

The Kiva questionnaire was conducted twice: before the intervention of the Metal Age programme and after it (table 4). The number of used questionnaires was 136. According to the Kiva questionnaire, the investigated workers had high satisfaction coefficients with work. In addition, the stress indicators did not evaluate the stress-levels high. Kiva methodology for deploying a team is intended to find solutions to the specific results of the work of the team. It is connected with the character of work to find suitable working arrangements to be developed and therefore it is suitable for work-related psychosocial and MSDs prevention.

Table 4. Responses to the Kiva questionnaire

The question	*Office 1 B/A**	*Office 2 B/A*	*Office 3 B/A*	*Office 4 B/A*
1	5.0/7.1	6.2/7.3	7.3/7.8	7.1/8.1
2	5.3/8.0	8.4/7.8	8.3/8.7	8.3/8.2
3	8.0/7.7	7.7/7.6	8.3/8.4	8.2/7.9
4	7.8/8.0	8.4/8.9	9.0/9.1	8.5/8.9
5	3.5/7.7	7.3/7.8	8.3/8.5	7.4/8.3
6	5.5/8.0	7.2/8.1	8.8/7.3	8.1/8.2
7	5.8/4.9	7.0/7.6	7.5/7.4	6.7/7.4

B/A*- the mean values before the Metal Age intervention/ after the Metal Age intervention

The changes in the Kiva questionnaire are shown in Figure 8. The relations between the employer and the employees were usually improved (questions 1-2, 5-7). A slight decrease appeared in questions 3, 4: Does your immediate superior help you develop your skills? Does your immediate superior tackle problems as soon as they surface?

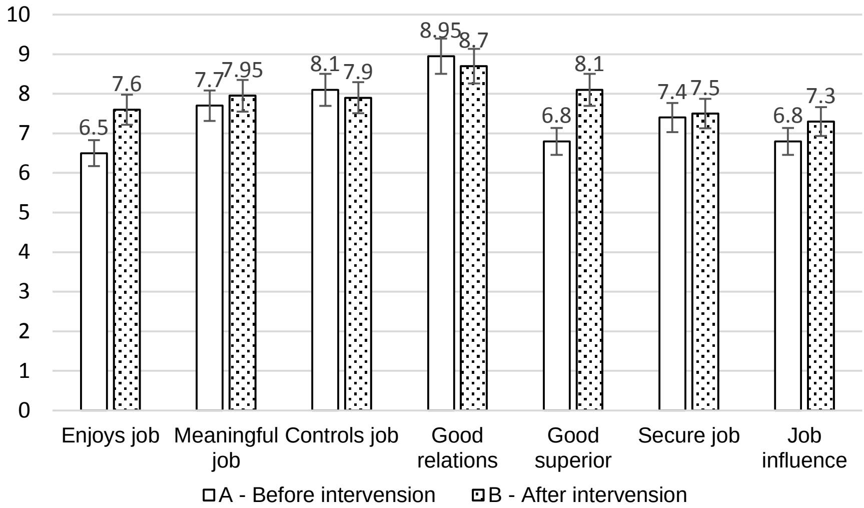


Figure 8. The responses to the Kiva questionnaire before and after the intervention (10-point scale)

The measurements of the cortisol level in saliva (table 5)

Table 5. The cortisol levels of office workers

Cortisol levels, nmol/l	*Office 1 M/L/E**	*Office 2 M/L/E*	*Office 3 M/L/E*
Before intervention	10.3/4.6/4.1	8.9/6.1/4.8	10.5/6.5/5.5
After intervention	10.2/7.6/7.0	9.5/7.0/6.0	10.2/6.5/5.5

M/L/E*- measurements of cortisol in M-morning/ L-lunch-time/E-evening

The changes in the cortisol levels in the three investigated offices are given in Figure 9.

The usefulness of the Metal Age programme is obvious, particularly in the office situated in a small town where the knowledge of ergonomics is poorer than in the capital (two other investigated offices).

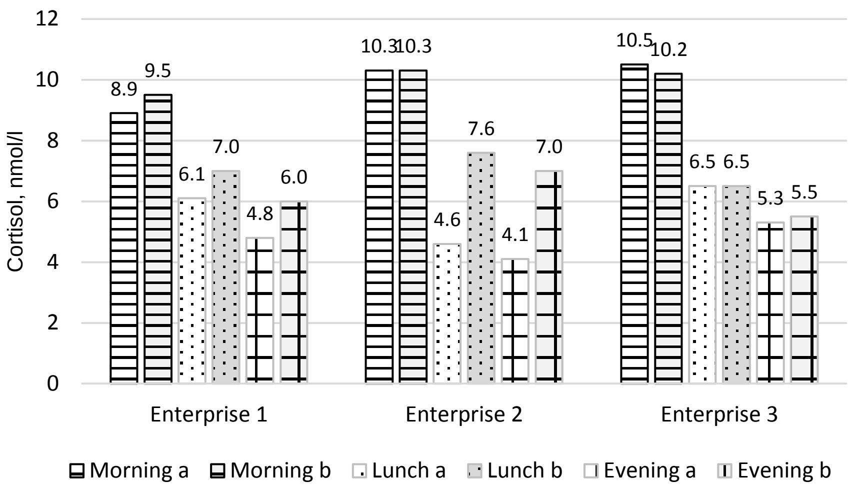


Figure 9. The change of the cortisol level in saliva before (a) and after (b) the intervention

Prevention and rehabilitation

The proposals were given for ergonomic improvements at workplaces (new ergonomic chairs, the possibility to change the height of the worktable; the change of the situation of a monitor etc.). The complaints in the air of the work environment and lighting deficiency complaints were reported to the employer. The rehabilitation of MSDs are possible using balneological methods of treatment and methods of physiotherapy (Tuulik et al., 2013; Visnola et al., 2010; Wargenaar et al., 2012).

Conclusions

The myometer allows the basic indicators of the skeletal muscle condition (stiffness and elasticity) to be determined. The data are valuable for the early diagnosis of possible health disturbances caused by work and for planning the rehabilitation treatment in an early stage of overload caused MSDs. MSD questionnaires, objective methods and environmental measurements are useful to plan prevention and early rehabilitation before the disability appears.

The indoor air and other problems in the same workroom could be defined individually in quite different ways. Therefore, an individual approach for every workplace has to be implemented considering the anthropological and other features of the worker who will work in the certain workplace. The info-technology workers often work in under-lighted working conditions although there is a possibility to raise the (artificial) lighting to the normal limits (400-500 lx).

The main conclusion from the investigation is that stress situations at workplace could be prevented by use of proper intervention programmes. It is most interesting (based on the study results for under 40 - year and over 40 – year workers) that there are different conditions for younger workers and for older workers or the same conditions are felt by the two groups of workers in different ways. In addition, the age and the sex of the employer and the employees have their influence on the questionnaire results.

References

Brauer, C. & Mikkelsen, S. (2010). The influence of individual and contextual psychosocial work factors on the perception of the indoor environment at work: a multilevel analysis. *International Archives of Occupational and Environmental Health, 83,* 639-651.

Eisenberg, N.I., Cole, S.W. (2012). Social neuroscience and health. Neuropsychological mechanisms linking social tied with physical health. *Nature Neuroscience*, *15*, 669-674.

Eltayeb, S., Staal, J.B., Kennes, J., Lamberts, H.G.P., & de.Bie, R. (2007). Prevalence of complaints of arm, neck and shoulder among computer office workers and psychometric evaluation of a risk factor questionnaire. *BMC Musculoskeletal Disorders*, *8*(68), 1-11.

Feuerstein, M., Nicholas, R.A., Huang, G.D., Dimberg, L., Ali, D., & Rogers, H. (2004). Job stress management and ergonomic intervention for work-related upper extremity symptoms. *Journal of Applied Ergonomics*, *35,* 565-574.

Hofman, L.F. (2001). Human saliva as a diagnostic specimen. *Journal of Nutrition, 131*, 1621-1625.

Kalman, B.A. & Grahn, R.E. (2004). Measuring salivary cortisol in the behavioural neuroscience laboratory. *Journal of undergraduate neuroscience education, 2,* A41-A49.

Kawasaki, T., Maeda, M., Tsuji, A. (1979). Determination of plasma and urinary cortisol by high-performance liquid chromatography using fluorescence derivatization with dansyl hydrazine. *Journal of Chromatography. Biomedical Applications, 163*, 143-150.

Kulin, J., & Reaston, MR. (2011). Musculoskeletal disorders early diagnosis: A retrospective study in the occupational medicine setting. *Journal of Occupational Medicine and Toxicology, 6,* 1-6.

Lahtinen, M., Huuhtanen, P., Kähkonen, E., & Reijula, K. (2002). Psychosocial dimensions of solving an indoor air. *Indoor Air*, *12,* 33-46.

Lindström, K., Elo, A.-L., Skogstad, A., Dallner, M., Gamberale, F., Hottinen, V., Knardahl, S., & Orhede, E. (2000). General Nordic Questionnaire for psychological and social factors at work. Copenhagen, Denmark: Nordic Council of Ministers.

Melamed, S., Ugarten, U., Shiron, A., Kabena, L., Lerman, Y., & Froom, P. (1999). Chronic burnout, somatic aerosol cortisol levels. *Journal of Psychosomatic Research, 46*, 591-598.

Miller, G., Chen, E., & Cole, S.W. (2009). Health psychology: developing biologically plausible models linking the social world and physical health. *Annual Review of Psychology, 60,* 501-524.

Näsman, O. (2011). Metal Age and Kiva-questionnaire. Assist in navigation towards well-being at work. Mediona OyAb. The Archipelago Academy for Well-being at work. Retrieved September, 14, 2013, from http://www.mediona.fi/pdf/KANSI%20Metal%20Age%20ja%20Kiva-kysely%20ENG.pdf

Oha, K., Viljasoo, V., & Merisalu, E. (2010). Prevalence of musculoskeletal disorders, assessment of parameters of muscle tone and health status among office workers. *Agronomy Research, 8*, 192-200.

Panari, C., Gugliemi, D., Ricci, A., Tabanelli, M.C., & Violante, F.S. (2012). Assessing and improving health in the workplace: an integration of subjective and objective measures with the Stress Assessment and Research Toolkit (St.A.R.T.) method. *Journal of Occupational Medicine and Toxicology*, *7*, 1-10.

Raja, A., Tuulik, V., Lossmann, E., & Meister, A. (1996). Neural network approach to classify the functional state CNS in case of neurotoxic diseases. *Medical and Biological Engineering and Computing*, *34*(suppl.1), 241-242.

Tint, P., Traumann, A., Pille, V., Tuulik-Leisi, V.-R., & Tuulik, V. (2012a). Computer users' health risks caused by the simultaneous influence of inadequate indoor climate and monotonous work. *Agronomy Research*, 10, 261-268.

Tint, P., Traumann, A. (2012b). Health risk assessment in atrium-type buildings. *International Journal of Energy and Environment*, *6*, 389-396.

Tuomi, K., Ilmarinen, J., Jahkola, A., Katajarinne, L., & Tulkki, A. (1998). Work Ability Index. Helsinki: Finnish Institute of Occupational Health.

Tuulik, V.-R., Tuulik, V., Pille, V., Tamm, M., Saarik, S., Vare, T., & Tint, P. (2013). Laser-doppler perfusion monitoring, myotonometry, and workplace risk evalution as assessment methods of musculoskeletal overuse syndromes in industry workers. *Journal of Rehabilitation Medicine*, *45*, 976-977.

Vain, A. & Kums, T. (2002). Criteria for preventing overtraining of the musculoskeletal system of gymnasts. *Biology of Sport*, *19*, 329-345.

Visnola, D., Sprudza, D., Bake, M.A., & Pike, A. (2010). Effects of art therapy on stress and anxiety of employees. *Proceedings of the Latvian Academy of Sciences, Section B, 64,* 85-91.

Wangenaar, A.F., Kompier, A.J., Houtman, I.L.D., van den Bossche, & S.N.J., Taris, T.W. (2012). Employment contracts and health selection. Unhealthy Employees out and healthy employees in? *Journal of Occupational and Environmental Medicine, 54,* 1192-1200.

Zakerian, S.A. & Subramaniam, I.D. (2009). The relationship between psychological work factors, work stress and computer-related musculoskeletal discomforts among computer users in Malaisia. *International Journal of Occupational Safety and Ergonomics* (JOSE), *15*, 425-434.

Stress and strain in mobile IT-supported work: an empirical study in the area of mobile services involving smartphones, notebooks and similar devices

Michael Bretschneider-Hagemes
Institute for Occupational Safety and Health (IFA of the DGUV)
Sankt Augustin
Germany

Abstract

Design and organizational issues and problematic aspects of mobile IT-supported work performed by *service technicians* were examined by a quantitative study of the associated stresses and strains. The study differentiated between stress dimensions of a technical and organizational nature on the one hand, and their consequences, i.e. strain, on the other. This differentiation in turn enabled conclusions to be drawn for specific recommendations for occupational practice. Reductions in impairing stress and impairing strain were to be achieved by selective prevention measures and improvements to the technical equipment used by the workers. Specific measures derived for implementation in occupational practice are outlined.

Introduction

The world of work is undergoing a sustained process of digitalization. The mobile use of digital systems in the form of mobile IT-supported work has numerous consequences, not all of which are immediately apparent. Beginning with the breakdown of established forms of work organization and the decreasing scope for individuals to identify with the contexts of traditional employment (cf. Bretschneider-Hagemes & Kohn, 2010), through to safety risks caused by the on-board use of mobile information and communications technology (ICT) in vehicles, numerous issues of major relevance to occupational safety and health can be identified. For some years, the Institute for Occupational Safety and Health (IFA) of the German Social Accident Insurance (DGUV) has conducted research into such work systems, which can be grouped under the heading of mobile IT-supported work. Publications and firm recommendations for various work scenarios with the focus upon safe and humane work have been produced as a result.

In order to counteract the existing dearth of quantitative data on the stress situation associated with mobile IT-supported work and to provide a basis for preventive activity, a theoretical standard paradigm of the labour sciences (stresses and strains, based upon EN ISO 10075-1, cf. also Rohmert & Rutenfranz, 1975 and Cox et al., 2000) was operationalized in a dedicated approach within the current research activity, and used as the basis for an empirical study. The groups studied comprised

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

mobile service technicians, responsible for example for maintaining telecommunications networks. Working over a wide geographical area, these technicians make intensive use of the most diverse mobile ICT equipment in order to complete and document maintenance tasks. Equipment typically includes netbooks and notebooks, smartphones, computerized instruments, satnavs, etc. Rather than being used sporadically as an aid as was observed in the early days of mobile ICT, the equipment is now used systematically; its users are now dependent upon it and its trouble-free operation. The suspected sources of stress were studied with regard to their relevance by the measurement of stresses (the total objective influences upon the individual) and strains (individual impact, moderated by personal resources), and the statistical relationships between the two. For this purpose, a broad discussion was conducted in which hypotheses were gathered, supported by field observations and assigned to potential stress dimensions. The aim was to identify which stress situations should be optimized, be they of a technical or organizational nature. In the interests of concision, the description below discusses only selected key aspects.

Operationalization

Operationalization involved the distinction between stresses on the one hand and strains on the other. The terms “stress” and “strain” as such imply no value judgement; the adverse forms of stress and strain are termed “impairing stress” and “impairing strain”. In order for the problematic forms of stress and strain to be identified, the analysis was structured by the creation of generic dimensions on the stress side (see Figure 1).

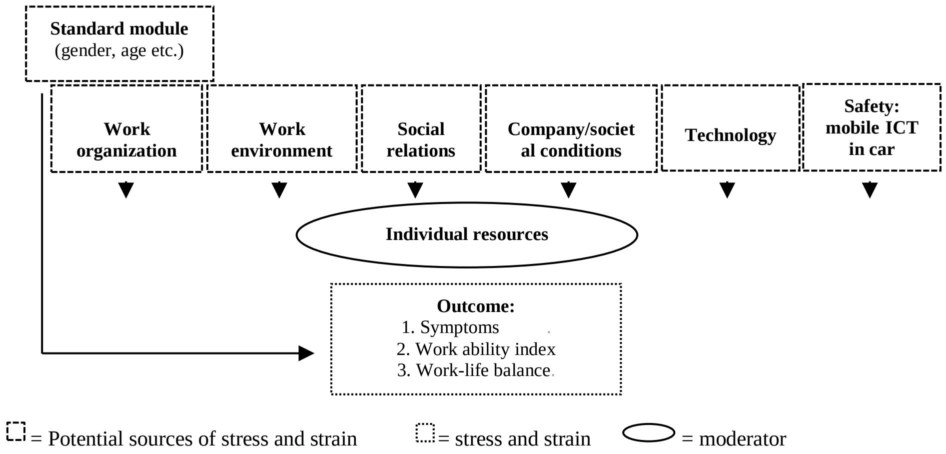


Figure 1. Research design/operationalization

The dimensions constructed in this way are specifically those of work organization, work environment, social relationships, company and social conditions, and technology. In many situations, a complex interaction was observed between the use of mobile IT and diverse objective factors influencing the work system. The research design is able only to operationalize this objectivity as adequately as possible and to study it with regard to its stress factors associated with the use of mobile IT.

All dimensions constructed in this context and populated with items were reviewed for their internal consistency and legitimized by Cronbach's α. In addition, the strain level was studied by means of three survey instruments employed in parallel (see Figure 1). These were the work ability index (2nd dimension: test subjects' own estimation of their work ability in relation to the demands of the work; cf. Hasselhorn & Freude, 2007); a proven index of symptoms (cf. Ertel et al., 1991) used to measure the state of health; and certain work-life balance items. This procedure enabled the effect of specific work conditions and work equipment to be compared.

Results

The study was based upon a random sample (N=233, returned questionnaires = 100%) of service technicians performing relatively homogeneous tasks. Mobile IT was used for the purposes of documentation, communication and navigation, besides the technical work itself. The data were obtained from a questionnaire completed by the test subjects themselves following an introductory talk at the workplace (team meeting). As expected, the composition by gender was very one-sided, with male workers accounting for 99%. The average age was 48.5 years. The individuals questioned had already been working with mobile IT for an average of ten years and with an average usage intensity of 3.6 hours per day. Up to 74.8% (with strong variations according to region and technical equipment) considered their work situation to have deteriorated on balance by the use of mobile IT. In this context however, the charge was disproved that older employees are less well disposed to the use of mobile IT. The variables did not indicate any such correlation. By contrast, poorer estimations regarding the potential for stress and impairing strain were invariably found among persons with a negative fundamental view of the use of mobile IT. For example, negative views both of the work organization and of the work-life balance correlated with the negative fundamental view (statement: The situation has improved/deteriorated as a result of mobile IT). This does not however mean that the former is the cause of the latter.

Such negative attitudes are not without justification. On the whole, initial motivation for the use of new technology disappears when technical failures begin to occur and the personal tolerance of frustration is exceeded (cf. Hoppe, 2010). In view of this, the introductory phase of new technology in particular can be considered decisive. A comprehensive pilot phase and granting the users a say in the planning phase therefore pay off in the long term. During observation of the interaction between generic items that do not correspond to any constructed dimension in the area of stresses (standard module) on the one hand and strains on the other, a notable observation was that neither the length of use (in years) nor the daily duration of use of mobile IT had a statistically significant effect upon the strain. The pointed emphasis that the longer IT-supported work equipment is used, the more stressed and more ill the users are is in the first instance shown to be a prejudice.

Health, WAI and work-life balance

Health: The health of the individuals surveyed was described by the recording of scales of symptoms. The symptoms were selected based upon a reference study conducted among service technicians (cf. Ertel et al., 1991). Musculoskeletal complaints were those most frequently stated (see Table 1). Approximately 76% of the individuals surveyed (averaged across each group of technicians) indicated that they suffered from back complaints; 69% indicated shoulder/neck complaints. These were followed by typical symptoms of mental strain, for which the values were also high: 64% suffered from inner restlessness and tension; fatigue/weariness was a factor for 55%; 48% complained of concentration disorders; and 43% experienced increased irritability. The levels of these conditions were generally more severe than those suffered by the reference group, specifically with regard to physical/ergonomic area aspects. In addition to work conditions that may have been less favourable – the survey of the reference group was conducted at the beginning of the 1990s, without the influence of ICT – one reason may be the relatively high average age of the random sample. The trend towards musculoskeletal disorders is at any rate confirmed clearly here and points to ergonomic deficiencies in the work equipment and workplaces of the mobile workers.

Table 11. Selection of symptoms, comparison

		Reference	**Random sample**
Inner restlessness	%	73%	64%
Back pain	%	64%	**76%**
Concentration disorders	%	64%	48%
Shoulder/neck pain	%	55%	**69%**
Irritability	%	55%	43%
Fatigue/weariness	%	55%	55%
Changes in eyesight	%	50%	**60%**
Headaches	%	27%	**39%**
Joint pain	%	14%	**34%**

Work-ability index (WAI): The work-ability index was recorded in its second dimension, the work ability in relation to the demands of the work. Work ability refers here to separately surveyed mental and physical criteria in relation to the demands of work, and is regarded as a reliable indicator that deviates only rarely from third-party estimations by experts (cf. Ilmarinen 2003, p. 94). Estimations of the physical work ability differed: a good 44% considered their physical work ability to be mediocre or more bad than good, 56% considered it good to very good.

Table 2. WAI values, comparison

Mean values of the WAI reference database (University of Wuppertal) and mean value of the random sample			
	Mean value	SD	N
Reference	7.44	1.501	8,330
Random sample	**6.89**	**1.562**	**209**

Around 56% rated their mental work ability as mediocre or more bad than good. A weighted total was calculated from the values for the mental/physical work ability according to task profiles. This total can be compared with the values from comprehensive reference data (see Table 2). If the respective mean is taken as the reference, the value attained for the estimated work ability is seen broadly on average to trend slightly negatively (random sample = 6.89, reference = 7.44 with a best possible value of 10 and a possible negative value of 2). In view of the clearly negative indications of health symptoms, the estimations regarding the WAI appear somewhat optimistic. The reactivity in response to the survey should be considered during interpretation of the data: possibly being unable to satisfy demands may give rise to negative feelings and fear of repercussions.

Work-life balance: The total value for the work-life balance was operationalized by the adjacent non-vocational areas of family, partnership and other leisure aspects. 54% of the workers stated that the mobile IT-supported work caused significant to high perceived stress in the non-vocational areas of their lives. Availability at all times as a consequence of mobile ICT was repeatedly raised as a major problem in both the questionnaire responses and interview situations. Many employees had difficulty separating working and vocational life. Some employees felt that they were being checked on by the use of mobile ICT, particularly by means of their mobile telephones. In addition, some individuals complained of being controlled by always being expected to respond immediately to telephone calls and e-mails. Criticism was also levelled in this context at the scheduling practices of some employers. The affected employees would like to have a greater say in this area. The industrial partners were advised to clarify issues of reachability and availability in company agreements. Procedures for the issuing of appointments and for planning journeys were also criticized, particularly with regard to their compatibility with family life.

Impairing stress and its relationship to impairing strain

The most important results for the dimensions created in the area of stresses (see Figure 1) are described in condensed form below, and their relationships to the defined strain values outlined.

Work organization: The dimension of work organization fared negatively in the estimation of the survey participants (see Figure 2). 77.4% considered themselves exposed to significant to high impairing stresses. The variable with the strongest negative influence upon this total value was the tightness of the schedule. 75.8% of those surveyed indicated the stress value here to be significant to very high. The individuals surveyed would appreciate a greater say in scheduling. The variable of multitasking is also conspicuous for its substantial stress values, which 71.3% of those questioned state to be in the range from significant to very high. One observation was continual interruptions to the primary maintenance tasks by calls by dispatchers and superiors.

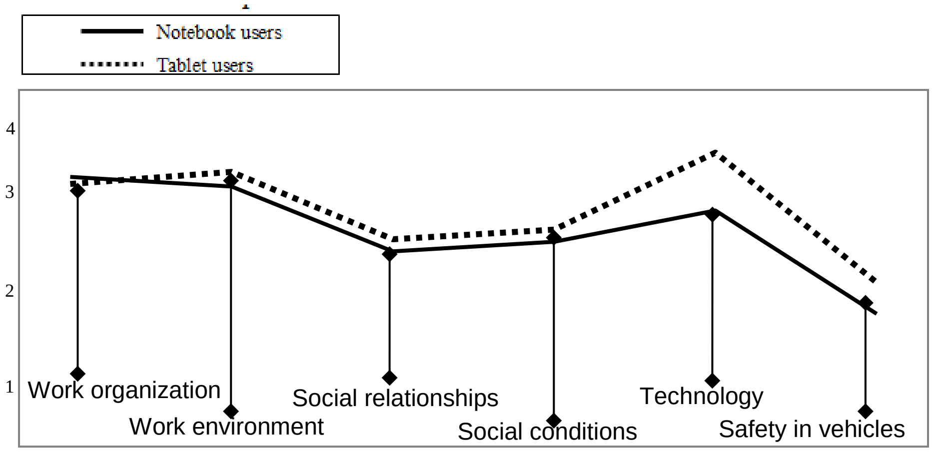


Figure 2. Region of the stresses, overview (clustered estimations of the subjects – 1 = no impairing stress; 2 = impairing stress; 3 = significant impairing stress; 4 = high impairing stress – groups differentiated according to work equipment of notebook and tablet PC)

A further complaint voiced in interviews was of typical work situations that were often performed under adverse conditions. These included working in dark and dirty and in confined spaces. Documentation work was frequently performed without desks or similar. Work with mobile IT was then carried out with no work surface whatsoever, in the vehicle or on unsuitable work surfaces.

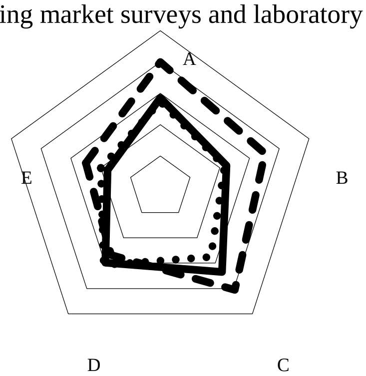
The most significant effect ($p < .05$) is that of the total value of the work organization upon the work-life balance strain (see Figure 4). The correlation here is moderately strong ($r = .400$). The strain elements of health ($r = .273$) and WAI2 ($r = .286$) also correlate significantly. It was thus demonstrated that unfavourable circumstances in the organization of mobile IT-supported work are closely related to impairing strain upon the individual (particularly that associated with the work-life balance).

Work environment: Among those surveyed, the work environment had not changed directly as a result of mobile IT. The demands resulting from the combination of use of the equipment and the environment had changed, however. Those affected may for example be responsible for the condition in which their equipment is kept. The equipment is however regularly used under adverse weather conditions. In the cases observed, the equipment was not adequately prepared for such use, even though the market provides a more than adequate range of products for this purpose. In these and similar cases, it is notable that equipment is frequently purchased and used without consideration or adequate acknowledgement of the actual working conditions. The work environment in general was rated by 77.4% of those surveyed as being significantly to very stressful. The strongest negative influence upon this total value was the variable of stresses arising from unfavourable work locations. 75.9% of those surveyed stated a significant to very high stress value in this case. The street, construction site and the vehicle were stated as examples.

The variable of creation of a new mobile office was conspicuously negative, with a value of around 70% (significant to very high stress). Complaints were voiced regarding the problems entailed by having to improvise a workplace under continually changing conditions. The clearest effect was that of the total value for the work environment upon the perceived work-life balance (see Table 2). A significantly ($p < 0.05$) positive relationship was observed here ($r = .351$). The strains of work-life balance and WAI2 also correlated, with weak significance. It was thus demonstrated that unfavourable circumstances in the area of the work environment of mobile IT-supported work are closely related to impairing strain upon the individual.

Social relationships: Reference studies, including studies in the area of home-based teleworking (cf. Kleemann & Voß, 1999), have repeatedly drawn attention to the issue of social isolation. The associated discussion tends at times to view work performed in isolation from the outset as negative. Interestingly, this depends however very strongly upon the attribution by the affected individuals, i.e. upon their personal resources. The preferences of individuals, i.e. whether they prefer to work alone or not, must be taken at face value. Mobile IT-supported work doubtless leads to formerly stationary forms of work now being confronted with this fact. The attribution of negative consequences solely to the use of mobile IT and the drawing of conclusions on a purely technical level would not however appear to be geared towards resources and solutions. The results of the present random survey further indicate that the enhancing of personal resources by targeted programmes for the development of skills and accompanied by structural provision for regular networking between the mobile workers could be more constructive in terms of reducing strain. Social relationships at the workplace were classified by the participants in this survey broadly on average, i.e. by 46.8% of those questioned, as being associated with significant to very high impairing stress. Up to 54% of those questioned (depending upon the sub-group) however associated the variable of isolation with significant to very high stress. The term loneliness appeared repeatedly in isolated free-text responses. Employers must also appreciate that much of the information communicated implicitly within an organization (unwritten rules) may pass mobile, isolated workers by. Informal chats are in fact frequently underestimated as a functional medium within companies. The clearest effect is that of the total value of social relationships upon the work-life balance and the workers' own estimation of their mental and physical work ability, WAI2 (see Figure 3). In this area, a weak but significant ($p < 0.05$) relationship is observed ($r = .335$ and $.405$). The related issue of scope for staff participation had already been raised by Klemens in relation to the resulting burnout risk and recognized as a considerable influencing factor (Klemens et al., 2003). It was therefore demonstrated that the quality of social relationships in mobile IT-supported work exhibits relationships with impairing strain upon the individual.

Technology: The stress dimension of technology can be interpreted as the origin of numerous forms of strain. It is important however to consider that many forms of strain attributed prematurely to technology become relevant only through the relationship in particular situations with the specific work organization and work environment. Almost the full range of assumed technical deficits were encountered. Input devices were often unsuitable, since they exhibited ergonomic deficits. Displays exhibited glare, were highly reflective, and often could not be adjusted to the lighting conditions. Data transmission was frequently unreliable, or no network access available. In addition, software applications were often not suitable for the practical work procedures. The source of strain of technology is notable for having comparatively clear negative aspects (with values for tablet users being even worse than the average value stated). 74.9% reported significant to very high impairing stress. The variable of software ergonomics is shown to have the clearest negative influence upon the total value. The questionnaire focussed upon the criteria of usability. The adaptation of the user interfaces to the small screens of mobile devices can be regarded here in particular as unsuccessful. The variable of technical failure is conspicuous, receiving a negative rating of 70.5%. The notebooks and tablets used, i.e. the hardware side of mobile IT, were not in any way equipped for use under adverse conditions, but were merely conventional business products. Total failure was correspondingly frequent, causing interruptions in the workers' productivity, to their disadvantage (time pressure). It was also demonstrated that where the technical equipment of mobile IT-supported work exhibited deficits, the result was impairing strain upon the individual. Significant correlations were seen with all surveyed forms of strain. Specific products for improving the situation were recommended following market surveys and laboratory tests.



	Work-life balance	Health	WAI2
A: Work organization	40.00%	27.30%	28.60%
B: Work environment	35.10%	21.60%	22.30%
C: Social relationships	40.50%	27.50%	33.50%
D: Company/Health condition	26.80%	30.70%	29.80%
E: Technology	25.00%	20.00%	17.70%

Figure 3. Relationships between stresses and strains, overview (r = 1 is treated as 100%)

Discussion

The present study represents a quantitative reference points in the study of mobile IT-supported work. An established standard paradigm of the labour sciences (the stress/strain model) was operationalized and applied in combination with observations in the field and interviews (method triangulation). Results were obtained for the specific fields concerned. The service technicians proved to be an occupational group yielding useful study results. The assumption of high mobility among these employees was confirmed, as was the diverse and intensive use of mobile IT. The interpretations and observations enabled a substantial need for improvement to be identified within all the sources of strain (stresses). The key factors leading to impairing strain were shown to be not, for example, greater age or the daily duration of use, but the way in which mobile IT-supported work is organized, the form taken by the work environment of mobile workers, their social relationships, and the quality and usability of the mobile IT (hardware and software). The most important individual variables were presented and related to the field. In view of the particularly critical strain source of technology, a surprisingly high relevance must be accorded to the work equipment. Product tests in the IFA's laboratories enabled recommendations to be made for ergonomic and usable equipment. The high luminance and the reflection properties of the screens in particular led to product recommendations in the area of rugged computing.

The statistical studies enabled the assessments of those surveyed to be related to the findings from the field studies. At the same time, it became possible to identify probable consequences of impairing stress in the form of impairing strain (outcome), and to compare these with reference groups, at least to some extent. A useful body of data is thus available for the pending prevention work and can be used to launch the necessary changes purposefully and selectively.

References

Bretschneider-Hagemes & M., Bökmann, M. (2010). Intervention in komplexe Systeme - Eine Systemtheoretische Betrachtung im Umfeld von Unternehmensberatung und Coaching, Carl-Auer Verlag, Heidelberg.

Bretschneider-Hagemes, M. & Kohn, M. (2010). Ganzheitlicher Arbeitsschutz bei mobiler IT-gestützter Arbeit. In C. Brandt (2010) *PaPsD – Arbeitsschutz und Arbeitsgestaltung von mobiler Arbeit* (pp. 33-52), ver.di, Berlin.

Cox, T., Griffiths, A. & Rial González, E. (2000). Research on work-related stress, Office for Official Publications of the European Communities, Luxembourg.

DGUV (2009). BGI/GUV-I 8696. Information - Einsatz von bordeigenen Kommunikations- und Informationssystemen mit Bildschirmen an Fahrerarbeitsplätzen, Deutsche Gesetzliche Unfallversicherung (DGUV), Berlin.

European Standard (2000). EN ISO 10075-1 – Ergonomic principles related to mental work load – part 1: General terms and definitions - German version.

Ertel, M., Wilkening, W. & Zwingenberger, J. (1991). Die Zeit ist weggelaufen - Servicetechniker in der Computerindustrie, DAG Ressort Bildungspolitik, Hamburg.

Hasselhorn, H.M. & Freude, G. (2007). Der Work Ability Index – ein Leitfaden, baua, Dortmund.

Hoppe, A. (2010). Komplexe Technik – Hilfe oder Risiko? Darstellung ausgewählter Ergebnisse einer Grundlagenuntersuchung zu Technikstress, BTU Cottbus.

Ilmarinen, J. & Tempel, J. (2003). Erhaltung, Förderung und Entwicklung der Arbeitsfähigkeit – Konzepte und Forschungsergebnisse aus Finnland. In B. Badura, H. Schellschmidt and C. Vetter (2003) *Fehlzeitenreport 2002*, Springer-Verlag, Berlin/Heidelberg.

Kleemann, F. & Voß, G.G. (1999). Telearbeit und alltägliche Lebensführung, http://www.tu-chemnitz.de/phil/soziologie/voss/download.htm, 5 January 2000, p.7.

Klemens, R., Wieland, R. & Timm, E. (2004). Was schützt vor Burnout in IT-Berufen? Belastungen und Ressourcen in Unternehmen der IT-Branche. In W. Bungard, B. Koop and C. Liebig, *Psychologie und Wirtschaft leben. Aktuelle Themen der Wirtschaftspsychologie in Forschung und Praxis* (pp. 64-71), Rainer Hampp Verlag, München.

Rohmert, W. & Rutenfranz, J. (1975). Arbeitswissenschaftliche Beurteilung der Belastungen und Beanspruchungen an unterschiedlichen industriellen Arbeitsplätzen, Bundesminister für Arbeit- und Sozialordnung. Referat Öffentlichkeitsarbeit, Bonn.

Can Immersive Virtual Environments make the difference in training industrial operators?

Salman Nazir[1], Annette Kluge[2], & Davide Manca[1]
[1]PSE-Lab, DCMIC "Giulio Natta", Politecnico di Milano, Italy
[2]University of Duisburg-Essen, Duisburg, Germany

Abstract

The role of the operator's performance for continuous, hazard-free and economically viable production in the chemical industry is of paramount importance. Even a minor error may result in a near miss, an abnormal situation, and even a catastrophic accident. This paper analyses and discusses the role played by distinct training methods performed in an Immersive Virtual Environment (IVE) to form and influence the performance of operators. An experiment was designed and conducted for a specific procedure of a polymerization plant, i.e. catalyst-injector switch, where the operator has to follow a sequence of actions. Two groups of participants (N = 24) were trained within the IVE according to different training methods. The Immersive Observers (IO) group observed a trainer explaining and performing the procedure, while the Immersive Actors (IA) group performed singularly and personally the procedure through a so-called guided-tour. The IO group showed higher accuracy, precision, process understanding, identification skill and lower help requirements than those of IA group, conversely IA group showed better speed when compared with the former. The results lead to the conclusion that IVE are not effective per se and need guidance by a competent trainer when the task is unfamiliar. The impact for future research and practice are discussed.

Introduction

Industry background

The contribution of chemical industry is enormous in terms of the quality of life that was achieved in recent past, the availability of goods, and reduction in costs of various necessary products, and optimization of energy consumption and production. Only in Europe, in 2011, the production of chemicals generated euro 539 billion in revenues. The chemical industry employed directly 1.1 million people working in a large number of companies (CEFEC, 2012). At the same time, each year millions of euros are lost in this sector because of accidents, unscheduled shutdowns or abnormal situations, for instance the notorious Deepwater Horizon oil spill caused an estimated economic loss of USD 90 billion. In terms of human lives, according to the International Labor Organization (2003), 2 million people are killed by their work every year, which means ca 5,500 people every day, ca 230 every hour and nearly 4 persons every single minute.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Most of the industrial accidents are triggered by some abnormal situation. An abnormal situation is referred to as a situation where the operating conditions deviate from the optimal range, which can result in the introduction of uncertainties to the system leading to forced shut-down of the plant or generating a serious accident scenario. Chemical processes run smoothly if operated within a safe region as shown in Figure 1. If the process falls outside that region, an abnormal situation occurs. If the abnormal situation is not managed properly, it can result in an accident.

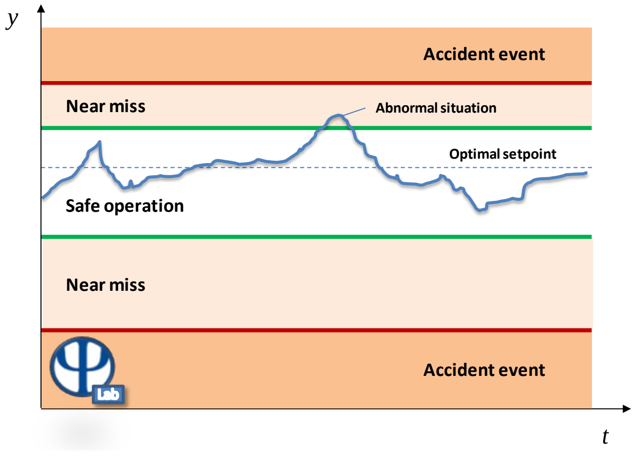


Figure 1. Qualitative trend of a chemical process variable, y, as a function of time, t. Since the process is intrinsically non-linear then the horizontal bands are usually not symmetric (PSE Lab, Politecnico di Milano).

Influence of training on safety

Most of the abnormal situations, which arose in a process plant, are less due to the failure of equipment, design or system and are more due to inappropriate reaction or understanding of the operator who transforms a possible little problem into a serious accident (Reinach & Viale, 2006). There are different stages before a major accident arises. Figure 2 shows the pyramid depicting different types of situations and incidents before the major accident takes place. Most of the situations including unsafe behaviours, near misses and minor incidents can be avoided if proper actions are taken in a timely and consistent way. If the situation can be controlled during the first three stages of the pyramid, then major incidents might be avoided.

In a recent study by Antonovsky et al. (2013) 45 petroleum industries were surveyed to understand the elements responsible for accidents using HFIT (Human Factor Investigation Tool). It was found that the most common reason among all failures was the wrong interpretation or assumption made by the operator. Therefore, the understanding, accuracy and errorless actions of operators are key elements in determining the safety and the continuous production from industrial plants. Various training methods are used in the industry with the aim of improving both operators' understanding and performance.

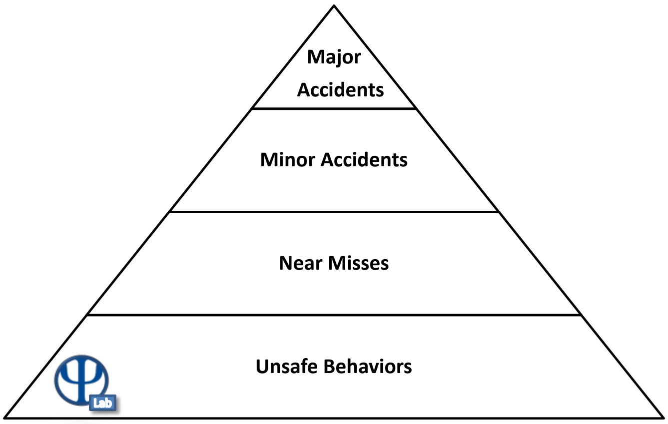


Figure 2. Safety pyramid reflecting different stages before reaching the major incident (PSE Lab, Politecnico di Milano).

Operator training

An industrial operator is the person who works in the plant. Depending on his/her location and nature of job, s/he works either as Field Operator (FOP, working in field) or as Control Room operator (CROP, working in control room).

Although there is not a specific operators' training method, nonetheless the concept of industrial training exists since decades (Kluge et al., 2009). Contrary to training methods for aviation industry, where homogeneity and consistency can be observed independent of the geographical and cultural differences, the process industry training methods face diversity and variance depending on country, nation, culture and stakeholder's interest. Unfortunately, the value and impact of training has not been fully understood and appreciated in case of process industry. This study does not focus on the importance of training, which is already extensively covered by Nazir et al. (2012) , Manca et al. (2013), Salas et al. (2013), and Kluge (in press).

The Authors have been working on advancement in training methods for industrial operators in order to improve their performance, reliability and ability to handle normal situation, abnormal situation and rare situations like start up or shut downs (Kluge, in press, Kluge et al., 2009, Manca et al., 2013, Nazir et al., 2013). Start-ups and shut downs are not the only rare events which take place in chemical plants, but other events like regeneration of some specific components at a periodic frequency are common events in industrial processes. The probability of running into an error in these rare events is higher as there is always a rather long time interval (usually and at least from 1 to 2 months) between consecutive events.

It has been shown that skilled performance can be directly co-related to quality and quantity of training (Kluge (in press), Proctor & Dutta 1995, Colquitt et al., 2000) and expertise (Ericsson et al., 1993). Burkolter et al. (2009) and Nazir et al. (2013) discussed the possible improvements in process control tasks attainable by different training methods. Most of the conventional tasks are performed in the field and not in the control room, where automation, graphical data, numerical values are explicitly displayed on the screens of the DCS (i.e. Distributed Control System) solution. Simple skills like identification, attention management, following a sequence of actions in a precise and timely manner are sufficient to accomplish such tasks successfully.

Training of CROPs is considered essential and vital in the world of process industry. Conversely, FOPs have been mostly neglected by the training programs of the Industry. As extensively discussed by the authors of this manuscript, the training tool for FOPs should be immersive and capable of reproducing the spatial layout of the plant. Manca et al. (2013) coined the term Plant Simulator (PS) to describe an advanced training tool that is centred on the trainee and addresses the most relevant human factors issues related to his/her formation and personal involvement. In less technical words, the PS is a training simulator using IVE augmented by computational software and algorithms to assess the performance in a systematic manner. As reported in Nazir et al. (2013), the understanding of the plant geometry and most relevant process operations was significantly improved with respect to the conventional methodologies adopted to train FOPs. The benefits and usefulness of IVE implemented in the PS training solution was tested and shown by Nazir et al. (2012) and Colombo et al. (2013). Those experiments simulated an accident event and compared two different training methods, i.e. PowerPoint presentation versus PS (IVE). Conversely, the present study emphasizes on deepening the use of IVE by analysing two training methods within the PS solution for normal operating conditions. In the study, a replica of catalyst injector switch was developed in the PS by following the real process parameters, geometry and constraints. Moreover, an algorithm for assessment of performance was developed and integrated.

The purpose of these experiments based on alternative training methods is reaching the best possible training practice for industrial operators so to mitigate the losses (in terms of finance and human lives) in the process industry. This paper represents a step towards filling the gap in the development of universally viable and consistent training methods. Specifically, two groups of trainees are trained by means of distinct methodologies inside the same training environment. The first group is trained in an IVE by an expert trainer. These trainees observe the specific procedure for catalyst injector switch (discussed later) and represent the immersive observers (IO) group. The second group is trained by directly performing the actions within an automatic guided tour. This is the immersive actors (IA) group. The performance assessment on both groups of trainees allowed extracting the following bits of information:

1) Observing the trainer performing the actions in an IVE improves the performance in terms of accuracy, number of helps required and precision.
2) Performing directly the actions in an IVE (without the previous example of the trainer) improves response speed at the cost of accuracy.

Methodology

Participants

A total of 24 university students (12 Males; age 22-24 year; mean 20.8 year; SD

0.94 year) participated to the experiment. The participants were from the third year of the bachelor degree in Chemical Engineering at Politecnico di Milano. They applied to an open and free call. No economic reward was provided. In advance, the participants were asked about any previous experience on similar experiments and their interest in 3D gaming to increase a priori the groups formation and equalize and balance the experiment. The experiment was performed in accordance to the ethical standards laid down in the 1991 Declaration of Helsinki. All participants signed a consensus form prior to their participation to the experiment and were informed of the freedom of leaving the experiment without any prior notice. Participants were randomly divided into two groups of 12 people each, equalized for gender and previous experience (if any). Each participant was assigned an ID from 1 to 24. The immersive observers (IO) group was given the IDs from 1 to 12 while immersive actors (IA) group received IDs from 13 to 24. All the participants were amateurs, as they had not any experience of working in industrial plants.

Experimental setup

The catalyst injector switch

The training task consists in switching the operation of two catalyst injectors. This is a periodic but not routine task. The Spheripol process, where the catalyst-injector switch occurs, is the most widely used process for the production of polypropylene (PP), with a total production of almost 80 million metric tons per year (Spaniol et al., 2008). The necessary steps of switching the catalyst at the plant are performed periodically every 2 months. The technical details of this process and the catalyst switch mechanism are out of scope of this paper and can be found in the literature (Urdampilleta et al., 2006, Shi, et al., 2010). The procedure for catalyst injector switch involves a total of 29 actions. A successful catalyst switch can be ensured only by following accurately such a sequence. Any errors in following the sequence can result in triggering an abnormal situation that may lead to a significant impact on the process operation, thereby, resulting in huge economic losses. Failure of catalyst switch is enough to bring losses of millions of euros as it requires complex maintenance activities based on the shutdown of the plant (and following start up when the malfunction is fixed) with consequent production losses and need for production rescheduling. Therefore, other than analysing the effect of training methods in IVEs on the performance of operators, this study focuses the attention on the effectiveness of training procedures for catalyst-injectors switch in the production of PP.

During the development of the 3D model for catalyst-injectors switch, the details of the real process, in terms of geometry, process dynamics, physical appearance, valves locations and functions, and piping, were considered (see Figure 3).

Figure 3. A subsection of 3D catalyst injector switch model (courtesy of Virthualis company).

Training

Pre – training

The participant (playing the role of FOP) experiences the PS as a 3D immersive environment (facilitated by 3D passive glasses) in a dark room with spatial sounds. The pre-training session is focused on the use of nun-chuck to all participants at the same time (see Figure 4a and 4b). Nun-chuck is the means of interacting with the IVE for current experimental setup. It allows moving around the process section where the catalyst injectors switch occurs.

Figure 4b Training for the use of nun-chuck.

Figure 4a. Nun-chuck used to interact with IVE.

Both groups were pre-trained for use of the nun-chuck. At this stage, participants had no prior knowledge about the function or significance of any valves and their location.

Training

In order to keep the time interval between training and performance assessment consistent, the IO group was trained and assessed in the morning and the IA group in afternoon of the same day. The pre-training session lasted the same for both groups.

IO the immersive observers group

The IO group was given a lecture using slides presentation, which consist of background information regarding the process, its significance. The sequence of the first 9 out of 29 actions (see above), which the participants will be performing, was also revealed in the last slide of the presentation (see Figure 5a). Both the training and assessment sessions limited the sequence to 9 out of 29 actions based on the available training time with relation to the capacity of working memory in the range of 7±2 chunks (Miller, 1956). At the end of the slides presentation, the IO group went to the 3D room (where the PS is installed). A trainer performed the 9 actions of the training procedure by means of the PS and the participants observed his actions. While performing the catalyst-injectors switch procedure, the trainer explained briefly the meaning of each action to disclose the not overt aspects of the task. The actions were performed at a slower pace than normal. Actually, the trainer took about 20 min instead of the usual 6 to 8 min in real operating conditions, so that participants could understand better the sequence and its meaning. All the participants were completely involved and immersed in observing and understanding the training session by wearing 3D glasses (see Figure 5b). The following step consisted in assessing the participants when performing the 9 actions individually. It was made sure that the discussions among participants, regarding the sequence and experiment, were avoided during the waiting time. Moreover, once a participant finalized the assessment procedure, s/he was suggested to leave the premises of the experiment.

Figure 5a. First part of training session for IO group using the slides presentation.

Figure 5b. Second part of training session for IO group. The trainer (yellow helmet) performs the set of nine actions.

IA the immersive actors group

In case of the IA group, the participants got a flavor of the experiment nature by means of a short oral introduction (without slides) that lasted 4-5 minutes and was followed by the training on how to use the nun-chuck (see above). Later, each participant was taken to the 3D room (PS) for the individual training session without the trainer support. The automated guided tour took the participant through the 9 actions by following a yellow coloured visual cue (Figure 6). Once a step was performed, e.g. the valve was either opened or closed, the yellow border surrounding the specific piece of equipment changed into green to produce a positive feedback about the action correctness. After training half of the participants, they were called for the performance assessment session individually so to keep some consistency with the IO group as far as the time interval between training and assessment is concerned. The same pattern was used also for the remaining half of the participants of the IA group to keep the time interval between training and assessment consistent.

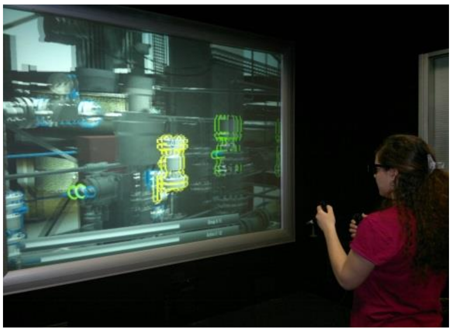


Figure 6. A Participant of the IA group involved in the training session by the fully automated guided-tour procedure.

The total time of training for the IO group was 40 minutes (15 minutes for slide presentation and 25 minutes for immersive training by the trainer), whereas each participant of the IA group took about 30 minutes for training.

Additionally, participants of both groups were informed about the procedure about how to request help. The help support was designed to avoid distraction during the performance assessment session. The help could be invoked by just raising one hand, which was read by a skeletal recognition algorithm based on the input from the Kinect device of Microsoft. Once the participant asked for help (during the assessment procedure) the following device to be operated in the actions sequence was highlighted in yellow. By doing so, the participant was able to continue the performance session without being stuck in the IVE.

Performance assessment

The operator/participant performance assessment is based on a hierarchical and computerized methodology that allows evaluating the effectiveness of a specific training method (Manca et al., 2012). Each action performed by the trainee during the assessment procedure is recorded and analysed. The assessment procedure is implemented in a computer program capable of evaluating not only the marks about the performance of operators but also registering, storing, and analysing the actions and decisions taken by the operator during the experiment with the help of a dedicated algorithm that runs in real time (Manca et al., 2012; Nazir et al., 2012).

With reference to the catalyst injectors switch procedure, the performance assessment is based on the following dependent variables, which are named as Operator Performance Indicators (OPIs) (see also Figure 7):

Overall mark: The total mark obtained by the trainee. It is evaluated by a dedicated algorithm and is based on the number and nature of errors conducted and the number of helps requested. In the experiment (based on the first nine actions out of the 29 total ones), the maximum obtainable (i.e. theoretical) overall mark was 30.04.

Number of helps requested: If and when the help(s) was requested by participants during the course of experiment; they were recorded by the assessment software. The possible range of score from helps requested can be from 0 to 9.

Number of errors: Each time the participant failed to perform the correct action the software registered it as an error. The total number of errors performed and their relative significance produce the most important contribute to the overall mark.

Total time taken: This OPI reflects the time taken by each participant to finalize the nine actions. The time taken in requesting and using the help contributed to the total time taken.

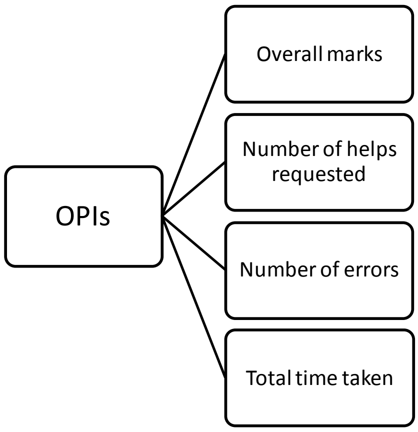


Figure 7 Operator performance indicators for catalyst injectors switch procedure.

Results

Overall marks

As shown in Figure 8, nine out of 12 (75%) participants of the IO group obtained the maximum overall mark (which means they did not make any errors and did not ask for any help and took a time to perform the nine actions below a suitable threshold), whereas only 4 out of 12 (33%) participants of the IA group were able to achieve the same optimal result. Moreover, the lowest score among the IO group is for participant 5 i.e. 16.55, while for the IA group participants 18, 19, and 24 obtained even lower scores. The overall marks for the participants of both the IO and the IA groups are shown in Figure 8.

With respect to our assumptions concerning the benefits of the IO training, these numbers reflect the better performance (in terms of overall marks) for the participants of the IO group when compared with that of the IA group, which is in accordance with the expectations of this study.

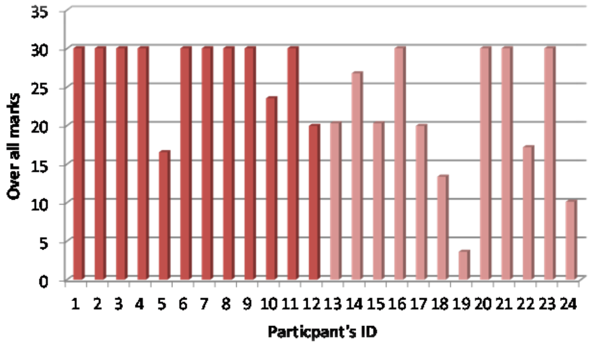


Figure 8. Overall marks obtained by each participant (IO group: 1 – 12, *IA group:* 13 -24*). Higher is better.*

Number of helps requested

In case of the IO group, only 3 participants (25%) needed any help for the completion of all the 9 steps. In case of the IA group there were 7 participants who asked for helps. Moreover, a deeper look into the data reveals that only one participant of the IO group requested help more than once (i.e. participant number 12 asked help 4 times). Conversely, four participants of the IA group requested more helps i.e. participants 17, 18, 19, and 24 requested helps 2, 2, 8, and 4 times respectively (see Figure 9).

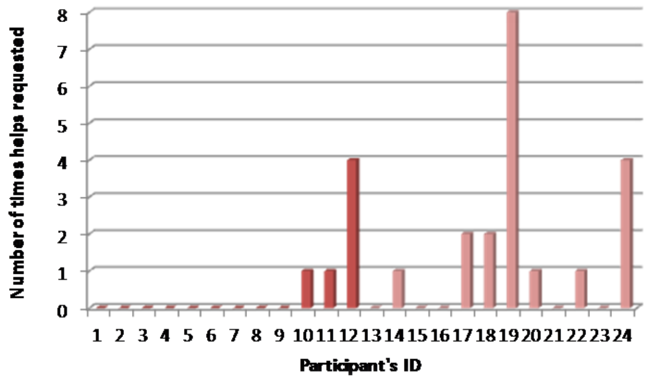


Figure 9. Number of helps requested by each participant (IO group: 1-12, IA group: 13-24). Lower is better

The total number of helps asked by the IO group was 6 as compared to 19 in case of the IA group. This shows that on average group IA required about 3 times more helps than the IO group to finalize the sequence of actions. This means that our assumption concerning the advantage of the IO training and its impact on the helps requested is confirmed.

Number of errors

Number of errors made by the IO group was 3.5 times less than those made by IA group; the IO group outperformed the IA group also in terms of error-free procedures. As a matter of facts, there were 9 participants of the IO group and only 4 participants of the IA group who did not make any errors during the assessment procedure.

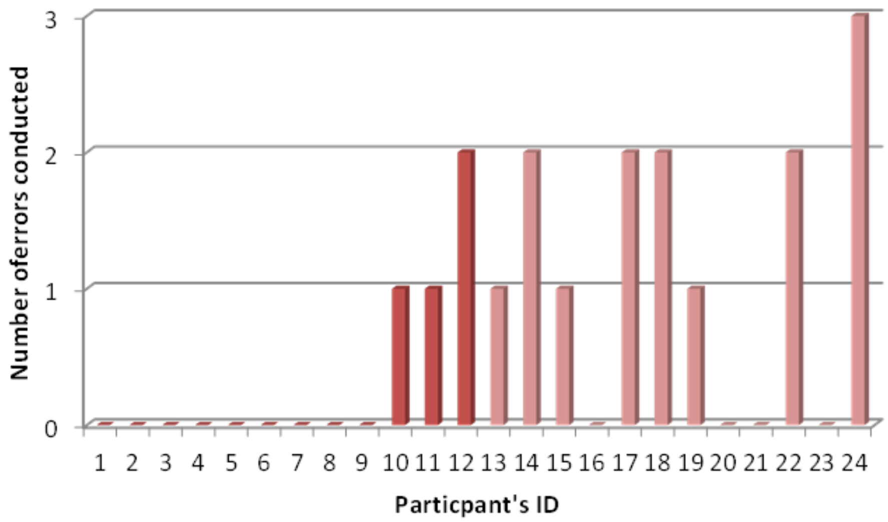


Figure 10. Number of errors made by each participant (IO group: 1 – 12, IA group: 13 -24, (lower is better)).

Total time taken

The last OPI calculated is the total time taken by each participant to finalize the experiment based on the 9 actions. The average time spent by the participants of the IO group was 407 s, that is 99 s higher than that of the IA group. Therefore, as per the assumptions of the study, the responsiveness and promptness of the IA group were higher than the ones of the IO group.

In summary, the IO group did perform better in accuracy and precision by running into fewer errors, whereas the IA group was faster in completing the same task but with a lower accuracy.

Discussion

This study has examined the impact of training methods, within an immersive virtual environment, on the performance of operators during a specific operating procedure. The overall mark reflects the accuracy with which the actions were performed. Moreover, it incorporates the number of helps requested and the relative significance of errors made. The IO group was more accurate and precise as was demonstrated by the individual overall marks. Therefore, the first assumption is correct and supports the earlier work (Nazir et al., 2013).

As mentioned in the results section, the number of helps requested by the IO group was far below the number of helps requested by IA group. The participants of the IO group probably developed a more accurate association for identifying the valves/devices and memorizing their location. They were supported in understanding the whole picture of the procedure and set of actions by observing the trainer and listening to his explanations. Similar reasons can be attributed to the number of errors made by the participants of each group. It is interesting to observe that 9 participants of the IO group did not make any error in identifying and operating the valves/devices. From a practical perspective, this is an important finding because during sensitive operations like catalyst-injectors switch, there is no space for any errors. One error in the sequence of valves/devices to be operated can lead to the emergency shutdown of the whole process. In case of error in the procedure sequence, the risk of undesired polymerization reactions inside the pipes (instead of the reactor) is rather high. When this happens, the whole plant must be shut down and the involved pipes and injectors must be maintained.

According to the authors' impression even if IA participants spent efforts and attention in following the colours to perform the sequence of actions they were probably distracted from focusing on the complete sequence, the location, and the arrangements of valves/devices in the IVE because a clear explanation of the meaning of the sequence was missing and learning was bottom-up. In contrast the IO group received a further explanation of the sequence meaning and reason. Therefore, the IO group was supported in memorizing, understanding, and following the sequence by observing the expert trainer performing the actions.

The total time taken for each participant of the IO group was higher than that of the IA group. Actually, the participants of the IA group performed all actions during their training by following the colours. Nevertheless, due to the lack of an extended view of the scenario, the precision and accuracy in following the sequence of actions were not comparable to the ones shown by the IO group (see Figure 10).

It is worth adding a further comment, which is based on the efficacy of training in IVEs. When the authors designed the experiment, there were vivid discussions about the possibility by participants to perform the whole sequence of actions without doing any errors. The overall task looked, at least theoretically, extremely challenging and complex. Actually, some authors were supposing the task almost impossible due to the difficulty of remembering a rather complex sequence of nine actions applied to valves and devices that are very similar and occupying a reduced volume (see Figure 3). The scepticism was finally washed away since there were a good number of participants (from both groups) who were able to achieve the maximum overall mark corresponding to an error-free sequence of actions. It is rather surprising thinking at how an immersive virtual environment can train in a short time (30-40 min) a trainee to perform a complex sequence of actions with a quite good efficiency. IVEs implemented within a PS show their efficacy as tools to improve the training of FOPs in complex environments.

The authors are aware of the fact that this study has limitations, e.g., that the participants were all students and not “real” operators. Students are more accustomed to receiving information in the form of lessons than that of being exposed to a new environment. This means that the results cannot be easily generalized to plant operators with a different educational background and learning capabilities. It can be assumed that operators are of a different age and differently educated than the participants involved in this study. It would be necessary to conduct similar studies with real operators and analyse their performance to draw valid conclusions for organizational applications.

Conclusions and future work

This paper proposed a comparison of two distinct training methods implemented in an IVE to understand and analyse their influence on the performance of operators based on the accuracy and speed reflected by proper OPIs. It was found that observing a trainer when s/he is completely immersed in the 3D plant with spatial sounds resulted in improved precision and performance as compared to the training method when the trainee follows the procedure and set of actions (once) in the same IVE by means of a fully automated guided tour. Conversely, the response time achieved by latter training method was better than the former one. This study reinforced the previous work (Manca et al., 2013, Nazir et al., 2013, Kluge, in press) that maintain that a PS improves performance of operators when they are facilitated by an expert trainer and opens the way to find more suitable methods to exploit IVE training applied to the process industry. Improved performance of operators results in averting emergency shutdowns, abnormal situations, and accidents, thereby, saving financial and human resources.

From a practical perspective, catalyst injectors switch takes place on average every two months in several plants around the globe. The operator must follow a well-defined sequence of actions. The impact of each action brings to different results in terms of its significance with respect to the process. Similarly, the error made among the sequence of actions is a function of its relative significance and impact on the overall process. Missing one action or doing an error, e.g., opening a valve and allowing flowing water or oil instead of catalyst has a completely different impact if the opposite action is made. Both errors are apparently similar i.e. operating a wrong valve, however, the damage in terms of cost and possible injuries associated can be quite different. This task is independent of teamwork; therefore, the FOP is responsible to carefully perform all the actions in a precise manner that requires cognitive skills like perceptual and spatial memory, attention management, responsiveness and self-reliance and should be well trained.

The use of IVE in training industrial operators is a relatively new concept. This study shows its advantages and paves the way to several questions which require further research and investigations, e.g., how much background knowledge of the process is necessary before IVE training? How many times the trainee should practice in the PS/IVE before being certified to work on a real plant? What is the most optimal way of interacting with IVEs during training session? Further investigations are necessary to find the most suitable training method within immersive environments and to find answers to the questions reported above.

References

Antonovsky, A., Pollock, C., & Straker, L. (2013). Identification of the human factors contributing to maintenance failures in a petroleum operation. *Human Factors*, 1-46.

Colombo, S., Manca, D. & Nazir, S. (2013). Virtual reality as effective tool for training and decision-making: preliminary results of experiments performed with a plant simulator. *Proceedings of SPE European HSE Conference and Exhibition Health, Safety, Environment and Social Responsibility*, 405-411.

Colquitt, J.A., LePine, J.A., & Noe, R.A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, 85, 678–707.

Ericsson, K.A., Krampe, R.Th., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.

ILO, International Labour Office, (2003). ILO Safety in numbers. Pointers for a global safety culture at work. Geneva.

Kluge, A. (in press) *The acquisition of knowledge and skills for task work and team work to control complex technical systems. A cognitive and macroergonomics perspective*. Dordrecht, The Netherlands: Springer

Kluge, A., J. Sauer, K. Schüler, & D. Burkolter. (2009). Designing Training for Process Control Simulators: A Review of Empirical Findings and Current Practices. *Theoretical Issues in Ergonomics Science 10*: 489–509.

Manca, D., Colombo, S., & Nazir, S. (2013). A plant simulator to enhance the process safety of industrial operators. Proceedings of SPE European HSE Conference and Exhibition Health, Safety, *Environment and Social Responsibility, 3*, 1-11.

Manca, D., Nazir, S., Lucernoni, F., & Colombo, S. (2012). Performance indicators for the assessment of industrial operators. *Computer Aided Chemical Engineering, 30*, 1422- 1426.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The psychological review, 63*, 81-97.

Nazir, S., Colombo, S., & Manca, D. (2013). Minimizing the risk in the process industry by using a plant simulator: a novel approach. *Chemical Engineering Transactions, 32*, 109-114.

Nazir, S., Colombo, S., & Manca, D. (2013). Testing and analyzing different training methods for industrial operators: an experimental approach. *Computer Aided Chemical Engineering, 32*, 667-672.

Nazir, S., Colombo, S., Gallace, A., Bordegoni, M., & Manca, D. (2012), Performance comparison of different training methods for industrial operators in D. De Waard, K. Brookhuis, F. Dehais, C. Weikert, S.Röttger, D. Manzey, S. Biede, F. Reuzeau, and P. Terrier, (Eds.) (2012), Human Factors: a view from an integrative perspective,. *Proceedings HFES Europe Chapter Conference Toulouse 2012*, 235-242.

Noe, R.A., & Colquitt, J.A. (2002). *Planning for training impact: Principles of training effectiveness*. In K. Kraiger (Ed.), Creating, implementing, and maintaining effective training and development: State-of-the-art lessons for practice (pp. 53–79). San Francisco, CA: Jossey-Bass.

Proctor, R.W., & Dutta, A. (1995). *Skill acquisition and human performance*. Thousand Oaks, CA: Sage.

Reinach, S., & Viale, A. (2006). Application of a human error framework to conduct train accident/incident investigations. *Accident Analysis & Prevention, 38,* 396–406.

Salas, E., & Rosen, M.A. (2013). Building high reliability teams: Progress and some reflections on teamwork training. *BMJ Quality and Safety, 22,* 369-373.

Shi, D., Li, X., Su, P., Zheng, Z., & Luo, Z. (2010). Modeling of loop reactor for polypropylene Spheripol technology. Shiyou Xuebao, Shiyou Jiagong/Acta Petrolei Sinica (*Petroleum Processing Section), 26,* 582-587.

Urdampilleta, I., Gonzalez, A., Iruin, J.J., De La Cal, J.C., & Asua, J.M. (2006). Origins of product heterogeneity in the spheripol high impact polypropylene process. *Industrial and Engineering Chemistry Research, 45,* 4178-4187.

Strain caused by Head Mounted Displays

Matthias Wille, Britta Grauel, & Lars Adolph
German Federal Institute for Occupational Safety and Health (BAuA)

Abstract

Head mounted displays (HMDs) have become smaller, lighter, brighter and less expensive during the last years. They are now on the threshold to mass market and a wider field of applications. During work they can be used as an adaptive assistant device showing context-sensitive information while hands are free for the main task. Within the project "head mounted displays – conditions of safe and strain-optimised use" of the German Federal Institute for Occupational Safety and Health (BAuA) a study was carried out: 41 subjects worked for four hours on a construction task and a parallel monitoring task. The content was either presented on a HMD or on a tablet PC. Besides performance within both tasks objective strain parameters (NASA-TLX, RSME) and a visual fatigue questionnaire were selected. Measurements at different timestamps gave insight into the strain trend over time comparing both display technologies. Results showed a higher strain and slower performance while working with a HMD compared to a tablet PC. Another outcome was a stronger increase over time in strain and visual fatigue while using a HMD, especially for elderly users. This implicates that working conditions, e. g. rest periods, have to be adjusted to fit to new technologies.

Introduction

In 1998, Keller and Colucci reported, that head mounted displays (HMDs) did not meet users´ expectations at that time. Nevertheless, they expected HMDs to be useful tools if design problems would be resolved. Now, 15 years later, the design of HMDs shows progress. HMDs are becoming lighter and smaller. Completely new solutions are at the prototype stage.

Monocular HMDs allow users to receive data from the real environment and additional information at the same time (Kawai et al., 2010). Thus, they can adopt the primary functions of job aids: provide information, prompt procedures of coach perspectives and decisions. The main intention of using job aids is to improve performance by reducing cognitive load (Rosset & Gautier Downes, 1991; Salvendy, 2012).

One advantage of HMDs over traditional job aids like manuals or new display technologies like handhelds is that they can provide the user with context sensitive information while working hands-free. Their usage is particularly suitable if tasks contain several steps, information can be provided in a graphical format, information is time critical and there is little tolerance for errors (for detailed information see

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Grauel, Kluge & Adolph, 2011). Nakanishi, Taguchi and Okada (2010) showed that visual on-task instructions using see-through HMDs are more suitable than auditory instructions for preventing careless skill-based errors. A comparison study (Glumm et al., 1998) showed that soldier performance of land navigating was better if using a helmet mounted display compared to using traditional navigating equipment. Soldiers also reported lower levels of mental workload.

However, HMDs can create problems which may not occur if traditional job aids or displays are used, because the viewing situation is not natural (Patterson, Winterbottom & Pierce, 2006). Kawai et al. (2010) stated that, in spite of technological progress, many questions regarding usability and usefulness remain unanswered.

To identify the conditions for strain-optimized use of HMDs to improve job performance a long term study which addresses the optimization of work and task load was conducted.

Method

To investigate strain and visual fatigue within longer sections of using head-mounted displays, a study was conducted in the laboratories of the BAuA. Participants worked in 3 sessions for 4 hours each with an HMD or comparatively with a tablet PC. Hereby the first session was always held with the HMD, while the repetition with the HMD and the comparison with the tablet PC were permutated as the 2nd and 3rd session among subjects. The repeated measurement with the HMD was arranged to estimate a first habituation to the new technology. Between the sessions were at least two and a maximum of 10 days.

Participants

A total of 41 subjects participated in the study with an age of 18-67 years. The sample was divided by median split into two age groups to investigate possible age effects. In the younger group were 21 subjects between 18-31 years (Mean = 23.48, SD = 3.341, 9 male /12 female). In the older group were 20 subjects aged 38-67 years (Mean = 50.55, SD = 9.214, 11 male / 9 female). None of the subjects had previously worked with an HMD.

Tasks

The subjects had to perform two tasks which were instructed as equal important (dual task paradigm): First, they should assemble model cars from building bricks, based on a famous construction toy. The assembly instructions were displayed on either the HMD or on the tablet PC. These image-based, step by step assembly instructions were chosen because it resembles industrial assembly instructions, such as those being used in maintenance. In a parallel presented monitoring task, subjects were asked to pay attention to three vertical bars on the outer edge of the screen and confirm certain state changes with a button press. The bars varied continuously, but very slowly in length and from time to time in its colour (blue / red). Colour change which includes a visual pop-out effect by the large change in presentation, should be confirmed by one button. On a second button each a change of position of the longest bar, caused by the continuous variation in length of the bar should be confirmed. The variation of the bars were programmed randomly, but on average a change in bar colour happened every 140 seconds and a change in the longest bar position every 95 seconds. As a further dependent variable in half of the blocks feedback was given on the last confirmation in form of a written notice on top of the bars (e.g. “centre - red”), while in the other blocks such feedback was not given.

Apparatus

The used HMD was a MAVUS-System from “Heitec” company, as it is currently used in industrial maintenance. It is a monocular look around display with a resolution of 800 x 600 pixels. The technique is fixed to a head carrier which includes a front camera and a headset. During the study the camera and headset had no function, while in industrial applications they are provided for communication with experts. The head carrier including all technology weighed 380 grams and was connected via a cable with a vest which included the radio technology for the transmission of data and the accumulator for power supply. As Tablet PC the CL900 by Motion was used. To ensure that the representation of the work content was comparable, only a window of 800 x 600 pixels was shown and the rest of the area was covered. All interactions - forward and backward switching the construction slides and buttons for confirming the bar tasks - took place over a converted number pad.

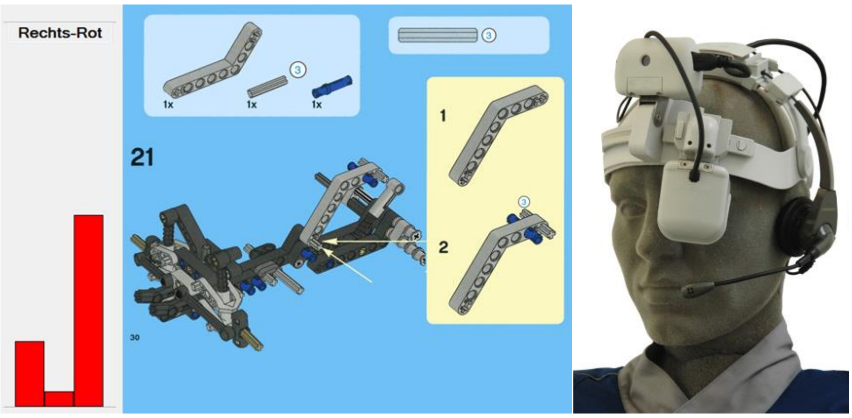


Figure 1. Left: work content as presented on the HMD. Right: HMD used in the study.

Procedure

The subjects were seated in a well-lit room in front of a height-adjustable table. At the beginning of the experiment, the table was set to the body dimensions of the participant. The experimenter sat behind a partition and monitored the session which was fully recorded on video. At the beginning of the first session participants filled out a questionnaire about demographical aspects and were informed about the test procedure in common and that they could cancel the experiment every time they want. After that, their experience with different kinds of technologies was asked and a questionnaire about technology affinity (TA-EG by Karrer et al., 2009) was completed. Then the RSME scale (Rating Scale of Mental Effort, Zijlstra, 1993) for monitoring mental strain was introduced. Subjects were asked to imagine different typical daily routine situations and to rate them on the RSME scale to get some practice in using the scale which was retrieved later at different points in time during the session. Afterwards, the instruction describing the tasks and the session procedure in common was given in a written form. If participants had any questions they were answered by the experimenter. Subsequently the HMD was shown to the subjects and all functions were explained. After that, the head carrier was adapted to the subjects and the HMD was put on. Hereby the monocular HMD was always mounted in front of the individual guiding eye of the subject and the bar-task always showed up in the outer region of the visual field (left side for left dominant eye, right side for right dominant eye). Then the combination of the two tasks was practiced for about 10 minutes in the first session. The following session was carried out in 4 blocks of 50 minutes each. There was a break of 5 minutes between the first and last two blocks where subjects keep the head carrier of the HMD on, but folded the display out of their sight. Between block 2 and 3 – in the middle of the session - was a major break from 15 minutes where the HMD was removed and the subjects were asked to stand up and move around a bit.

During the session, several subjective stress parameters were measured: The RSME, which reflects the perceived strain on a scale of 0-150, was administered in the middle and at the end of each block during the work. So in total eight measurement points are available to give insight in the development of strain over time. The NASA - TLX (Hart & Staveland, 1988) with its six subscales was not presented directly in the work situation, but right at the beginning of the big break and at the end of the experiment, so here are two measurement time points available. The pair comparison for the individual weight of the subscales was carried out at the end of the experiment. Furthermore, a questionnaire on visual fatigue (Visual Fatigue Questionnaire after Bangor, 2000) was administered at 5 time points: immediately before the experiment to determine the individual initial situation and directly after the end of each block. (As an objective strain parameter the heart rate was collected during the entire experiment, too. These data are still in analysis.) At the end of each test session, an interview was conducted with the subject.

Experimental design and variables

This study followed a repeated measurement plan with multiple dependent and independent variables. Independent variables are the experimental session - first measurement HMD ("HMD1"), second measurement HMD ("HMD2") and tablet PC ("Tab") - and the feedback in the monitoring tasks (blockwise change). The age and technology affinity of the subjects are between subjects factors (each median split into 2 groups). Dependent variable was the number of construction slides per session, as a performance indicator in the assembly task. This was possible because the complexity of each slide is similar across different car models and complex models primarily consist of more slides. Further dependent variables were the hit rate and reaction time in the monitoring tasks, RSME values at 8 times, NASA - TLX values at 2 times and Visual Fatigue Questionnaire at 5 times.

The data was analyzed in SPSS 20 using ANOVA with repeated measures. Here, two experimental sessions were always compared in pairs. A comparison over all 3 sessions at once would be not allowed with regard to contents because they do not represent levels of one factor, but in one case it is a repeated measurement to derive habituation effects and in the other case a comparison of different measurement technology. For purposes of clarity this article is primarily concerned with the comparison between the second HMD measurement and the Tablet-PC measurement, representing differences between display technologies. The second measurement with the HMD was chosen, because here a sufficient familiarization with the new technology already was given. The comparison between first and second measurement with the HMD which can describe habituation effects with the technology is also described if significant effects appeared.

Results

Construction task

As a performance indicator in the construction task, the number of construction slides processed per session was used. If comparing across both age groups with a HMD significantly less steps were done than when using a tablet PC (HMD1 = 130.9 , HMD2 = 128.8; Tab = 157.7). Taking the age factor into account the main effect of display stayed significant [$F(1, 37) = 24.285$, $p < .001$] and furthermore a highly significant main effect of age showed up [$F(1, 37) = 23.647$, $p < .001$]: Younger participants processed more constriction slides per session (see Figure 2). A habituation effect on the HMD by the comparison between the first and second measurement was not detected [$F(1, 37) = .208$, $p = .651$].

Monitoring task

The monitoring task showed a highly significant difference between the task types “bar colour” and “bar length”: changes in bar colour always had a higher hit rate [$F(1, 36) = 84.690$, $p < .001$] and lower response time [$F(1, 36) = 17.560$, $p < .001$] as the change of position of the longest bar (see tables 1 and 2). This effect illustrates the effect of a visual popout by the large variation in colour change and was expected. Similarly, the feedback on the last confirmed position led to a highly significant increase in the hit rate [$F(1, 37) = 58{,}257$, $p < .001$] in both bar tasks. But the reaction time showed no main effect of feedback [$F(1, 37) = .078$, $p = .782$]. The display type showed no significant effect on the hit rate [$F(1, 37) = 3.635$, $p = .065$] or reaction time [$F(1, 37) = 2.858$, $p = .100$]. However, the tendency is towards better values for the tablet PC. A habituation effect for the HMD by comparison of 1st and 2nd measurement showed up in terms of the hit rate [$F(1, 37) = 5.395$, $p = .026$], but not for the reaction time [$F(1, 37) = .196$, $p = .661$]. An effect of age was not found in the monitoring task.

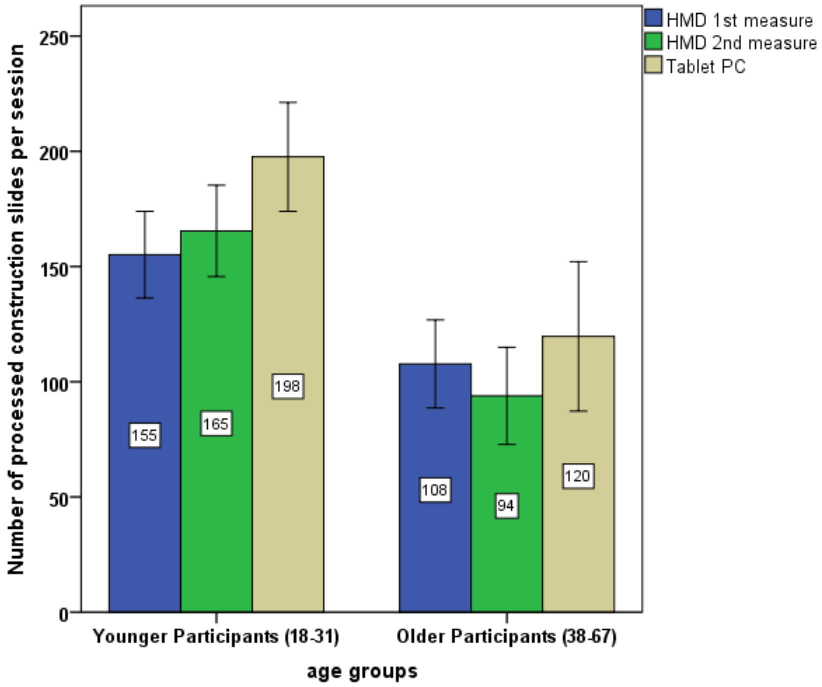


Figure 2. Number of processed construction slides per session by age groups. Error bars reflect the 95% confidence interval.

Table 1. Average hit rate in percent for the monitoring tasks by test session and feedback (standard deviation in parentheses).

	Feedback	*HMD 1st measurement*	*HMD 2nd measurement*	*Tablet PC*
Colour changes	no feedback	78.17 (15.91)	81.02 (14.99)	80.86 (15.12)
	given feedback	87.30 (15.27)	89.85 (13.25)	92.62 (11.18)
Length changes	no feedback	63.25 (23.43)	66.11 (21.96)	71.39 (18.45)
	given feedback	73.99 (21.58)	79.75 (18.07)	81.23 (14.07)

Table 2. Average reaction time in seconds for the monitoring tasks by test session and feedback (standard deviation in parentheses).

	Feedback	*HMD 1st measurement*	*HMD 2nd measurement*	*Tablet PC*
Colour changes	no feedback	9.60 (5.91)	11.50 (11.21)	11.86 (9.47)
	given feedback	13.68 (8.26)	12.85 (8.98)	11.50 (6.65)
Length changes	no feedback	20.29 (21.64)	20.93 (21.36)	17.66 (19.21)
	given feedback	19.42 (12.48)	19.99 (14.17)	16.14 (10.61)

Subjective Strain

The subjective strain measured by the NASA-TLX (see figure 3) showed a significantly higher score for the HMD compared to the tablet PC [$F(1, 37) = 26.952$, $p < .001$]. The increase of perceived stress over time is also significant [$F(1, 37) = 14.267$, $p =. 001$], but no interaction between display type and measurement time showed up [$F(1, 37) = 1.375$, $p =. 253$]. No main effect of age was found, but an interaction between display type and age reached statistical tendency [$F(1, 37) = 3.172$, $p =. 083$], based on a higher increase in strain over time on elderly participants. A habituation effect between 1st and 2nd measurement with the HMD can not be proved.

For the first time an influence of technology affinity showed up in interaction with the display type [$F(1, 37) = 5.000$, $p = .031$]: Participants with lower scores on technological affinity showed not only a higher stress score in general (HMD: 73.18, tablet PC: 66.17), but also less of the discharge in strain by the tablet PC compared to users with higher technology affinity (HMD: 68.96, tablet PC: 51.35).

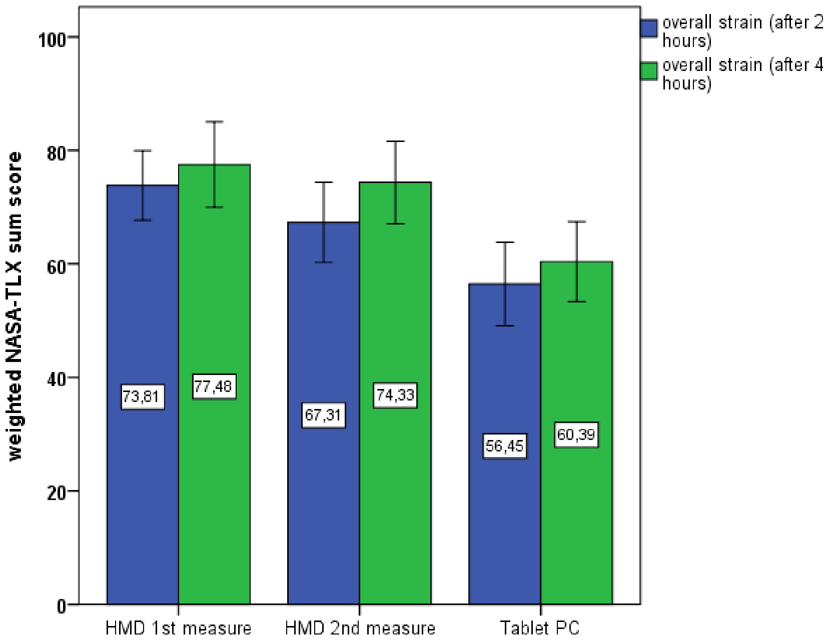


Figure 3. Weighted NASA-TLX sum score after 2 and 4 hours of working by session. Error bars reflect the 95% confidence interval.

The RSME value (see figure 4) also showed a highly significant increase in strain for the HMD compared to the Tablet PC [$F(1, 35) = 84.435$, $p < .001$], an increase over time [$F(1, 29) = 6.956$, $p < .001$] and a main effect of age [$F(1, 35) = 6.684$, $p = .014$], which again indicates that seniors are more stressed. The development of strain over time also showed the influence of breaks on the strain rating: The short breaks of about 5 minutes after 60 and 180 minutes – where the HMDs head carrier were kept on – seemed only to have little effects. However, the longer break after 120 minutes – where the HMDs head carrier was taken off – seemed to result in lower values on next strain request for younger participants, while older participants did not profit from that break.

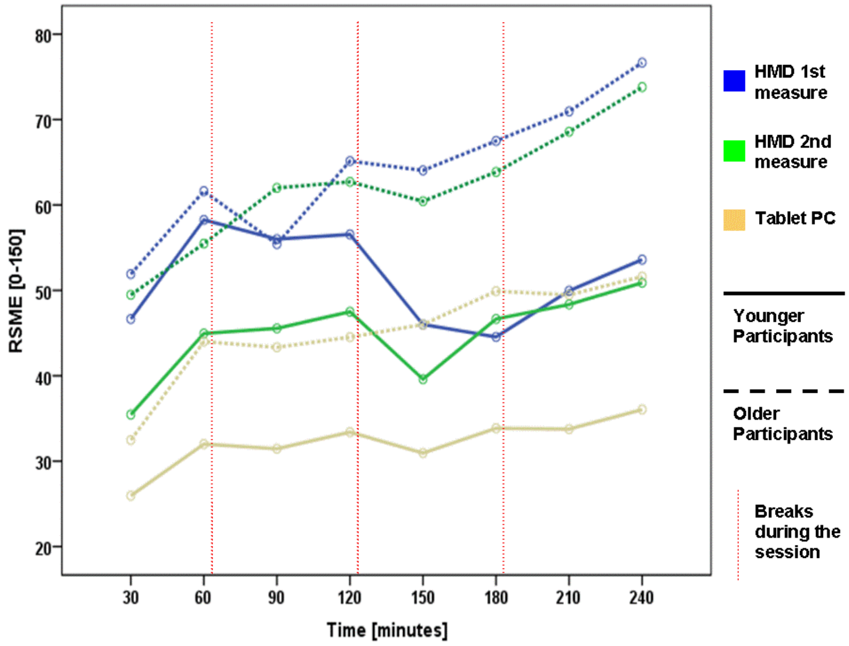


Figure 4. RSME score over time by session and age groups.

Visual fatigue

The Visual Fatigue Questionnaire (VFQ) collects various aspects of visual fatigue within 16 items. Each item is recorded on a ten-point scale, a sum score does not exist. The questionnaire was given before the start of the session and then hourly. Results showed a clear, highly significant change of almost all item scores, which however remain at a low average level of 1-3.5. All items showed an increase over time and higher strain levels on the HMD. For many items, the interdependency time x display was significant which reflects a greater increase over time during the HMD sessions.

As an example the item "difficulty in seeing sharp" showed a main effect of display [$F(1, 36) = 28.662$, $p < .001$] , a main effect of time [$F(1, 33) = 8.510$, $p < .001$] and an interdependency display x time [$F(1, 33) = 7.668$, $p < .001$]. But here also a reduction of the item scores in comparison between 1st and 2nd measurement with the HMD showed up [$F(1, 37) = 9.319$, $p = .004$] indicating that participants getting used to the display technology. Equally noteworthy are the items that are not directly attributable to visual fatigue: “headache” showed higher values for HMD [$F(1, 37) = 21.435$, $p < .001$], and an interdependency display x time [$F(4, 34) = 5.546$, $p = .002$] indicating a greater increase of headache over time while working with the HMD compared the tablet PC.

Effects on “neck pain” went into same direction with a main effect for display [$F(1, 37) = 22.000$, $p < .001$] and an interdependency display x time [$F(4, 34) = 5.341$, $p = .002$]. And also “mental fatigue” showed significant effects (display [$F(1, 37) = 7.956$, $p = .008$] , time [$F(4, 34) = 16.665$, $p < .001$], display x time [$F(4, 34) = 3.688$, $p = .013$]), but here also a habituation effect in the comparison between 1st and 2nd measurement of HMD showed up [$F(1, 37) = 4.175$, $p = .048$].

In a final interview, many participants complained about the weight and discomfort of the HMDs head carrier. The monocular representation in front of only one eye was not a problem for most subjects after a short period of habituation. Only one of 41 subjects preferred the HMD for work, while the others found working with the Tablet PC more comfortable.

Discussion

The results generally show poorer performance and higher strain when working with a HMD compared to a tablet PC: participants work slower in the picture based construction task if presented on a HMD and the hit rate in a parallel monitoring task is also worse compared to presentation on a tablet PC. That last point illustrates that parallel monitoring is not necessarily easier, if the stimuli is presented always within the field of view. As attention needs to be focused on that area, too, even a presentation on a tablet PC, which is not always in direct sight, can result in a better reaction to stimuli.

One typical problem if investigating such new technologies like HMDs is that participants are often unfamiliar with this technology. Although they train tasks and handling of technology for a short moment before the measurement starts, their work performance and ratings will still be under the influence of getting used to the technology. Therefore in this study the repeated measurement with the HMD in a second session was done. However, habituation effects occurred only hardly in that second session with HMD. But habituation to such new technology may take longer than just 2 sessions of 4 hours each investigated during this study. Here a long term investigation will be useful that measures how people get used to that technology over weeks or months.

Some of the reported results are probably due to the hardware design of the used HMD: While most current HMDs still use a head carrier - similar to a bicycle helmet - to which the equipment is fixed, newer HMDs, which will be available during the next months or years, put all this technology in some slightly larger glasses. Head- and neck pain measured during this study could be based on the weight of the head carrier and be less dominant if using those upcoming HMDs. Also it is to mention that in this laboratory study, the benefits of HMDs - the mobility and the hands remain free - did not come into play, which could have a negative impact on the here reported mainly subjective parameters.

The approach of a long-term investigation and process-accompanying strain survey has proven itself and could often show a stronger increase over time in addition to higher strain with the HMD. This is especially important, if HMDs are used as a working assistance system. Based on these findings, recommendations for break times have to be reconsidered. For a more detailed presentation of the results and further discussion please also refer to the BAuA research report that will be published in 2014.

References

Bangor, A. W. (2000). *Display technology and ambient illumination influences on visual fatigue at VDT Workstations.* PhD thesis, Virginia Polytechnic Institute and State University, USA.

Glumm, M.M., Marshak, W.P., Branscome, T.A., Mc Wesler, M., Patton, D.J., Mullins, L.L. (1998). *A comparison of soldier performance using current land navigation equipment with information integrated on a helmet-mounted display.* Army Research Laboratory, Aberdeen Proving Ground.

Grauel, B., Kluge, A. & Adolph, L. (2011). Unterstützung makrokognitiver Prozesse in Teams durch Head Mounted Displays: Methoden der Eignungsermittlung und Gestaltung. In M. Grandt (Ed.): *Ergonomie im interdisziplinären Gestaltungsprozess: 53. Fachausschusssitzung Anthropotechnik der Deutschen Gesellschaft für Luft- und Raumfahrt Lilienthal-Oberth e.V. 27. und 28. Oktober 2011, Neu-Isenburg*. (pp. 71-79) Bonn: DGLR,

Hart, S.G., &Staveland, L.E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock & N. Meshkati (Eds.) *Human mental workload* (pp. 139-183). Amsterdam: North-Holland.

Karrer, K., Glaser, C., Clemens, C., & Bruder, C. (2009). Technikaffinität erfassen– Der Fragebogen TA-EG. In A. Lichtenstein, C. Stößel & C. Clemens (Eds.): *Der Mensch als Mittelpunkt technischer Systeme. 8. Berliner Werkstatt Mensch-Maschine-Systeme* (pp. 196-201). Düsseldorf, Germany: VDI Verlag GmbH. Online available: http://www.tu-berlin.de/fileadmin/f25/dokumente/8BWMMS/13.3-Karrer.pdf [August 2013].

Kawai, T., Häkkinen, J., Yamazoe, T., Saito, H., Kishi, S., Morikawa, H., Mustonen, T., Kaistinen, J. & Nyman, G. (2010). Ergonomic evaluation of ubiquitous computing with monocular head-mounted display. *Proc. SPIE 7542, Multimedia on Mobile Devices 2010*, 754202.

Keller, K. & Colucci, D. (1998). Perception in HMDs: what is it in head-mounted displays (HMDs) that really make them all so terrible? Proc. SPIE 3362, Helmet- and Head-Mounted Displays III, 46.

Nakanishi, M., Taguchi, K. & Okada, Y. (2010). Suggestions on the applicability of visual instructions with see-through head mounted displays depending on the task. *Applied Ergonomics, 41*, p. 146-155.

Rosset, A. & Gautier-Downes, J. (1991). *A Handbook of Job Aids.* San Francisco: Jossey-Bass Pfeiffer.

Salvendy, G. (2012). *Handbook of Human Factors and Ergonomics.* Hoboken: John Wiley and Sons, Inc.

Patterson, R., Winterbottom, M.D. & Pierce, B.J. (2006). Perceptual Issues in the Use of Head-Mounted Visual Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48, 3,* pp. 555-573.

Wille, M., Grauel, B. & Adolph, L. (2011). Methodische Vorgehensweise zur Beanspruchungsermittlung bei Head-Mounted Displays. In *Reflexionen und Visionen der Mensch-Maschine-Interaktion – Aus der Vergangenheit lernen, Zukunft gestalten. 9. Berliner Werkstatt Mensch-Maschine-Systeme*. (pp. 84-85) Düsseldorf, Germany: VDI Verlag, Fortschrittsberichte, Band 33.

Zijlstra, F.R.H. (1993). *Efficiency in work behavior: an approach for modern tools.* PhD thesis, University of Delft. The Netherlands. Online available: http://repository.tudelft.nl/view/ir/uuid:d97a028b-c3dc-4930-b2ab-a7877993a17f/ [August 2013].

Wearable crew support technology on the International Space Station: the mobile Procedure Viewer (mobiPV)

Keshav Chintamani [1], Boris Van Lierde[1], Shane Maloney[2], Paul Kiernan[2], David Martinez Oliveira[3], & Mikael Wolff[3]
[1]Space Applications Services, Belgium
[2]Skytek, Ireland
[3]European Space Research and Technology Centre (ESTEC), The Netherlands

Abstract

Astronauts on the International Space Station (ISS) constantly rely on procedures for system and payload operations. These procedures based on the Operations Data File (ODF) standard have been traditionally visualized using print media and fixed laptops. Typical crew tasks on the ISS demand mobility and there is an increasing need for wearable, hands-free technologies that support procedure visualization and execution, simultaneously enabling collaboration with flight control teams on Earth. The European Space Agency (ESA) has identified this technology gap and is developing the Mobile Procedure Viewer or mobiPV. A user-centred crew support tool that relies on mobile computing hardware; mobiPV enables procedure execution, 3D data visualization in augmented reality, note-taking and provides peer-to-peer multimedia communications for collaboration, including text, audio and video communication and white-boarding. These features are provided in a mobile and hands-free manner relying on wearable displays and speech recognition technologies for user interaction. We walked through system concepts and initial UI mockups with end-users at the European Astronaut Center (EAC) and gathered their feedback on the system's features and expected usability. This article presents the state of the art in wearable systems in space operations, introduces the mobiPV system and discusses the main findings of the usability workshop.

Background

Human spaceflight and the International Space Station (ISS) program are enormous endeavours where governments, space agencies, industry and academia across many nations work together to ensure that the environment of space can be understood for the benefit of mankind. International crew-members live on-board the ISS for varying periods of time performing complex tasks ranging from assembly, inspection, maintenance, system operations and scientific investigations. ISS tasks can range from laptop usage procedures, robotic teleoperation, biology experiments to ultrasound diagnosis (non-exhaustive) and are at times time-critical, such as diagnosing a malfunctioning payload or performing an emergency routine. Although the crew are trained for most of these tasks on-ground, they are still required to execute these tasks following strict procedures on-board which are based on

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

standard task description schemas such as the Operations Data File (ODF) (NASA, 2010).

1.101 KUBIK 6 SETUP
(KUBIK/ALL/FIN/HC) Page 7 of 8 pages

4.2 Align Centrifuge Insert (KUB-FM1-CI) on top of Kubik 6 (cut corner of the base plate, on upper left position relative to the front panel, Figure 9)

4.3 Insert Centrifuge Insert (KUB-FM1-CI) into Kubik 6

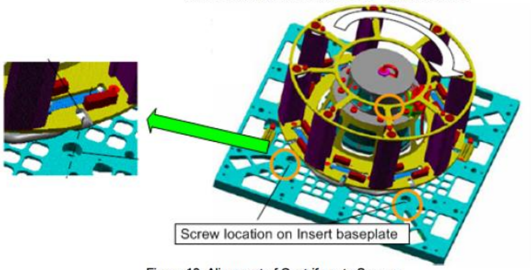


Figure 10. Alignment of Centrifuge to Screws

4.4 Align the holes on the Centrifuge towards the screws on the baseplate of the Insert (Figure 10)

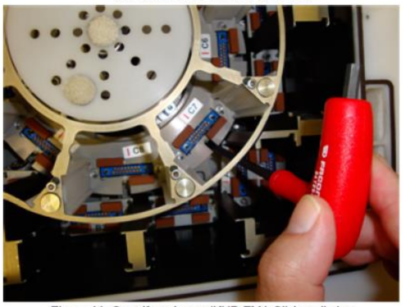


Figure 11. Centrifuge Insert (KUB-FM1-CI) Installation

Figure 1. Payload ODF procedure with reference images (Courtesy: European Space Agency)

1.403 HIGH DEFINITION VIDEO CAMERA ASSEMBLY (HD-VCA) INSTALLATION AS VCA1
(MSM/INC34/FIN/HC) Page 9 of 9 pages

3.3 Fix Fibreoptics Converter, part of HD-VCA SUP Cable, to SUP handrail. Use velcro ties as required (Refer to figure 10).

CAUTION
Over-bending of HD-VCA SUP Cable may result in damage of fibreoptics. When coiling the minimum bend radius is 50mm.

3.4 HD-VCA SUP Cable P08 →|← SUP1 J08 DATA5
HD-VCA SUP Cable P06 →|← SUP1 J06 DATA3

Carefully coil remaining cable length of Fiberoptics Cable and fix as required, using velcro ties.
Avoid Bending Radius < 50mm

4. CLOSEOUT

4.1 Take picture of installed HD-VCA and cable routing (Digital Camera).

4.2 VCA1 lens connector ←|→ VCA1 body

Figure 2. Payload ODF procedure with warnings and symbols (Courtesy: European Space Agency)

The ODF standard is implemented for almost all ISS crew procedures including subsystem, science, crew-health and payload operations. The standard is implemented for ground operations as well and used by flight controllers, payload instructors and operators. Figure 1 and 2 provide examples of system ODFs.

It might surprise some readers who expect that the astronauts working in microgravity execute these complex tasks aided by advanced displays and wearable computers that the crew still rely on print media and fixed laptops to view procedures (Figure 3).

These choices are not because space agencies want to keep their crew from using the latest, cutting-edge computers. Safety regulations and tough hardware qualification tests for space flight can at times prevent or significantly delay new technology from reaching the ISS. Today, dedicated software on these laptops provides access and renders the procedures which are uploaded to a flight database ahead of a crew member's activity.

On-board tasks are rarely performed alone by the crew. Flight controllers in Europe and the United States oversee and support the crew on a 24/7 basis. There are limitations here as well. Planning information and digital content needed is prepared in advance of the actual tasks because communication coverage, data approval and verification processes and available technical infrastructure impedes and delays the availability of real-time support from the ground. The only real-time channels are full-duplex voice and unidirectional video downlink. Even these are constrained to specific mission control centers, preventing other end-users, such as scientists or payload specialists from working directly with the crew. In the technical environment of the ISS, there is a strong need for real-time collaboration during task execution.

Figure 3. ESA Astronaut Thomas Reiter executing a checklist procedure in the US Destiny laboratory (image courtesy: ESA/NASA Image Archives)

These limitations have not gone unnoticed. Both ESA and NASA are constantly investigating process and technology changes to improve crew task performance and to support collaboration between crew and ground specialists.

In recent years, increased attention has been paid towards wearable systems, which are well-suited for the ISS environment. A simple use case is provided for clarity. The ESA crew work in the Columbus module where a number of payload racks are installed. These racks comprise of instruments and hardware for on-board experiments and assembly, operations and maintenance of these systems require simultaneous payload procedure access and manual tool manipulation. The crew often have to work in difficult-to-access locations on the ISS and laptops are not always available here.

In such scenarios, wearable displays in combination with mobile computers can eliminate the need for cumbersome and bulky laptops and provide critical task-relevant procedural information on-demand to the crew. Wearable systems can dramatically reduce valuable crew time and in combination with interactive technologies such as speech recognition and augmented reality, improve task performance levels on-board spacecraft.

Wearable Systems for Human Spaceflight

In recent years, ESA has initiated a number of technology development activities to evaluate wearable procedure support systems. The WEAR++ project (Figure 4b,c,d) demonstrated by Astronaut Frank De Winne in 2009 evaluated a head-mounted system that allowed procedures to be viewed and navigated using speech recognition with information relevant to the procedure overlaid on real-world objects in 3D using augmented reality (De Weerdt et al., 2010). Despite some limitations, due in most part to the unavailability of mobile computers, operational challenges with accurate augmented reality rendering and immature speech recognition technology; the experiment highlighted the benefits of mobile procedure displays in flight operations. More recently, the voice activated procedure viewer (vaPV) system was tested on-board (Figure 4a) to evaluate the benefits of speech recognition as a hands-free input modality for procedure navigation (Wolff, 2013).

Similar concepts have been applied in other space related scenarios, such as medical diagnosis and spacecraft assembly. The Computer Aided Medical Diagnosis and Surgery system (CAMDASS) demonstrated the benefits of head mounted displays and 3D augmented reality guidance in supporting untrained crew with health diagnosis using ultrasound (Nevatia et al., 2011). The Portable Virtual AIT Visualiser (PVAITV) system (Naessens, 2011) demonstrated the benefits of augmented reality and procedure guidance during the assembly of mechanical components in spacecraft cleanrooms.

Although these systems have helped demonstrate the benefits of heads-up procedure support, it has taken some time for the required technology to reach a level deemed feasible for spaceflight.

In recent years, improved optics and display technologies have resulted in a number of light-weight and ergonomic head mounted displays. Simultaneously, the mobile technology market is now flooded with powerful tablets and smartphones with display resolutions higher than conventional desktop monitors. A number of companies such as Vuzix and Google are now manufacturing wearable computing solutions that combine head worn computers with integrated displays, audio and trackers. In comparison with heavy HMDs and bulky computers, these devices can help address technology needs on the ISS and support the development of operational systems to improve crew tasks.

Figure 4 Clockwise from top-left: a) Astronaut Chris Hadfield testing voice recognition on the CRUISE experiment (credit: ESA/CSA); b) Astronaut Frank De Winne training to use the WEAR system (credit: ESA image archive); c) The WEAR head mounted procedure viewer; 4) WEAR user interface rendering procedures and guidance information using augmented reality

In an on-going activity titled mobile procedure viewer (mobiPV), the ESA and industry are creating a system that aims at solving these challenges faced by the crew such as just-in-time information access, mobility and collaboration. In the following sections, we introduce the system concept and features, outline additional needs identified by end-users at a recent workshop and highlight the human factors challenges needed to make this system an integrated crew tool in ISS operations.

The Mobile Procedure Viewer Project (mobiPV)

mobiPV is a system designed to assist the crew-member(s) during task execution on-board the ISS, enhance collaboration between crew members or ground stakeholders and is intended as a permanent crew assistant during operations. Primarily a procedure viewer, it provides the crew with a range of complimentary features aimed at improving efficiency during on-board tasks. Its main features are listed in Table 1.

Table 1. Key mobiPV user-oriented and system features

	Feature	Usage
Software	**Procedure Execution**	Access procedures stored in flight databases and rendering of step by step instructions to the crew
	Note-taking	Create and save task critical text, image and video notes during task execution for post-analysis
	Reference materials	Access to information needed during tasks such as multi-media, documents and images
	Task Planning	Access to daily task schedules and plans prepared by ground personnel
	3D Rendering	Rendering of animated 3D CAD models to support assembly, inspection and maintenance tasks
	Communication	Peer-to-peer communication including text messaging, audio and video conversations and white-boarding
	Collaborative Procedures	Shared control of procedures between ground and space, and on-the-fly modification of procedure content
User Interfaces	**Head-mounted displays**	Heads-up rendering of procedure and advanced 3D content
	Speech Recognition	Hands-free navigation of graphical user interfaces and data input
	Wearable computers	Mobile processing units with integrated touchscreens for manual data entry and user interface navigation
	Head-worn cameras	Support first-person image and video sharing with collaborators for situational awareness and task support

On-board the ISS, mobiPV allows the crew-member to access procedures that are located in flight databases. Before the start of an activity, it can provide access to the crew-member's daily task plan and critical information that the crew needs to know before the start of the activity, either of which may have been recently updated.

Once a procedure is loaded, the crew-member navigates this procedure either using voice commands or the touchscreen. The current step can appear highlighted or displayed independent of other steps. While on a particular procedure step, the crew-member can take notes or start a voice or video conversation with a ground expert to troubleshoot a problem or ask for support. He can also stream video to the ground from his head-mounted camera to improve remote operator situational awareness, i.e. let the ground-based experts look over the shoulder of the astronaut. Multi-media, documents and 3D content associated with the procedure can be called upon at any time and rendered on either the head mounted or mobile displays. In more complex scenarios where both the crew and ground teams need to execute parts of the same procedure, the system allows shared control of procedures between these teams.

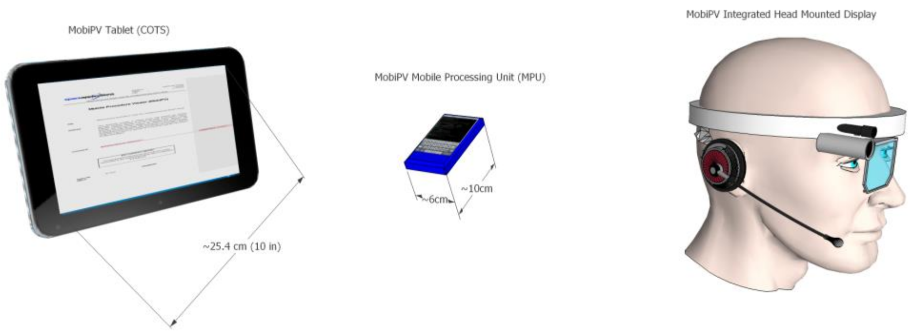


Figure 5. Initial mobiPV concept with an optical see-through HMD and customized mobile processing unit

In the initial concept (Figure 5), mobiPV was intended as a customised wearable computer connected to a monocular, optical-see through[3] head mounted display (HMD), an audio headset and a head-worn camera. The HMD is the main display, providing a heads-up view of procedures and procedure content. With augmented reality features in mind, an optical see-through HMD would enable the user to see virtual scenes and the real world seamlessly fused together. A secondary hand-held tablet was also envisioned to provide a large format display for specific content such as documents which are difficult to read on the HMD.

With this concept in mind, in March 2013 the mobiPV engineering team organized a user workshop at the European Astronaut Centre in Cologne to which ISS stakeholders including crew-members (experienced and in-training), flight controllers and training specialists were invited to provide feedback on various aspects of the mobiPV system. Areas of focus were operational scenarios and use cases, hardware and software features and walk-throughs of early graphic user-interface mockups designed for the HMD. Conversations with the end-users led to the discovery of a number of value-adding capabilities and features.

[3] An optical-see through HMD contains optics that allows computer-generated video signals to be overlaid on the visual field seen by the user. In contrast, a video-see through HMD renders computer generated video signals combined with real-world imaging from cameras.

End-user Feedback

During the workshop, users were walked through the key features of the system using GUI mockups. Each use case led to many in-depth discussions which are impossible to cover here. For brevity, we will only highlight the key discussions covering system usability and specific value-added capabilities that the end-users asked the design team to include.

Displays: The benefit of a monocular HMD as the primary display was recognised by all but a number of concerns were raised. Monocular displays require the user to re-focus one eye on the display and surroundings frequently. Even if the display is off-set from the normal locus of vision, ocular fatigue can occur quickly due to frequent changes in view direction while reading the display.

While a monocular optical-see through HMD is better suited for augmented reality, or where information is required periodically or with minimal display occupation, non see-through heads-up display were seen as more beneficial for reading procedures which are predominantly text or graphic-based. Some end-users even questioned the need for a HMD as a tablet could provide similar or even better display capabilities. To address challenges with the proposed displays, the use of other heads-up displays such as Google Glass in combination with smartphones displays are being investigated along with methods to organise and balance visual information presented to the user on these displays.

Speech Recognition: Navigation of the mobiPV graphic user interfaces is expected to be driven primarily by speech. One of the biggest concerns about this technology was the selection of a speech vocabulary that is simple to use while covering all GUI actions. For example, saying “next” provides the following procedure step or “create audio note” allows an audio note to be created. The vocabulary and the speech recognition system must be designed to handle and filter misinterpretations arising from unintentional utterances and background conversations. To avoid such errors, Apple and Google add keywords such as “Siri” or “Glass” to filter unintentional speech recognition. However, frequent repetitions of the same phrases can quickly become frustrating. The vocabulary must be easy to remember and not require user-manuals to use the system and on the ISS, it must be sensitive enough for accurate recognition with 50dBA of background noise which can increase to 70dBA in specific zones. Most importantly, recognition accuracy should be very high to prevent users from rejecting the technology. Hands-free system interaction is a very desirable feature in the ISS environment; designing a speech recognition system to work in the ISS environment however, is one of the biggest challenges in this project.

Note-taking: Feedback collected from the crew indicated that they frequently need to create notes, specifically photo and videos of their task environment and share it in real-time with the ground. Additionally, payloads are maintained by various crew-members on different missions and information during these processes needs to be transferred to future missions. For example, a critical parameter change or ad-hoc modification to a system should be available to other system users. Similarly, the same notes should be available to payload operators and instructors to evaluate if systems are functioning nominally or to gather crew feedback and improve future processes and procedures. The current process involves using cameras, transferring data to a hard drive and downloading it to the mission control rooms. This process can take several days. In our design, dynamic note sharing was considered for a future implementation but given its ability to improve task execution, it is now a priority feature.

Communication: Features such as text messaging were welcomed by the crew who felt that non-critical information could be exchanged by text instead of occupying the on-board voice system. Peer-to-peer audio-video communication using the mobiPV system was identified as useful for crew-members working on collaborative tasks in different parts of the station however it was clear that peer-to-peer communication between space and ground would be difficult to implement as the current infrastructure did not permit this capability. Whiteboarding was also appreciated particularly by flight controllers and operators who have no effective mechanism to troubleshoot problems remotely.

3D Rendering: Given that many on-board procedures involve 3D CAD content (see example in Figure 1), the benefits of this feature was apparent to the crew. The ability to visualize graphic detail on small form displays such as HMDs and smartphones was identified as a challenge. Rendering 3D content using mixed reality techniques was harder to describe without a demonstration of an example system, although it did pique the interest of the audience. Augmented reality was tested previously by one of the astronauts who highlighted the challenges involved in preparing the system for on-board operations (calibration, tracking, etc.).

Design and Human Factors Challenges

Designing user-centered systems for human spaceflight such as mobiPV presents a number of unique challenges. References and benchmarks are few and far in-between. Space engineering processes impose a number of safety and security restrictions which means solutions that are taken for granted on the ground need to be tailored to the ISS environment. We list some of the key challenges below.

User Groups: In comparison with ground-based studies where user groups are very large, gathering crew feedback early in the design phases is a challenge as crew-members have busy schedules and cannot always dedicate their time to specific activities. Feedback can vary due to mission experiences, socio-technical and cultural backgrounds. For mobiPV, the involvement of crew, both experienced and in-training across the project phases from requirements to design and testing has been made a priority and up until now, four crew-members are actively supporting the design effort.

Wearable Hardware: A tablet, smartphone or HMD for use in space will most likely require its hardware to be re-engineered. Internal batteries may need to be replaced by external qualified batteries, which can result in external holders and wiring (therefore electrical hazards). Displays can shatter; a major risk to the crew in the confined ISS environment and these need to be protected and can affect the optical quality of the displays. Hygiene requirements for a single system shared by the crew require a modular design allowing components to be removed, disposed or cleaned. Body worn hardware should not impede dexterity and mobility and safety standards mandate that body-worn displays (auditory, visual or even tactile) must not hinder crew performance. Given the limitations of resupplying the station with upgraded versions, careful thought is needed while considering the physical ergonomics of the system.

User Interfaces and Interaction: Due to the scarcity of benchmarks and limited user experience, UI and interaction design are at present exploratory. Assumptions need to be made on user needs and the first versions of the system are expected to remain simple, yet functional. In fact, the crew expects such simple user interfaces as the priorities are time and task performance and not always attractive visual design. Therefore, the UIs navigation and information architecture must be intuitive and easy to learn and most importantly, reliable.

Design Updates

Based on the end-user feedback received, the design philosophy of the mobiPV system has been reevaluated. Instead of developing custom solutions and reusing heritage technologies, the design team is now considering the use of off-the-shelf technology such as smartphones and tablets to provide the computing solution for the wearable system. Google's Glass is being considered as a fully integrated mobile computing solution (Figure 6, option 2); however the device is not yet available for consumers. As an alternative, the crew may be provided a wrist worn smartphone as the primary display with a head-mounted camera and audio device (Figure 6, Option 1). During the user evaluations that follow, hardware prototypes and mockups will be used to evaluate and select the best solution preferred by the crew.

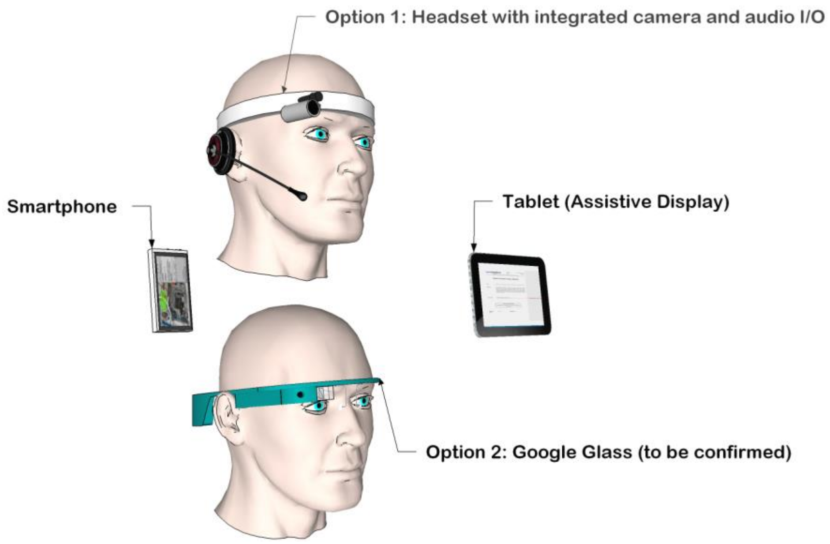


Figure 6. The updated mobiPV system concept relying on off-the-shelf technology

Conclusions

This article, the authors expect, has provided the reader with an overview of the challenges faced by the crew during task execution on spacecraft and the efforts being made by space agencies and industry in developing effective crew support solutions. The mobiPV system is at the time of this writing in its implementation phase and will soon be evaluated by the crew on-board the ISS on a mission planned in late 2015. This valuable opportunity will demonstrate path breaking task support and crew-ground collaboration paradigms and pave the way for mobiPV to be used in future missions as a permanent crew assistant.

Acknowledgements

The authors extend their thanks to the European Astronaut Corps, EuroCom and payload trainers at the European Astronaut Centre for their valuable contributions during this project. This work is funded by the European Space Technology and Research Centre (ESTEC) under contract 4000107062/12/NL/LvH and is executed by Space Applications Services NV, Belgium and Skytek, Ireland.

References

De Weerdt, D., Ilkovitz, M., Ambrus, R., Nevatia, Y., Martinez Oliviera, D., Arguello, L. (2010), *WEAR: WEarable Augmented Reality*, In 11th International Workshop on Simulation & EGSE facilities for Space Programmes (SESP 2010), Noordwijk, Netherlands: European Space Agency.

Naessens, Benedikt. (2011) *Portable Virtual AIT Visualizer*, Report PVAITV-WP4-TDP8-D1, Brussels, Belgium, Space Applications Services NV.

National Aeronautics and Space Administration (2010) *Operations Data File Standards*, Report SSP50253, Revision R, Houston, Texas, NASA Johnson Space Center.

Nevatia, Y., Chintamani, K., Meyer, T., Runge, A., & Fritz, N. (2011) *Computer Aided Medical Diagnosis and Surgical System: Towards Automated Medical Diagnosis for Long Term Space Missions*, In 11th Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA), Noordwijk, Netherlands: European Space Agency.

Wolff, M. (2013) *CRUISE Final Report*, Report ESO-IT-RP-0668, Friedrichshafen, Germany, EADS-ASTRIUM,

Exceptional performing screener project: methods and findings from the field study

Gerald D. Gibb[1], Karen D. Gibb[1], J. Michael Barrientos[2], Jesse D. Flint[1], Kelly Hale[1], & Darren Wilson[2]
[1]Design Interactive and [2]Department of Homeland Security
USA

Abstract

Earlier work (Gibb, et al., 2007) discussed methodological approaches for improving threat detection skills of the U.S. airport security workforce. This paper presents the unique techniques developed and findings from an extensive field data collection effort examining differences between exceptional performers (EPs; ≥97th percentile) in the screening workforce and other screeners. The study was conducted using operational X-ray machines and sixty test bags with 119 security officers from nine U.S. airports. The empirical methods used to identify EP participants and threat stimuli are described along with the cognitive task analytical techniques used for capturing the high level skills of the exceptional performers. The findings indicated that EPs use subtle cues, distinctive techniques and approaches, refined image analysis methods and recognition of anomalies that have led to the successful detection of simulated threats (IEDs [Improvised Explosive Device], guns, knives, and other weapons). The transfer of acquired knowledge, expertise, and techniques of EPs to the workforce at large was possible because the information was transformable into discrete training elements. An illustration of key concepts is included. Recently a prototype training program based on this analysis was developed and piloted.

Introduction

The U.S. Transportation Security Administration (TSA) and Department of Homeland Security (DHS) are working to improve aviation security. The performance of the Transportation Security Officer (TSO) workforce is paramount to meeting that mission. While major advances have been made in developing and deploying advanced technologies improving detection of threats targeting the air transportation system, enhancing the performance of the workforce has been more challenging. One key performance area is the capability of the workforce to effectively use X-ray technology to detect threat objects. Since X-ray technology is the principle method to screen carry-on belongings at U.S. airport checkpoints it is critical that workforce performance is maximized. This investigation focused on identifying what effective threat detection practices are employed at the individual operator level that contribute to threat detection. The two key facets were to empirically identify the subset of the workforce with exceptional threat detection

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

skills and use established methodologies to capture their skills, abilities, and techniques.

Over the past decade several field studies and investigations (Barrientos, et al., 2011) have demonstrated there are workforce personnel that consistently excel above their peers relative to Threat Image Projection (TIP) performance measures. TIP is a comprehensive, objective and quantitative system that assesses threat detection in the operational environment.

Research methods

Apparatus and participants

There were three major elements to successfully execute this study: a) selection of "discriminator" test stimuli (threat images), b) identification of personnel from three performance tiers, and c) design and crafting of test bags. The test stimuli was a limited threat library that included only discriminator TIP images as determined by an item analysis (Barrientos, et al., 2011). Discriminator TIP images are threats that exhibit among the lowest overall detection rates but EP screeners detect at significantly higher levels than the workforce. As there were two different Threat Image Ready X-ray Systems (TRX) deployed across U.S. security checkpoints, analytical work indicated two separate threat image sets were appropriate. The Smiths Detection test Threat Image Library contained 32 images whilst the Rapiscan test Threat Image Library included 44 images.

In determining how EP personnel differed from the workforce three study groups were established – EP (Exceptional Performers) who were defined as at or above the 95^{th} percentile; mid-tier performance where the mean performance is within the 45^{th} to 55^{th} percentile; and lowest-tier where the mean performance is at or below the 10^{th} percentile, over a six-month window based on TIP performance (Barrientos, et al.; 2011). The percentile bands represented a rank ordering and do not imply substandard or inadequate performance as officers are evaluated annually and meet accepted standards. EP Screener groups were identified for each of the Smiths Detection and Rapiscan TRX operator populations; one each representing large domestic or international airports and one representing regional airports. The distinction between large and small sites is based on the operational TIP:Bag ratios set on the TRX systems and not passenger throughput or physical size of the airport. Two mid-tier and two low-tier groups, one each for the Smiths Detection and Rapiscan TRX operators were also identified. Airport sites were selected on the basis of having the highest number of TSOs within a defined tier.

There were 60 test bags created using rollaboards, shoulder bags, duffles, and backpacks; representing moderately cluttered, typical passenger carry-on items. No liquids, gels, pastes, or aerosols greater than the limit of 3.3 ounces were included. No prohibited items (e.g. knives, flammables) were used. The bags contained sufficient light organic clutter (e.g. clothing, plush dolls) to mitigate the contents from shifting during transport.

Prior to finalizing the test bag set each bag was screened by the test team and the level of clutter adjusted to provide a challenging detection environment. Test bags were ‘themed’ in order to closely mimic typical passenger baggage. For example, “female bags” contained stereotypical gender-specific clothing, curling irons, shoes, and jewellery whereas a child’s bag would contain toys, smaller footwear. This was intentional to examine if TSOs were sensitive to relationships of articles within bags.

Specialized software and projection parameters were developed to support obtaining sufficient data for the protocol analysis (Ericsson & Simon, 1986) on the operational checkpoint TRX units. Therefore operational TIP system settings were not used as few threat stimuli would be projected across the test bag set for each participant. This innovative approach enabled participants to screen the bags using the system they normally worked. The following TRX settings and parameters were used:

R-V-R (Ratio-Variance-Random): This setting guides the frequency of TIP projections. It was set to 2-1-50 for the study. The setting generates one TIP for every two bags screened. The setting also provides for fifty percent of the TIPs as randomly projected. These settings allowed a TIP projection on every other bag screened, however, due to the random nature of the presentations, there are instances where no TIP is projected for 3, 4 or 5 bags and in other cases where a TIP is projected for every bag for 3, 4, 5 or more bags.

SDT (Secondary Decision Time): This parameter was set to 45 seconds. The SDT is the time the X-ray conveyor belt is stopped and the TIP button is selected before the TIP is considered a “Miss” and the feedback message is displayed. The maximum was selected to allow as much time as possible for the TSO to convey the cognitive, perceptual, decision, and analysis processes used screening each bag and for adequate data capture.

In some cases the TSO was talking through the review of the bag and “missed” the threat even though they had detected the threat and were talking about it, including positive identification of the location and the distinguishing characteristics when the 45 second time limit expired. In cases where there was no TIP presented and therefore no threat in the bag, the TSO continued verbalization for a time greater than 45 seconds until prompted by the facilitator to move the belt for the next bag.

Threat Category Distribution: These settings determine the projection distribution among the four threat categories. The settings differed between the Smiths Detection and Rapiscan systems to approximate the distribution of threats by threat category of the image libraries used.

Operator Logon ID and Password: A unique operator logon (e.g. 77777777) and User Name (EP Screener) was established to maintain anonymity.

Knowledge elicitation methods

Cognitive engineering, cognitive task analysis (CTA), and knowledge elicitation methodologies have advanced substantially (Hoffman, et al., 1995; Woods, 1993;

Salas & Klein, 2001; Someren, et al., 1994) since Klein, et al.'s ground-breaking study (1986). A history on the early stages of naturalistic decision making (an application of these methodologies specific to problem analysis and decision processes) is found in Lipshitz, et al. (2001). Hoffman, et al. (1998) in their examination of CTA cite numerous examples of successful applications in critical care, systems analysis, medicine, and military operations. These approaches demonstrated that highly refined skills and knowledge acquired by experts through experience can be identified and elicited. Moreover, the techniques are superior to conventional interviews, surveys, task analyses, and observational methods when the desired information is based on cognitive processes as opposed to behavioural practices (Dubois, 2002).

Approaches that have attempted to garner knowledge that relied on experts recalling past events, through a multitude of techniques, have often been met with failure to develop information in sufficient detail to develop training, build assessment tools, or establish performance criteria. Attaining vital detail requires eliciting implicit knowledge and transforming it into explicit information by exposing experts to the very tasks that require the use of the knowledge (Dubois, et al. 1995; Johnson & Johnson, 1987). Dubois (1995) maintains that verbalizations alone, if done outside the context of performing the actual job or tasks, are unlikely to generate the type of data desired. Consequently the most comprehensive and accurate data regarding cognitive processes is best attained as the expert is performing the task or a reasonable simulation of the same.

General methodology

All participants were briefed, completed background inventories, guided through a Think Aloud technique (Someren, et al., 1994) warm-up exercise, and performed the X-ray image screening task individually by scanning test bags one at a time. The flow of test bags was controlled so that only one bag image would display at a time and presented in the same sequence to each participant.

Screening each test bag consisted of two different stages characterized by the type of interaction between test facilitators and participant. The first stage was the onset of the bag image scrolling onto the display and terminated when the TIP system generated a feedback message. If no feedback was generated (Clear Bag), the facilitator urged the TSO to continue to the next bag. If the TSO pressed the TIP button before 45-seconds, indicating a threat was possibly detected, a feedback message appeared and this terminated the initial stage. During this stage the facilitator's primary function was to encourage the participant's verbalization of cognitive, perceptual and decision processes. This is the Think Aloud technique and resembles one verbalizing their thoughts as if talking to themselves. For example, a verbal protocol fragment might be "That orange area looks dark; I don't like it; I'll strip out organics and see if that wire leads from the cell phone to it". The facilitator did not probe for explanations. The technique is strictly having the participant verbalize thought processes. The facilitator called out each image analysis tool to the data recorders to reduce participant workload.

The second stage began as the TIP feedback message appeared and varied dependent on: a) whether a threat was detected, b) the nature of the feedback message (hit or miss), and c) content of the verbal protocols during the initial 45-seconds. Several cognitive engineering and knowledge elicitation techniques were employed by the facilitator during this stage and was characterized by direct interaction between facilitator and participant using specific probes, Teachback technique, or demonstration by the TSO.

Teachback and Critical Incident Technique were the primary methods and were specific probes targeted at TSO actions, cognitive processes or decisions made while screening the bag. The focus was to isolate the cues, anomalies, and threat characteristics that were perceived while viewing an image, the image analysis tools used and the purpose for using them. The verbal protocols, responses to probes, and interaction during Teachback dialogs were recorded independently by two trained observers. The data was maintained separately for each bag screened. Because of the randomness of TIP projections it was not uncommon to see the same threat on more than one occasion as a result of the limited size of the test set. Even though the same TIP was presented more than once it was presented in a different bag and in a different orientation,– thus constituting a unique presentation.

Results

Job and work experience

The mean job longevity for EP Screener personnel was 80.02 months (n = 66, SD = 29.80 months) and 71.30 months (n = 53, SD = 32.55 months) for mid-tier and low tier combined. These were not significant differences ($F_{1, 117} = 2.314$, $p < .13$). There were no significant differences in mean job longevity between Smiths Detection (mean months = 77.86, SD = 31.02, n = 35) and Rapiscan (mean months = 82.45, SD = 28.67, n = 31) EP Screeners ($F_{1, 64} = .39$, $p < .53$).

Information was collected on the participant’s previous employment prior to working as a Transportation Security Officer (TSO). This was an exploratory component to determine if there were trends regarding prior law enforcement, security or X-ray experience (e.g. food services or industrial inspection). Based on the findings of this limited sample the workforce is comprised of individuals who came from a broad range of previous job careers.

Threat detection performance

Threat detection metrics, although captured both manually and as part of the TIP automated database during the study, are not reported or analyzed. It became apparent that hits, misses, and false alarms in many cases may not have been accurate markers of threat detection performance. The source of error variance was two-fold: a) the knowledge elicitation and cognitive engineering techniques employed, and b) the validity of the metrics themselves.

Verbal protocols

Verbal protocols generate a substantial amount of information that require using segmenting, coding and analysis procedures (Ericsson & Simon; 1996). This study is no exception and produced a comparatively larger database than similar efforts using comparable techniques. This resulted from relatively high number of participants, task complexity, inclusion of two different systems, and number of events (bags screened). Each bag constituted an experimental trial and generated a unique set of verbal protocols. Consequently it is impractical to provide an exhaustive compilation of the verbal protocols. The protocol analyses were executed separately for each system. The data were sorted by threat category (e.g. IED, knife, etc.) and by specific threat image. This process facilitated organizing, segmenting and interpreting the vast number of verbal protocols. Each image analysis function used was coded sequentially as occurred such that they were lined with decision and perceptual processes. Encoding participant group enabled a comparison between groups across each specific threat type (see Table 1).

Whilst not practical to include an extensive summary of protocols from 6,000 events, illustrative verbal protocol from the three performance groups are offered to demonstrate that different performance groups use very different cues for threat detection and X-ray image analysis. Table 1 provides these comparisons along key areas of interest:

Table 1. Overall differences and sample verbal protocols across key dimensions

	Performance Tier Groups		
Dimension	**EP Screener**	**Mid Tier**	**Low Tier**
Cues	Thin black straight line indicative of a knife	9-volt battery is automatic bag check	Many organics are a problem
Major Difference	EPs respond to subtle, distinctive cues that are linked directly to threats whereas the other performance groups respond to generic cues		
Anomalies	Changes in the texture and density of organics occurs with underlying or overlapping explosive masses	Too much organics	Too much clutter
Major Difference	EPs respond using rule-based, specific visual elements while other performance tiers respond generically to global assessments such as colour		
Knowledge	Greenish-blue Polaroid batteries are rare and would not be carried without associated equipment	9-volt batteries are often found in TIPs	Too many batteries
Major Difference	EPs determine the validity and reason for an item in context whereas others apply general rules. The process is more analytical and hypothesis based.		
Analogues	Curling irons don’t have thin non-voltage carrying wires or components in handle	Camera is too green	Camera has no lens
Major Difference	EPs recognize when common items vary and have been tampered with or contain additional, un-expected materials; then isolate the cause for the discrepancy		
Appraisal	Assessing relationship among bag contents (e.g. Men don’t normally carry curling irons)	Shoes come in pairs	No or little evidence at determining relationships among articles in a bag

Major Difference	EPs assess the "theme" of a bag; detecting subtle deviations or anomalies from the norm		
Tools	Use of black/white tool to highlight small wires or "High/Low" incrementally to define the edges of organics	Black/white tool lightens image. Tool use generally to treat image holistically and remove colour.	Rigid, consistent use of the same pattern of tool use regardless of image dynamics
Major Difference	EPs select tools to achieve specific purposes related to the image or information desired; others use tools procedurally		
Strategies	Often systematic and methodical method; although individualized	Procedural applications	Often focus on specific objects/areas and not systematic approach
Major Difference	EPs have well defined strategies that are diverse and involve testing assumptions		
Approaches	Scanning as image scrolls on; noting anomalies and areas of concern – examine entire bag then focus on "hot spots"	Generic approaches as presence of batteries and organics or organics and electronics are considered a threat.	If cannot immediately clear a bag then suspect it
Major Difference	EPs use well defined and disciplined approaches.		
Techniques	Focus on items that don't "fit' with contents. Use of image tools to extract additional information to confirm or dismiss suspicions.	Typically a consistent use of the same tools in the same sequence	Typically a consistent use of the same tools in the same sequence
Major Difference	EPs have substantial command over the image analysis features, using them to analyze and test assumptions about the contents of bags.		

The synchronization of TIP databases with the verbal protocols provided a unique opportunity to examine differences between EP personnel and other participants to specific threat images (TIPs) because it was possible to isolate the verbal protocols in response to each TIP projection (Note: A TIP projection could occur in any test bag, consequently each TIP appeared in different bag environments). In some cases a threat was salient and un-obscured whilst in other instances it could be masked within dense clutter.

Table 2 illustrates the differences of perceptual cues or physical characteristics of an IED across different groups. Representative verbal protocols are provided for each group. The sample sizes are number of participants within each group that were presented with the image.

The most prominent feature of this IED was a curling iron containing a small metallic detonator in the handle. EPs focused on anomalies with the curling iron, emphasizing physical characteristics that drew comparisons with typical curling irons. The texture, hue, and presence of unexpected elements (including additional wiring) were physical cues that led to successful threat detection despite the entire IED was not visible. The query regarding presence of two curling irons in the bag indicates another anomaly as did the association of the curling iron to an unknown dense organic. These protocols are indicative of examining the suspect item in relationship to other bag contents.

Table 2. Differences among groups (threat characteristics) for an IED

EP (n = 20)	Mid Tier (n = 9)	Lowest Tier (n = 10)
▪ Extra wire coming out of curling iron ▪ Why two curling irons? ▪ Two different shapes in handle – doesn't belong ▪ Additional component in handle; det clear with B/W ▪ Texture of barrel different than normal ▪ Two metal things in handle; why two wires? ▪ Heavy organic next to curling iron ▪ Two many cords around it ▪ Darker handle than normally see, should not be blue inside ▪ Something concealed in handle	▪ Mass right here could be something, curling iron looks ok ▪ Just the shoes with massive shanks ▪ Automatic TIP; looks like pocket knife ▪ See curling iron; not a normal curling iron ▪ Don't know what to make of this bag; curling iron – nothing coming off of it ▪ Dark, indicator of being Disney gun – saw trigger real fast ▪ We have couple of pairs of shoes, curling iron, wires ▪ Straight TIP – the thing had a detonator in back	▪ There is something in shoe (*was not in shoe*) ▪ Looks like we have a mass here, looking for wires ▪ Looks ok to me, clear ▪ Too many wires going to nothing, could be a knife ▪ All electronics one side, all shoes on the other ▪ Some kind of hair thing; don't think there is a TIP ▪ Long metal bar going through here, plus whatever so dark in the bag

In comparison, only occasionally did mid-tier personnel recognize anomalies with the curling iron. Generally the curling iron is verbalized but dismissed as a potential threat. There was only one verbal protocol indicative of tampering. Low tier participants failed to identify anomalies and did not detect the threat. In nearly all cases attention to the curling iron was tangential.

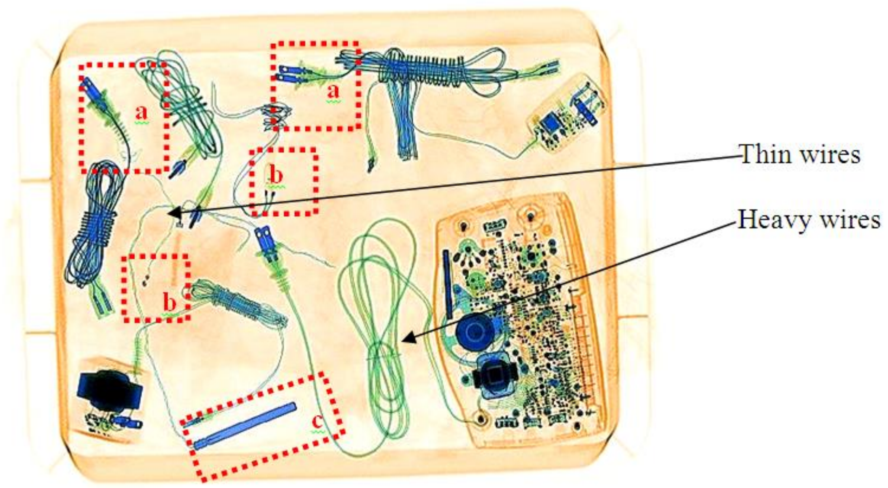


Figure 1. Illustration of visual cues used by EP Screeners

Differences between EP personnel and other groups were representative across the full range of threat objects. There was considerable consistency of the findings within groups. For example, specific knowledge of the physical cues and anomaly of

Polaroid batteries was almost universal among EPs whilst uncommon among other participants. Figure 1 provides an illustration of the unique knowledge and visual cues employed by EPs in detecting detonators. Visual cues such as: a) ordinary use wires usually terminate in plugs or jacks, b) fine wires that end in dots indicate a possible detonator, and c) presence of a, straight, slender cylinder are coupled with knowledge (e.g. thick wires are not capable of carrying the small amount of current needed for detonators) to successfully identify detonators. Figure 1 was prepared for the prototype training program. Whilst the differences are subtle and seemingly small, collectively they represent knowledge that improves threat detection.

A prototype IED training module for new hires was developed from a modest part of the protocol analysis results. The training used elements that discriminated between EPs and other participants. Training images used uncomplicated bags and threats that highlighted features used by EPs. Whilst the content was based on the protocol analysis, the research team built and captured images that increased the saliency of the cues and approaches suitable for an inexperienced group.

Summary

Although individual differences exist amongst EP Screeners in the use of image analysis tools, strategies, and routine approaches to screening, there is a genuine consistency distinguishing them from other screening officers. EP Screeners often began screening as the leading edge of the bag scrolled into view. They examine bags for relationships existing among the contents, identify anomalies, and focus on items that appear out of place. For example, recognition that electronics have only minimal, low density organics. They apply knowledge that 9-volt batteries are uncommon in typical electronics carried aboard aircraft.. They appreciate hair management products (hair dryers, curling irons) don't have dark components in the handles or include fine wires incapable of handling AC electrical current.

EP Screeners are attuned to subtle perceptual features and anomalies within bags (e.g. thin, straight black lines are not characteristic of most carry-on items; typical organics are not dense and have varied textures; highly saturated orange hues without definitive shape is uncommon; multiple pairs of shoes are not miss-matched in size or vary significantly in colour or density). Perhaps one of the most illustrative examples of recognizing perceptual nuances is when an EP Screener, while screening a particularly cluttered test bag with considerable organics, detected a change in the texture and shading within an area of overlapping organic material. By incrementally changing the contrast of the image using the High/Low image analysis function the shape of a glove began to emerge. The officer changed the image using the Black/White (monochrome) functionality, removing colour, and the outline of the "glove" became pronounced against the background of other organics. After confirming the anomaly existed the officer questioned "why is only one glove packed?", "why would the glove appear stiff as if it something was inside it?" and "what is that small metal piece inside?" These perceptual and cognitive processes were representative of EP Screener expertise throughout the study and were identified successfully using the applied knowledge elicitation techniques.

Acknowledgment

This work was funded by the U.S. Department of Homeland Security.

References

Barrientos, M., Gibb, G.D., & Eldredge, D. (2011). *Test and Evaluation Report: Exceptional Performing Screeners Study - Analysis of the National Threat Image Projection Database Characterizing Transportation Security Officer and Threat Image Performance*, Washington, D.C., USA: U.S. Department of Homeland Security.

Dubois, D.A., Shalin, V. L., & Borman, W.C. (1995). *A Cognitively-Oriented Approach to Task Analysis and Test Development*. Arlington, VA, USA: Office of Naval Research, (Contract No. N00014-93-C-0199).

Dubois, D. (2002). Leveraging hidden expertise: Why, when and how to use cognitive task analysis. In D. Dubois (Ed.), *Creating, implementing, and managing effective training and development*, Jossey-Bass Publishing of Wiley Company, San Francisco, CA., USA.

Ericsson, K. & Simon, H. (1996). Protocol analysis: *Verbal reports as data*. MIT Press, Cambridge, MA., USA.

Gibb, G, Fischer, B. & Cabeca, B. (2007) Methodological approach to advancing airport screener X-ray threat detection skills. In D. de Waard, B. Hockey, P. Nickel, & K. Brookhuis, *Human Factors Issues in Complex System Performance* (pp. 443-454). Shaker Publishing , Maastricht, The Netherlands.

Hoffman, R. Shadbolt, N, Burton, A. & Klein, G. (1995). Eliciting knowledge from experts: A methodological analysis. *Organizational Behavior and Human Decision Processes, 62*, 129 – 158.

Hoffman, R. Crandall, B. & Shadbolt, N. (1998). Use of the Critical Decision Method to elicit expert knowledge: A case study in the methodology of Cognitive Task Analysis. *Human Factors, 40*, 254 – 276.

Johnson, L & Johnson, N. (1987). Knowledge elicitation involving Teachback interviewing, (91-108). In A. Kidd, (Ed.), *Knowledge Elicitation for Expert Systems: A Practical Handbook*, Plenum Press, New York, USA.

Klein, G.A., Calderwood, R. & Clinton-Cirocco, R. (1986). Rapid Decision Making on the Fire Ground. In Proceedings of the 30th Annual Meeting of the Human Factors Society, (pp. 576-580), Dayton, OH., USA.

Klein, G., Calderwood, R. & MacGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man, and Cybernetics, 19*, 462-472.

Lipshitz, R, Klein, G. Orasanu, J & Salas, E. (2001). Focus Article: Taking Stock of Naturalistic Decision Making. *Journal of Behavioral Decision Making, 14,* 331-352.

Salas, E. & Klein, G. (2001). Expertise and naturalistic decision making: An overview. In E. Salas & Klein, G. (Eds.), *Linking Expertise and Naturalistic Decision Making* (pp. 3-10). Lawrence Erlbaum Associates, Mahwah, NJ., USA.

Someren, W., Barnard, Y, & Sandberg, J. (1994). *The Think-Aloud Method: A Practical Guide to Modeling Cognitive Processes*, Academic Press, London, UK.

Woods, D. (1993). Process-tracing methods for the study of cognition outside of the experimental laboratory. In G. Klein, Orasanu, J, Calderwood, R. & Zsambok, C. (Eds.), *Decision Making in Action: Models and Methods*, Ablex Publishing Co., NJ, USA.

VOD-CA - Assisting Human Flight Performance and Situation Awareness in Lateral Deconflicting

Peer Manske, Frederik Meysel, & Moritz Boos
German Aerospace Center (DLR), Institute of Flight Guidance
Germany

Abstract

Velocity-obstacle based collision avoidance displays can provide the pilot with information about potential conflicts and assist manual avoidance manoeuvres. Lateral deconflicting assistance is a significant field in aviation research since it complements vertical conflict resolution provided by available collision avoidance systems (ACAS/TCAS). Avoiding conflicts in the horizontal plane increases the number of options for a more efficient deconfliction. VOD-CA (Velocity-Obstacle based Display for Collision Avoidance) is a display implementation developed by DLR to enhance pilot awareness of potential conflicts in velocity space and provide necessary information to derive lateral conflict resolving manoeuvres (Peinecke et al., 2013). It represents conflict geometries in a track-up velocity space, centred on own-ship speed and heading. In a first explorative study participants with flight simulation experience flew several scenarios using VOD-CA to avoid conflicts, while their situation awareness was assessed. The results indicate adequate situation awareness. Participants could stay sufficiently separated from other aircraft by choosing safe headings and speeds. It is discussed that some aspects of the display could be misinterpreted when confused with spatial hints. Nevertheless essential information could be derived to stay clear of potential conflicts.

Intoduction

Background

Midair collisions are still a serious safety threat in aviation today. Several layers of protection are in effect, which were established in history of aviation to keep aircraft safely separated during flight. The pilot of an aircraft is still required to maintain situation awareness, to observe the surrounding traffic at all times, obey right-of-way rules and take the necessary measures to avoid conflicts and prevent collisions. In standard cruise flight aircraft must stay separated several miles laterally and at least 2000 feet (1000 feet in reduced vertical separation minima RVSM) vertically. Air traffic control ensures these limits in Instrument Flight Rules (IFR) operations. Still, representing only one safety layer, IFR-separation may fail, which is why additional technological safety layers have been introduced, and the pilot is still required to observe surrounding traffic ("See and Avoid", Federal Aviation

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Administration, 1983), even when not flying under Visual Flight Rules (VFR). The display technology assessed in this work aims to support situation awareness for a human pilot, whether flying under IFR or VFR, in terms of recognizing conflicting traffic and, if separation loss is imminent, having adequate information to steer clear of a conflict.

In this paper a new display concept for self-separation and deconfliction is introduced. Also a study is presented in which the display's comprehensibility and pilots' situation awareness were tested.

Collision avoidance and self-separation

Separation of aircraft is a safety layer in aviation to keep aircraft far enough apart to leave time for controllers and pilots to communicate, assess and plan the situation. Cases of separation losses are called conflicts in the scope of this work. The risk of a collision is imminent, whichever protection layer, that was in place to ensure separation, failed. The pilot of an aircraft is required by law to observe the surrounding traffic at all times, obey right-of-way rules and take the necessary measures to avoid conflicts and prevent mid-air collisions. This principle is known as the See and Avoid-rule in manned aviation (Federal Aviation Administration, 1983). Without the assistance of technological devices and support functions, the pilot has to watch surrounding traffic with the naked eye. There are limits to See and Avoid however, given by the atmospheric conditions and human perception limits, e.g., the resolution of the human eye, attention span, alertness and the ability to create a mental representation of the current traffic scenario. Many incidents took place, even in the best of conceivable viewing conditions (Hobbs, 1991).

Many ways were sought to assure aviation safety even in adverse conditions, including but not limited to operational measures, procedures and air navigation services. Technical assistance functions were invented to alert the pilot in case that all other safety layers are failing and an imminent collision requires immediate action.

The invention of the Airborne Collision Avoidance System (ACAS) and the introduction of the Traffic Alert and Collision Avoidance System (TCAS; as a concrete implementation) in the 80s of the last century posed a milestone in airborne collision avoidance systems. With all aircraft TCAS-equipped, it became possible to collect information on aircraft nearby, even without visual confirmation. Furthermore, through its sophisticated logic, TCAS is not only alerting involved airspace users of an imminent conflict, it also resolves the conflict by negotiating the respective manoeuvres aboard both aircraft. Thus, the pilot is informed about close traffic with traffic advisories (TA) and in situations where action has to be taken to resolve the conflict, pre-calculated (and automatically negotiated) resolution advisories (RA) are issued. TCAS contributed dramatically to a very high level of aviation safety since its introduction and it prevented many collisions where other safety layers failed. However, TCAS resolves conflicts only vertically, and it has a significant bearing error in localizing other aircraft (e.g. Law, 2005). Furthermore, TCAS is a last layer safety net, and never was intended to assist situation awareness. It was designed to inform the pilot and tell him how to resolve a dangerous situation.

It decides about collision avoidance manoeuvres without keeping the human, neither air traffic control operator, nor the pilot, in the loop. In cases where pilots tried to extend their involvement in conflict resolution based on TCAS, dangerous situations were produced (Law, 2005). Although it is technically possible for the pilot today to see surrounding traffic on a TCAS-display the information provided is coarse, and reflects only the current positions of all surrounding aircraft. Moreover, the display may be mistaken for a radar display, which shows dynamic events in front of a static background. The user might suspect, that he observes the absolute movement of an intruding airplane, where he in fact sees only relative movement. This is due to the fact, that the display is centred on the own ship position, which is not static while in flight. In (Law, 2005), a situation is described, where the pilot had the impression, the traffic was steering directly into him, and the controller was about to direct him right into the traffic, where in fact both airplanes were flying almost orthogonally and the controller chose the right manoeuvre to steer immediately clear of conflict. Lessons learned from this event were operational adjustments and the procedural instruction that “the TCAS traffic display must not be used for self-separation” (Law, p. 4).

Still, the possible benefit of additional information to assist the correct prediction of the outcome of a conflict remains. In remote oceanic flight for example, TCAS is carefully used to allow for more effective climb manoeuvres, adding local TCAS information to the rather vague positional information air navigation services can provide from the distance. Displays that visualise traffic positions on a moving map are not adequate for self-separation, but even dangerous. But other ways of visualization may well prove beneficial for situation awareness and conflict avoidance. For example, displays showing conflicting velocities of other aircraft with the own ship velocity could assist the pilot to foresee a separation loss as well as an imminent collision and help to avoid it early. Such a display could add significantly to aviation safety and efficiency at the same time, e.g., by showing horizontal conflict solutions. Early information about a future conflict in cruise flight would allow for slight adjustments of the speed or heading, which is referred to as lateral deconflicting, to stay safely separated and clear of conflict without leaving the flight level.

Velocity-obstacle based displays are developed to provide such information in a different visualization concept. Velocity obstacles are a representation of velocities within a velocity space and thus can represent a conflict (detailed description below). The concept of velocity obstacles (VO) is known for several decades and was extensively researched in robotics, path-planning, collision free movement (see Peinecke et al., 2013). However, in recent years, its possible application in traffic and aviation scenarios has come into scope. In recent years, velocity space based display prototypes have been proposed, e.g., display studies concerning lateral deconflicting were presented by Technical University Delft (NL) showing velocity obstacles in a spatial navigation context (e.g. Van Dam, Mulder, & Van Paassen, 2007). However, the Delft displays overlay spatial position information with the velocity space, which may bear a risk of confusing both information and misinterpretation, a problem that lead to the banning of TCAS for self-separation. The pilot should not be easily allowed to confuse the indication of own ship position with a very different indication of own ship velocity in a completely different space. Furthermore a velocity obstacle could be mistaken for a spatial obstacle and lead to an impulse to outmanoeuvre it, e.g., flying around the assumed dangerous area whereas adjusting the velocity would have been correct. The VOD-CA display assessed in this work was designed as a stand-alone additional system and not as a feature of existing displays. It strictly does not contain any spatial position reference to overcome the danger of misinterpretation.

VOD-CA

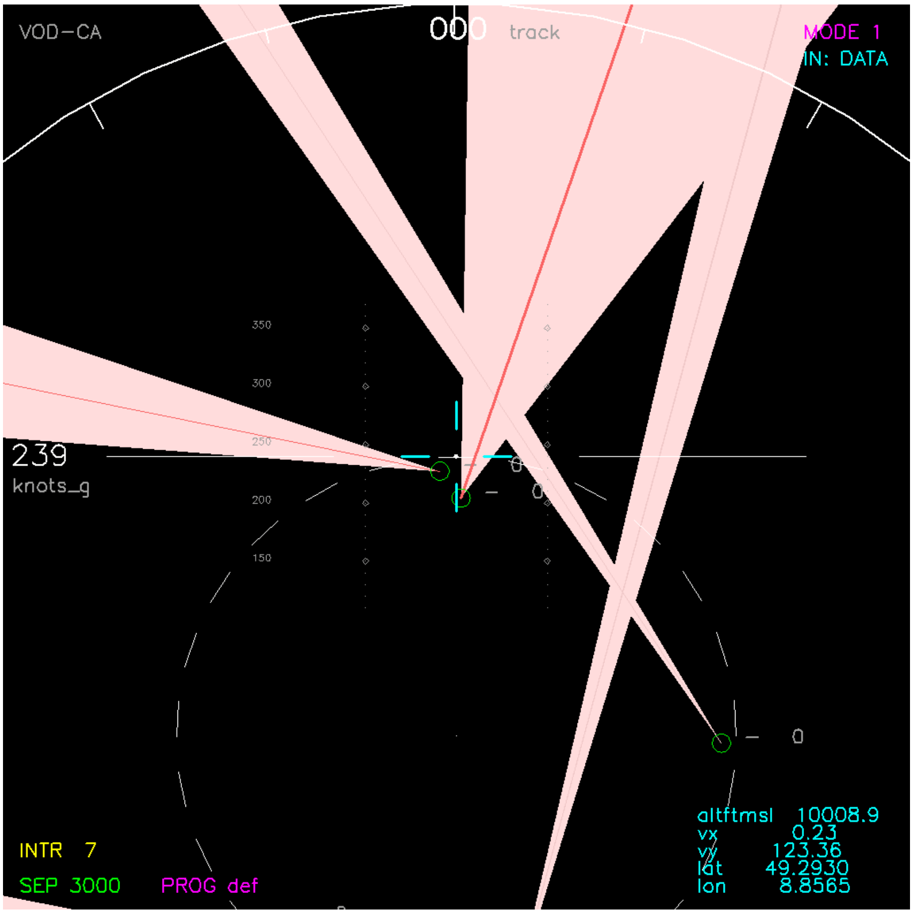


Figure 1. Velocity-Obstacle based Display for Collision Avoidance (VOD-CA)

The VOD-CA display implementation for collision avoidance has a strong focus on velocity space visualization. The main part of the display depicts the velocity space as seen from above. Velocity obstacles are drawn into that space. The y-axis of the display relates to the forward speed (“track”), the x-axis to speeds orthogonal to the current movement direction. Since the display is designed to assist pilots in their awareness and assessment of the current and future situation, the velocity space is always rotated such that its y-speed is aligned with the own-ship movement, similar to a track-up display for navigation. Furthermore, the velocity space is continuously centred on the own ship speed.

The display shows track (course over ground) and speed information similar to the established appearance in the common flight displays (see Figure 1). Acceleration of the own ship by thrust input shifts the velocity space up and down while a direction change rotates the display. Note, that there is no spatial reference to the own ship position in this display. The cross-hair always remains in the middle of the display. The velocity obstacles (VOs) are drawn into the space as triangular shapes marking areas of speed and heading that lead to future conflicts if the dot which marks own ship velocity remains within one of these areas. The tip of a VO marks an intruder aircraft's velocity, i.e., from the distance of the cross-hair to the pointy ends of the VOs, the relative velocity can be assessed. As the VO edges are the tangents to the other aircraft's separation zone the VO of nearer planes have a broader angle than those of distant aircraft. Again, this is only indirectly conveying information of the intruder aircraft positions.

The nature of a velocity space implies that it features speeds and directions of movement, not spatial relationships. They are of course implicitly contained, as much as relative movement is implicitly conveyed by a spatial display updating in time. Both paradigms however allow for immediate understanding of one of these aspects and a more trying determination of the other. VOD-CA explicitly is designed to provide a velocity space representation to see conflicting speeds instantaneously and is not meant to replace existing spatial displays.

Research question

It is assumed that pilots can use VOD-CA to stay safely separated from other aircraft. That requires adequate situation awareness. According to Endsley (1995) there are three levels: first the "Perception Of Elements In Current Situation", secondly the "Comprehension Of Current Situation" and the "Projection Of Future Status" at last (p. 35). As the VOs contain a representation of current and future state it should provide for situation awareness on all levels. Because the display has not been tested before, it is of interest if people actually understand when they are in conflict. Thus the question was whether people derive alternatives of reactions to avoid a conflict using VOD-CA and if they react accordingly.

Method

Setup

An explorative study was conducted to investigate people's perception of VOD-CA and their ability to fly safely using the display. Nineteen students mostly with a background in engineering (one female and 18 male) at a mean age of 25.2 years participated in the study. They had at least basic experience in flight simulations. In the study the participants flew at a desktop based simulator with two screens one above the other. A yoke, pedals and a thrust lever were available for flight control.

X-Plane 10 was used as the flight software and displayed on the top screen while VOD-CA received live data from X-Plane and was directing its output to the lower screen. Participants were provided with an outside view, the standard flight instruments in a glass cockpit and VOD-CA in addition. A Cirrus Vision SF50 jet, which is the standard aircraft in X-Plane, was used as flight model. Participants were not provided with a navigation display and had no spatial reference.

Scenarios

There were three tutorial scenarios and seven test scenarios. Every scenario was initialized in mid-air near Frankfurt am Main (Germany) travelling with a speed of 210 knots to the north. To reduce variance, the altitude of intruder and own aircraft was fixed at 10 000 feet so only lateral changes of track and speed were allowed. In the tutorial scenarios there was only one other aircraft that either stayed on course for the whole time or turned after a while. In both cases, an own ship reaction of the participant was necessary to prevent the conflict. The traffic in the following test scenarios consisted of six other aircraft with at least one future intruder. Every scenario lasted about two minutes.

Figure 2 shows an example scenario with the tracks of four intruders and the own aircraft. The aircraft pass without any separation loss.

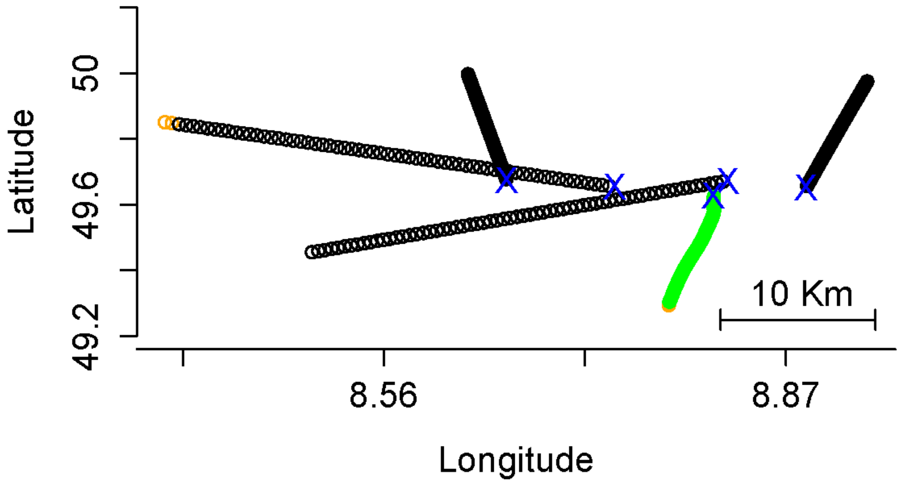


Figure 2. Example of a test scenario (showing only 4 intruder aircraft). Circles represent the locations of the aircraft at discrete time points, each five seconds apart from the last. The green colored tail illustrates the own ship track. The orange parts indicate the times of being in a conflict with another aircraft.

Experimental procedure

Participants familiarized themselves with the handling of the simulation for about five minutes with standard flight instrumentation. After that, they were briefed about the functionality of velocity-obstacles, followed by the four tutorial scenarios to get to know the new display. Participants were instructed to keep clear of conflicts and to stay separated from other aircraft by at least 3000 meters as the highest priority (3000 meters being the parameter used by VOD-CA to calculate velocity obstacles in the simulation).

The test scenarios were split into two parts. In the first four test scenarios participants' conflict-specific situation awareness was assessed using a SPAM technique (Situation Awareness Present Assessment Method; Durso & Dattel, 2004). In the beginning of every run the participants were asked whether they were currently in conflict and if so how they could deconflict. In case of no present conflict participants were asked to identify manoeuvres which would result in a conflict. The simulation was paused in the meantime. After a specific time, in which the participants were allowed to deconflict the situation if necessary, they were asked questions about direction, relative position and conflict potential of intruder aircraft, assessing further situation awareness and display comprehension. Additionally, participants were instructed to notify at the same moment they spotted a velocity change of one of the surrounding aircraft while flying. The second part of the test consisted of tree scenarios with an additional task. Participants should keep their direction to the north if possible, but prioritize staying clear of conflicts like before. Thus it was assured that the scenarios came out as intended. Every run lasted about two minutes. The overall sequence of scenarios was the same for each participant. The whole testing session took at least 90 minutes

Measures

For situation awareness measures, the answers to the SPAM questions were analysed. In the beginning of every run it was asked whether there was a conflict or not (conflict detection) thus assessing the participant comprehension of the situation. Additionally, participants should tell how to resolve the situation in case of a conflict (conflict resolution) and if there could be a future conflict (conflict anticipation) to assess their projection to the future.

For further display comprehension measures participants were asked about the relative position and track of one of the intruder aircraft. Additionally they should track other aircraft's velocity changes and tell the experimenter as soon as they spotted a change.

Besides the question measures, objective flight performance was measured. The number of conflicts and separation losses as well as the time in a conflict were counted. Moreover the deviation from the initial north track by degrees was calculated.

All measures dependent variables are shown in Table .

Table 1. Measures and dependent variables.

Measures	Dependent variables
Situation awareness	Conflict detection
	Conflict resolution
	Conflict anticipation
Display comprehension	Intruder relative position
	Intruder track
	Detected velocity change
Flight measures	Number of conflicts and separation losses
	Time in conflict
	Deviation from heading to the north

Results

The participants' situation awareness was assessed by SPAM questions. The results are plotted in the upper section of Figure 3. It is shown that every starting situation was perceived correctly. When the simulation was paused while flying the scenario participants answered questions assessing further comprehension of the display and the situation. They should tell the relative position of a certain intruder aircraft and its track which means the direction it was heading for. The relative position was named correctly for 97% of the cases. Only 86% of the questions regarding aircraft tracks were answered appropriately (see Figure 3). As a final display comprehension measure, participants were required to spot when an aircraft changed its direction or its speed. As Figure 2 shows at least 81% of the velocity changes were detected.

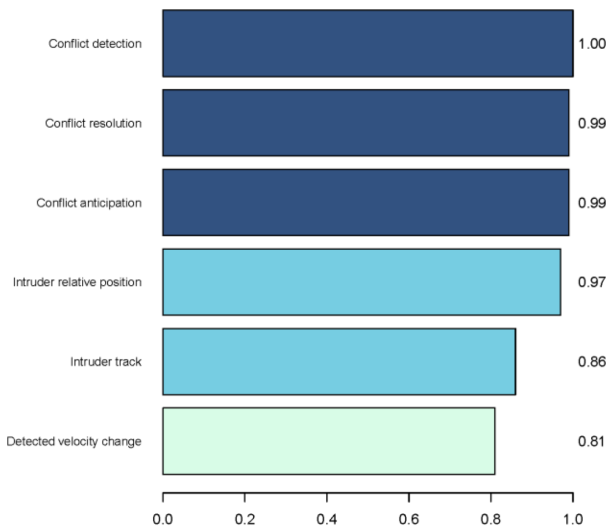


Figure 3. Situation awareness: Proportion of correct answers

There was also a learning effect. Participants got familiar with the display and answered more questions correctly in the later scenarios.

Moreover it was tested whether participants could fly the scenarios without separation loss. Only one participant went below the set separation minimum of 3000m only in the first test scenario. The rest of the time every participant kept the own aircraft well separated.

Participants chose different ways to deconflict. Figure 4 shows the tracks of aircraft within the same scenario flown by two participants showing two different solutions to deconflict. In the left solution the participant turned hard and flew around the conflicts whereas the right solution was more efficient due to less velocity changes but caused another conflict after one intruder aircraft turned.

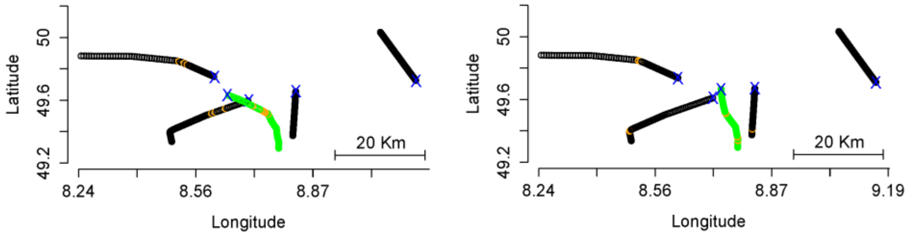


Figure 4. Two different solutions in one of the question scenarios (again showing only 4 intruder aircraft). Circles represent the locations of the aircraft at discrete time points, each five seconds apart from the last. The green colored tail illustrates the own ship track. The orange parts indicate the times of being in a conflict with another aircraft.

In average, there were 3.74 conflicts in the question scenarios and 5.19 in the scenarios with the additional task to keep to the north as long as practicable (see Table 2). However, participants stayed only 4.54 seconds (mean value) inside the conflict area. In comparison, it took them almost two seconds longer to resolve a potential conflict in the earlier scenarios. This might also arise from learning effects.

Table 2. Mean Number and duration of conflicts.

	Question-Scenarios		Additional task scenarios	
	M	*SD*	*M*	*SD*
Mean number of conflicts	**3.74**	3.62	**5.19**	4.70
Mean time in conflict (seconds)	**6.50**	4.25	**4.54**	4.30

At last it was compared how accurately participants stayed to the north. As a measure the mean deviation from zero degree was used. It turned out that the participants deviated by 24.58 degrees in the four questioning scenarios and slightly less in the task scenarios with 23.33 degrees. However, the scenarios differed in complexity so a comparison might be non-informative.

Discussion

The results show that participants can identify conflicts using VOD-CA and react appropriately so they keep themselves safely separated from other aircraft. Information about future hazards and how to avoid them can be derived from VOD-CA directly. In terms of deconflicting and self-separation VOD-CA supports adequate situation awareness on every level of Endsley's model (e.g. Endsley, 1995) indicated by the proportion of correct answers to the SPAM questions. Pilots spot and understand when they are in a conflict, they can identify their alternative manoeuvres to resolve the conflict and act accordingly. The latter can be inferred from the flight data. Only a single separation loss was registered in a first test run. Besides that it could be demonstrated that flying without any loss of separation is possible using VOD-CA.

Furthermore the finding of a learning effect indicates that participants can increasingly familiarize themselves with the display in a test period of 45 minutes. While participants encountered at least one conflict in each scenario, the number of conflicts and the imminence of a separation loss were highly specific to each scenario, making comparisons between scenarios difficult.

Although the display provides the user with an unfamiliar conflict representation, participants learn to use it in a short period of time for self-separation and collision avoidance. Participants were able to spot changes in the velocity of intruding aircraft and act accordingly. The detection of such velocity changes increased with familiarity of the display, e.g. more participants spotted the velocity changes in later scenarios.

Outlook

It was demonstrated that people are able to retrieve the relevant information from the display, and correctly answer the situation awareness questions, future studies could employ for example the reaction time till the relevant information is retrieved and changes are spotted. Another research question is whether the addition of VOD-CA to a cockpit already equipped with a TCAS display increases workload or pilots' head-down time due to attention capture. Also the assessment of users' usability experience could help to revise design aspects.

Some participants stated that they found it hard to derive position information, which they wished to have. As VOD-CA is not designed to display every aspect of the outside world - like exact information about aircraft's positions – but primarily to indicate future conflicts there might be a potential synergy of using it in addition to a spatial positioning display like TCAS. Further research is necessary to investigate how pilots conceive both a spatial and a VO display at once and if there could even be costs due to concurrent display concepts with different qualities of information.

All participants had flight simulator experience, but only 10 out of 19 participants had flying experience, and no professional pilots participated. Whether the display is well received by pilots is therefore a future research question.

Since VOD-CA provides the pilot with additional options for deconflicting, it has the potential to allow for more efficient changes in flight path, as well as more flexible conflict avoidance.

References

Durso, F.T., & Dattel, A.R. (2004). SPAM: The real-time assessment of SA. In S. Banbury, & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp.137-154). Hampshire: Ashgate Publishing Limited.

Endsley, M.R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, *37*, 32-64.

Federal Aviation Administration (1983). *PILOTS' ROLE IN COLLISION AVOIDANCE* (Advisory Circular No: 90-48C). Washington, D.C.

Hobbs, A. (2004). *Limitations of the See-and-Avoid Principle* (Research Report). Adelaide: Australian Transport Safety Bureau. [on-line]. Available: http://www.atsb.gov.au/media/32918/limit_see_avoid.pdf [13.10.2013].

Law, J. (2005). *Incorrect use of the TCAS traffic display* (ACAS II bulletin No. 6). Brussels: The European Organisation for the Safety of Air Navigation (EUROCONTROL). [on-line]. Available: http://www.eurocontrol.int/msa/gallery/content/public/documents/ACAS_Bulletin_6_disclaimer.pdf [13.10.2013].

Peinecke, N., Uijt de Haag, M., Meysel, F., Duan, P., Küppers, R., & Beernink, B. (2013). Testing a collision avoidance display with high-precision navigation. In D. D. Desjardins, & K. R. Sarma (Eds.), *Proc. SPIE 8736, Display Technologies and Applications for Defense, Security, and Avionics VII*.

Van Dam, S.B.J., Mulder, M., & Van Paassen, M.M. (2007). Airborne self-separation display with turn dynamics and intruder intent-information. IEEE International Conference on Systems, Man and Cybernetics, 2007 (pp.1445-1451). IEEE.

Teamwork: collaborative activities of French fighter pilots in joint operations

Anne lise Marchand
CReA – French Air Force Research Centre
France

Abstract

In literature tending to Team Cognition (Salas & Fiore, 2004) and in Activity Theory studies (Caroly, 2010) the definition of "team" is quite variable (Patel, Pettitt, & Wilson, 2012; Wildman, Thayer, Rosen, Salas, Mathieu, & Rayne, 2012), but it is generally associated to already stable teams (working collective) whom members are used to work together in their actual activities (collective activities). Nevertheless, some of these teams are composed by individuals who don't know one another (knotworking), working on reaching a common goal in a disrupted spatiotemporal environment (e.g. aircraft collaboration in military joint operations). This kind of team activity is called collaborative activity (Montferrat, Poirier, & Coppin, 2009). Using both quantitative and qualitative methodology, this study distinguishes clearly different kinds of team, working collective and knotworking, and different team activities, collaborative activities and collective activities.

Theoretical literature and operational context

The aim of this study is to distinguish collective activities from collaborative activities (two kinds of team activities) thanks to the difference between working collective and knotworking (two kinds of team). First of all, these concepts will be defined, then a specific methodology will be exposed to test two hypotheses, and finally the results will be discussed before concluding.

Teams are usually identified after their different general contexts: e.g. sports (Bourbousson, Poizat, Saury, & Sève, 2011), military (Fiore & Salas, 2004), medical (Cuvelier, 2011). Patel, Pettitt and Wilson (2012) specify that no consensus have been established yet about what a team actually is. A team is usually already stable and trained when it is studied by the Team Cognition's literature. Its performance is due to multiple factors: shared mental models (Van den Bossche, Gijselaers, Segers, Woltjer, & Kirschner, 2011), team situational awareness (Rentsch & Woehr, 2004) and the quality of communications (Paris, Salas, & Cannon-Bower, 2000). However, Team Cognition apprehends the team concept only in its cognitive dimensions (Salas & Cannon-Bowers, 2000).

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

The Activity Theory proposes a definition of the concept of team that expands to its activity. De facto, Caroly (2010) names it collective activity when a shared and coordinated job (collective work), is operated by a working collective. The working collective is a team of individuals who trust one another, who share values, common language and goals, and who negotiate and adjust their professional and operating rules. This notion of working collective has to be kept in mind in order to identify the stable and trained teams.

The concept of knotworking comes as an opposition to the concept of working collective. Knotworking treats about more temporary and less stable teams. It is described by Engeström (2008) as “the active contribution of combined persons and artefacts in constant reorganization and operating according to wide temporary paths, and that are widely distributed in space. (...) This notion refers to the orchestration of a collaborative performance, distributed and partly improvised between actors or activity systems that are in other respects weakly connected to one another” (our translation). One of the knotworking’s peculiar aspects is that it can be temporary when restricted to one single event, but it can be stabilised if its members get to work often together on the same situations’ category.

The knotworking concept will be kept as a key to characterise occurring or emerging teams, as groups of individuals called up to joint operations for a common purpose: when they come from different countries or professions, individuals rarely share common mental models or situational awareness. Individuals also face difficulties to communicate intensively with their partners (restricted and secured communication medias). Likewise, they are barely able to talk about their action or professional rules, especially if they don’t speak the same native language. Though they efficiently cope with strong technical and management constraints in network-centric warfare[4] and in interoperability. In particular, they deal with the spatiotemporal splitting of their actions (e.g. buddy lasing[5] between a fighter and an Unmanned Aircraft System, i.e. a drone). These activities cannot be anymore conceived or managed separately but necessarily jointly (individuals, teams, units, armies, nations). Riley, Endsley, Bolstad, and Cuevas (2006) indicate that controlling the military drills of planification process depends on a distributed team activity. This kind of activity is identified as a collaborative activity (Montferrat, Poirier, & Coppin, 2009). Collaborative activities call up teams that achieve their mission even though they don’t have at their disposal the same advantages as the working collectives. Indeed, the working collective has specific advantages of reliability (De Terssac & Chabaud, 1990), for it is particularly able to deal with unexpected events (Grosjean, 2005). Unexpected events can be managed by means of:

- Enhancement of decisions by sharing information and insights (Cooke, Salas, Kiekel, & Bell, 2004; Patel et al., 2012)

---

[4] Military doctrine or theory of war. It seeks to translate an information advantage, enabled in part by information technology, into a competitive advantage through the robust networking of well-informed geographically dispersed forces

[5] Coordination of two aircraft to treat a ground target: one of the aircraft is responsible for pointing a laser at the target to guide a weapon, while the other must deliver this weapon.

- Creation of action areas in- and outside the potential leeway (Caroly, 2010)
- Possibility to re-define the rules to reduce the goal conflicts (Caroly & Clot, 2004)
- Better capacities for adaptation and flexibility (Cooke et al., 2004).

Thus, reaching the set goals despite a knotworking configuration in collaborative activity empirically implies that to be effective, knotworking in collaborative activity uses different resources from those used by the working collective in collective activity. As a result, collective activity and collaborative activity appear to be two different kinds of team activities.

The literature (e.g., Caroly, 2010; Hsu, Chang, Klein, & Jiang, 2011; Wildman et al., 2012) allows to list a set of six constitutive dimensions to be accorded to any team activity. If collaborative activity and collective activity are two different kinds of team activity, these dimensions have to be differently rated. Individuals who work on collaborative activity will differently rate their satisfaction about these dimensions than individuals who work on collective activity. Those dimensions have been categorised from literature and six semi-structured interviews of French Air Force pilots who had worked with an Unmanned Aircraft System on a joint operation[6]:

- The "joint action" includes the constitutive elements of the live-activity when realised jointly, complementarily and articulated between team members. Also, there are elements linked to the interdependence of each one's part in the action, to the temporal and/or spatial (non) co localisation, to the simultaneous and updated knowledge of the state of the system, to the Team SA, and to the shared goals.
- "System and technical" supports include all the aspects linked to the characteristics of the organisational system (technical support included) communication excepted. This item concerns the awareness of the nominal functioning of the system, the constraints of the environment, the most frequent storylines, the decision-making chain, and finally the limits, errors and unexpected obstacles of the system.
- The "shared identity" concerns all the elements that impact the composition of the team: collective memory and knowledge sharing, systems of affiliation, tutorship or training, frequency of shared experiences, shared cultures and values, types of relationship.
- "Communication" refers to all the elements that allow communication within the team. It includes interfaces and means to get or transmit information, the interaction modes, the communication chain, the information delivery and flow (depending on the technical means at disposal), the existence of a shared language, the dedicated media.
- The "shared repository" covers all elements related to the prescribed rules (norms, procedures, rules of engagement) as well as the factual activity (daily violations, temporary uses). This dimension also includes every possibility of team regulations: leeway, hot and cold regulations, collective reflexive activities, rules' negotiation, etc.

---

[6] Exploratory pre-study, realised in June 2012, unpublished

- “Knowing your partner” concerns aspects bond to the knowledge of the other partner, and to build a strong representation of his/her identity, cultures and values: his/her part, responsibilities, skills, attitudes, preferences, tendencies, hierarchy level, etc.

These six dimensions to be applied to any team activity are proposed in order to predict two points concerning the differences between *collaborative* and *collective activities*:

1. If collaborative and collective activities are two different kinds of team activity, satisfaction about those team activity’s dimensions will be rated differently whether in collaborative activities or in collective activities.

Furthermore, Caroly (2010) precises that the emergence of a working collective is necessarily based on three factors: Negotiation of operational and professional rules, acknowledgement of skills, and mutual trust. These factors are respectively related to the dimensions “shared repository”, “knowing your partner” and “shared identity”.

2. Thus, differences between collective and collaborative activities will be found in the three pre mentioned dimensions linked with the working collective. They will be rated as more satisfying in collective activity.

Method

The study was set up on Nancy Ochey Airbase. Twenty-five fighter Pilots and Weapon System Officers (or WSO) answered this survey. Participants were confronted to six scenarios of team activity. If they already had lived the proposed scenarios, they had to evaluate their satisfaction about the six dimensions presented above. They also had to give an evaluation of the general satisfaction for each experienced scenario. These evaluations were marked on Visual Analogue Scales (Figure 1) from “totally satisfied” (grade 0) to “absolutely unsatisfied” (grade 10). In this example, the participant gives a 1.5 to the quality of communications in the activity.

Figure 1. Visual Analogue Scale evaluation of a dimension.

Six scenarios were proposed: collaboration with 1/ an American Unmanned Aircraft System Predator, 2/ a French Unmanned Aircraft System Harfang, 3/ a Forward Air Controller or FAC[7], 4/ Ground Troup 5/ a Fighter Patrol Force, and 6/ a C2[8] type structure. Amongst these six scenarios, only one (Fighter Patrol Force) concerned a collective activity that took place in a similar context of Joint Operations. This peculiar scenario was inserted in the survey in order to compare collective and collaborative activities. The other five scenarios complied with the characteristics of collaborative activities, and were gathered in order to guarantee that each participant had already been involved in one or more of the collaborative activities' scenarios. The survey was given to the participants during a briefing that started with clear instructions about the six dimensions. Twenty-five forms were collected out of the fifty ones that were distributed during the briefing. In addition, semi-structured individual interviews were performed with eight fighter pilots or WSO who had worked in Joint Operations. The point was to qualitatively strengthen the data from the surveys about specific collaborative activities, such as buddy lasing with Unmanned Aircraft System.

Results

The five collaborative activities scenarios (Predator, Harfang, Ground troup, FAC and C2) were aggregated (averaged) in order to get one "collaborative activity" category with six dimensions and a general assessment. In this way, each participant is related to one collaborative activity category with six dimensions and one general assessment. As a result one data table for the collaborative activity and one data table for the collective activity (Fighter Patrol Force) were created. Subsequently the ratings concerning the common dimensions of each category (collaborative vs. collective) were compared. For example, the average of "joint action" dimension in the collaborative activity category was compared to the average of the "joint action" in the collective activity category using Sudent's Test.

The results (Table 1) show a significant difference between collective activity (mean = 3.95, SD = 2.03) and collaborative activity (mean = 2.94, SD = 1.35) in the "joint action" dimension, $t(22)= -2.45$, $p< ,05$. This one is more positively rated in collaborative activity. The difference is statistically significant for the "shared identity", $t(22)=2.61$, $p< ,05$, between collective activity (mean = 2.53, SD = 1.93) and collaborative activity (mean = 4.09, SD = 1.76). However, "shared identity" is rated as more satisfying in collective activity. As well, the difference is statistically significant for the "knowing your partner", $t(22)=3.20$, $p< ,01$ between collective activity (mean = 2.98, SD = 2.31) and collaborative activity (mean = 4.38, SD = 1.72). "Knowing your partner" is more positively rated in collective activity. These results indicate that those dimensions ("joint action", "knowing your partner" and "shared identity") are rated differently by the individuals according they are set in a collaborative or in a collective activity.

7 Man on ground, who guides Close Air Support mission
8 Command and control system within a military operation.

These results have to be connected to the general assessment of each activity that is significantly different, t(22)= -4.12, p< ,001, between collective activity (mean = 5.13, SD = 1.96) and collaborative activity (mean = 3.17, SD = 1.13). Four out of the seven rated dimensions show a significant difference between a *collective* and a *collaborative activity*.

Table 1. Compared averages of collective/collaborative activities.

		joint action	communication	shared identity	knowing your partner	shared repository	system and technical supports	general assessment
Collaborative activity	mean	2.94	4.44	4.38	4.08	2.72	3.54	3.17
	SD	1.35	3.91	1.72	1.76	1.2	1.64	1.13
Collective activity	mean	3.95	5.01	2.98	2.53	2.03	4.33	5.13
	SD	2.03	2.63	2.31	1.93	1.74	2.71	1.96
Student's T	*t*	-2.457	-0.629	3.204	2.61	1.6	-1.26	-4.12
	dl	22	22	22	22	22	22	22
	p	0.022	0.535	0.004	0.01	0.121	0.22	0.0004

In contrast, the dimensions: "communication" t(22)= -0.62, p ns, "shared repository" t(22)= 1.6, p ns, and "system and technical supports", t(22)= -1.26, p ns, are not rated as significantly different between collective and collaborative activities. The rate of satisfaction of these three dimensions (communication, shared repository, system and technical supports) cannot allow maintaining that there is a difference between collective and collaborative activity.

Among the different rated dimensions, three are linked to the existence of a working collective: "shared repository", "shared identity" and "knowing your partner". Two out of these dimensions, "shared identity" and "knowing your partner " are rated significantly different between collective and collaborative activity. Only "shared repository" is not significantly different.

- "Shared identity" *t*(22)=2.61, p< ,05 is rated more positively in a collective activity (mean = 2.53, SD = 1.93) than in a collaborative activity (mean = 4.09, SD = 1.76).

- "Knowing your partner" $t(22)=3.20$, $p< ,01$ is rated more positively in a collective activity (mean = 2.98, SD = 2.31) than in a collaborative activity (mean = 4.38, SD = 1.72) .
- "Shared repository" $t(22)= 1.6$; p ns is rated more positively in a collective activity (mean = 2,03, SD = 1,74) than in a collaborative activity (mean = 2,72, SD = 1,2).

These results indicate that a working collective is more built in collective activity than in collaborative activity, and implies a significant difference between the two kinds of team activities.

The qualitative analysis of the interview aimed to define more precisely which factors enhance or hinder collaborative activities of the fighter pilots and WSO in joint operations. Interviews were analysed by a classical content analysis (Krippendorff, 2004). After transcribing the eight interviews, transcripts were divided in meaning units: In all, two hundred and fifty-eight meaning units were identified and connected to different topics. In a quality aim, the defined encoding scheme consisted in an ecological split of the remarks. The selected categories were: communication in action, situation awareness (SA), complementarity of the missions between aircrafts, representation of the partner, technical capacities of the partner, to dispel a doubt, trainings, workload, opportunism missions, and information sharing between fighter crew. Each meaning unit could have been connected to several categories. According to its topics, each category were after that connected to the six dimensions of team activity already used in the survey. The results are presented from this final classification.

Concerning the meaning units about the "joint action", the participants mostly recall the technical complementarities of the aircrafts (for instance the drones' camera or the fighter's speed) for 12 out 258 statements, the complementarities of the shared SA (34/258 statements) and the issues of taking the leadership in low-normed situations or opportunistic missions (13/258).

"Communication" is revealed as a sine qua non condition for the feasibility and success of a collaborative activity. 98/258 statements in the interviews gave specifications about this process. These specifications concern the operative language and the spoken language, but also information sharing (40/258) in order to build a Team SA or to dispel a doubt (43/258).

The "shared identity" is hardly informed in the interviews. It is detected in the need of a shared operative language (7/258). This working collective item is rarely found in the collaborative activities participants were asked about.

Regarding the dimension "knowing your partner", interviews revealed that the pilots and WSO have a limited trust in their partners in collaborative activities (21/258). The skills of their partners are little known, called into question, and the participants even try to regulate those through their choice to arise their workload rather than sharing their tasks with their partners. All the meaning units associated to the dimension "knowing your partner" reveal quite a hesitation of the participants in trusting their partners, especially due to the lack of knowledge of their training, of their profile, or of their geographical situation (for the Drone System Operators). This hesitation was declared to be due to the impossibility to clearly identify the partner’s skills. Thus, two peculiar trends are to be held: on one hand, the dimension “knowing your partner” is closely related to the idea of trust ; on the other hand, this dimension is specific to the working collective and seems to be more limited in a collaborative activity.

Concerning the dimension “shared repository”, negotiating the rules and regulations are revealed in collaborative activity but they only concern the operational rules and not the professional rules (48/258). Three different strategies of regulation were identified between the partners of a collaborative activity: the first one consists of setting up a common briefing stripped back to the essentials if necessary during the mission. The second strategy consists of misusing the technical capacities of an Unmanned Aircraft System and turning it into a radio relay in low covered areas. And the third one consists of to rely on the piloting skills of some Drone System Operators: indeed, the French ones are mostly former pilots and consequently have a real knowledge of the fighters they work with. So three strategies of regulation exist for only operational rules and not for professional rules. Regulations of professional rules are a proof of the existence of a working collective. The lack of professional rules regulation here indicates that there is no working collective in collaborative activity.

The dimension “system and technical support” is also a genuine worry (59/258). The categorised meaning units essentially concern the commanding process, or the limits of the available technical tools.

Discussion

The first hypothesis in the present study assumed difference between collective and collaborative activities. By comparing the averages from the survey of the collaborative and collective activities on each side, it was observed that these averages are significantly different for three out of the six dimensions and that they are different for the general evaluation of the perceived satisfaction in a collective or collaborative activity. These joined statistic index show a difference between both kinds of team activity, which tends to support our first hypothesis.

Furthermore, the results obtained with the survey as well as with the interviews, reveal the existence of a team that is more built or stable in a collective activity than in a collaborative activity. Moreover, a continuum of emergence seems to exist, that would allow an emerging team (a knotworking) to stabilizes itself in a collective working depending on the circumstances. The existence of this continuum seems to point out a major difference between both team activities. Since the working collective only emerges in collaborative activities as a knotworking, and as the estimated activities were all successfully achieved, then we can assume that the collaborative activities lay on other factors than the working collective to be efficient. Further studies will aim to identify these factors.

These further studies will attempt to strengthen the results. So the declarative data collection will be completed by more objective indexes of performance or workload, especially within the different populations (Rafale crew, A400M crew, etc). The next study's sampling will be operated on a similar quantity of collaborative and collective activities, which will enable us to focus on another factor that was often mentioned during the interviews: trust. And finally, since all the dimensions are not significantly rated different, it will be necessary to refine their classification.

The next study will investigate more intensively on the emerging processes of the working collective. It will bring elements of response to the question asked by Wildman and al. (2012) on the dynamic construction of the team knowledge through the evolution of the shared mental models.

As a conclusion, it seems necessary to split and consider differently these two kinds of team activities, in a purpose of activity analysing as well as in a typical Synthetic Task Environments of the Team cognition. Developing these concepts will enable eventually to guarantee the efficient and non-antagonistic articulation of individual, collective, and collaborative activities based on broken spatio-temporal span (e.g. an Unmanned Aircraft System flying in a COMposite Air Operation).

From an operational point of view, these results are the first step for the conception of an in situ evaluation tool. This will evaluate the quality of the collaborative context; during missions briefings for instance, this tool would anticipate and identify contextually the elements in favour and those more "fragile" for the up-coming collaborative activity. In the military aviation system for instance, this very tool would help for a better training of the operators for complexes and risky activities. In a long-term view, the trainings and tools dedicated to joint operations would take advantage of these results.

References

Bourbousson, J., Poizat, G., Saury, J., & Sève, C. (2008). Caractérisation des modes de coordination interpersonnelle au sein d'une équipe de basket-ball. *@ctivités, 5,* 21-39.

Caroly, S. (2010). *Activité collective et réélaboration des règles: des enjeux pour la santé au travail.* Doctoral dissertation, Université Victor Segalen-Bordeaux II. Bordeaux, France.

Caroly, S., & Clot, Y. (2004). Du travail collectif au collectif de travail: développer des stratégies d'expérience: Autour de l'expérience et de sa validation. *Formation emploi, 88,* 43-55.

Cooke, N.J., Salas, E., Kiekel, P.A., & Bell, B. (2004). Advances in measuring team cognition. In E. Salas and S. Fiore (Eds), *Team cognition: Understanding the factors that drive process and performance* (pp. 83-106). Washington, D.C., USA: American psychological association.

Cuvelier, L. (2011). *De la gestion des risques à la gestion des ressources de l'activité. Etude de la résilience en anesthésie pédiatrique.* PhD thesis, CNAM. Paris, France.

De Terssac, G., & Chabaud, C. (1990). Référentiel opératif commun et fiabilité. In J. Leplat & G. De Terssac (eds). *Les facteurs humains de la fiabilité dans les systèmes complexes* (pp. 111-139). Marseille, France : Octarès Entreprises.

Engeström, Y. (2008). Quand le centre se dérobe: la notion de knotworking et ses promesses. *Sociologie du travail*, *50*, 303-330.

Fiore, S.M., & Salas, E. (2004). Why we need team cognition. In E. Salas et S. Fiore (Eds), *Team cognition: Understanding the factors that drive process and performance* (pp. 235-248). Washington, D.C., USA: American psychological association.

Grosjean, M. (2005). L'awareness à l'épreuve des activités dans les centres de coordination. *@ctivités*, *2*(1), 76-98.

Hsu, J.S., Chang, J.Y., Klein, G., & Jiang, J.J. (2011). Exploring the impact of team mental models on information utilization and project performance in system development. *International Journal of Project Management*, *29*, 1-12.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology.* Thousand Oaks, USA: Sage.

Montferrat, P., Poirier, F., & Coppin, G. (2009). Étude d'activité collaborative en situation opérationnelle. *MAJESTIC 2009's proceedings*, Avignon, France: MAJESTIC.

Paris, C.R., Salas, E., & Cannon-Bowers, J.A. (2000). Teamwork in multi-person systems: a review and analysis. *Ergonomics*, *43*, 1052-1075.

Patel, H., Pettitt, M., & Wilson, J.R. (2012). Factors of collaborative working: A framework for a collaboration model. *Applied Ergonomics*, *43*, 1-26.

Rentsch, J.R., & Woehr, D.J. (2004). Quantifying congruence in cognition: Social relations modelling and team member schema similarity. In E. Salas et S. Fiore (Eds), *Team cognition: Understanding the factors that drive process and performance* (pp. 11-31). Washington, D.C., USA: American psychological association.

Riley, J.M., Endsley, M.R., Bolstad, C.A., & Cuevas, H.M. (2006). Collaborative planning and situation awareness in Army command and control. *Ergonomics*, *49*, 1139-1153.

Salas, E.E., & Fiore, S.M. (2004). Why team cognition? An overview. In E. Salas and S. Fiore (Eds), *Team cognition: Understanding the factors that drive process and performance* (pp. 3-8). Washington, D.C., USA: American psychological association.

Van den Bossche, P., Gijselaers, W., Segers, M., Woltjer, G., & Kirschner, P. (2011). Team learning: building shared mental models. *Instructional Science*, *39*, 283-301.

Wildman, J.L., Thayer, A.L., Rosen, M.A., Salas, E., Mathieu, J.E., & Rayne, S.R. (2012). Task Types and Team-Level Attributes Synthesis of Team Classification Literature. *Human Resource Development Review*, *11*, 97-129.

The Intermodal Bike: Multi-modal integration of cycling mobility through product and process innovations in bicycle design

Francesca Tosi[1], Alessandra Rinaldi[1], Alessandro Belli[2]
[1]Department of Architecture, University of Florence
[2]Tecnologie Urbane
Italy

Abstract

The paper presents the early results of the UE-FP7 project "The Intermodal Bike". The research aim is to provide a super-compactable, super-lightweight folding bicycle as a realistic solution to graft the cycling mode onto the root of the public or private transportation systems. The folding bikes now on the international market reach weighs between 12-15 kg, with a variable footprint but occupying – when compacted – an average volume of about 100 litres. To encourage the use of this vehicle and to extend it to a larger number of users with different characteristics, the research project has set its goal in increasing as possible compactness and light weight, creating a bicycle with a volume when compacted of 20 lit*r*es and a weight of 6 kg max, ensuring stability and improving vehicle usability and efficiency, during the ride and in the phase of bike folding. To achieve this goal ergonomic and usability tests have been carried out. The tests allowed finding a posture that would ensure efficiency and comfort in the ride to as many users as possible. Parallel tests were made on the vehicle usability in the urban transport system and intermodal. The need for light weight has required special studies on the optimization of the vehicle's architecture and research on super-lightweight materials.

Introduction

The overall research – financed by the EU through the FP7 instrument – is concerned with an updating of the industria approach to the production of "folding bikes" or, as they are increasingly defined, "intermodal bikes", to stress the fact that their principal use is a multi-modal duty, where the bike and other means of transportation collaborate synergically to enhance the global efficiency of the urban movements.

We are driven by the idea that the synergy between cycling and public transportation systems (train, tram, subway, buses) or private (car, taxi) would increase the share of both modes in the total of urban and suburban trips, and act as a booster also for the private transportation.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Bicycles are non- or sparingly- motorized “minimal vehicles”, able to handle very efficiently the small scale individual movements, in the range of the short trips typical of capillary urban mobility. As efficient as it is, it loses however its appeal on longer commute distances, for a number of reasons. The symbiosis of the bicycle with public or private transportation, more efficient at the medium or long distances, could enhance each mode’s strong points.

The synergy between fine grain local mobility and long-haul capacity could lower the access threshold to both cycling and public transport, and make cycling easily and immediately intermodal, with no need of special normative or infrastructural provisions.

Bicycle intermodality could therefore constitute a “third pillar” of incentivation of the cycling mode in the urban mobility (an often stated EU priority in all of its publications on the subject of city mobility) along with the other two more established instruments at “policy” level (modal incentives and prohibitions) and “infrastructural” level (parking, bike lanes). The folding bike is the new frontier of the intermodal urban rider.

In leading countries, like Japan, Holland, Great Britain, Belgium, USA, and model cities including London, Paris and Brussels, to name a few, the folding bike is in growing use to get around the city, along with other intermodal public/private transport, as for instance the exchange parking lots. Along with bike-sharing and bike access to trains and the underground, the folding intermodal bike is an integral part of measures and strategies to address issues and challenges of urban transport problems (AA.VV., 2008).

The EU air quality (reduction of CO_2) and noise directives require specific actions in urban areas. Improving the urban integration of the bicycle, making it interact with other modes of transport, not only affects the environmental situation, but will also improve accessibility to urban centers, with the prospect of improving their living conditions. Intermodal transport and folding bikes are a practical solution for all situations where classic cycling is impossible or difficult: for those users who daily flock to urban centers from the peripheral areas, covering distances that would be difficult to cover with the bike alone; for those who live and work in areas not reached in a capillary way by means of public transport; for those who do not know where to leave their bicycles at night and prefer not to use it, discouraged by the continuing thefts; in hilly towns, where climbs and descents discourage bikers, or urban centers where it is difficult to find bicycle parking; or,in those cities where you cannot go on public transport with a city bike, not only because of crowding, but also for the type of platforms that make access impossible. For these and other cases, the folding bike is a real solution. These advantages, if properly communicated, could greatly increase the number of users willing to leave their cars at home and choose to move with a folding bike (Mapes, 2009).

Ergonomics work package goals: Usability, technology and design

Object of the Ergonomics Work Package of Intermodal Bike is to apply the “Ergonomics for Design” and “User Centered Design” principles to the two areas of methodological research on the usability of industrial products, and of experimental design (Tosi, 2005, 2006):

a) providing all necessary knowledge needed for the configuration of a folding bike characterized by the maximum of comfort and ergonomics;
b) evaluating and optimizing its usability, effective and perceived, in relation to the context of use of the product.

As mentioned above, the latest generation of folding bikes are revolutionizing the way to get around by bike in the great international cities. Equipped with a gearshift up to 6-8 speed, the folding bicycles on the market today weigh between 12 and 14 kg, and fold in a volume between 100-200 cubic decimeters (liters). The research project has set its goal in increasing as possible lightness and compactness.

The goal is creating a bicycle with a max weight of 6 kg and a volume of 30 liters when compacted, ensuring stability and improving vehicle usability and efficiency, during the ride and in the phase of bike folding.

Figure 1 Volume comparison of folded existing bicycles. "Bike Intermodal" is the red one. All bicycles, when open, are the same size

To achieve this goal ergonomic and usability tests have been carried out. We ran a survey through a multiple-choice questionnaire on a sample of experienced urban cyclists, found mainly through cyclist associations. In total, 200 respondents were experienced users of different age groups. According to the survey, the folding bikes are considered lightweight and compact enough, but are still perceived by most consumers as strenuous, unreliable and unstable. It is therefore necessary to improve not only the lightness and compactness of the vehicle when closed, all elements that would expand its use in general and in particular by weaker users as women, but also improve usability, comfort and pleasure in use, real and perceived.

Regarding task a) the objective was to identify the postural configuration to guarantee maximum pedaling efficiency, protection from stress effects, problems, or damage to the column, neck and arms, even over short distances, the efficacy and in the natural control of the vehicle, with the consequent increase in the perception of comfort and pleasure.The research has focused on aspects other than the traditional ergonomics or physical or related to human factors, extending the work to areas such as anthropometry, biomechanics of posture and movement (Pheasant, 1997; Tilley, 2002).

It was useful in particular to identify and analyse user needs especially related to physical factors and consequently to define the requirements of the product and the acceptability thresholds under which the product will be evaluated on an objective basis; that is, referring to quantifiable and measurable metrics. The method of ergonomics intervention was aimed at assessing the compatibility between the characteristics of the abilities and characteristics of the product in the usage context, and to ensure and increase both the level of welfare of individuals, that levels performance of the final product. The biomechanics and anthropometry were used as monitoring tools such as posture and information tools that can provide data and evaluation procedures necessary for the approach and the definition of possible design. Based on the posture, that we identified as the ‘ideal’ one, or rather on the range of optimal angles of joints involved in cycling, was designed the bicycle frame. Therefore we have defined an optimal “ergonomic triangle” (handlebar-seat-crank) so as to make the experience of the ride comfortable for the higher percentile of users as possible (from 5° to 90° percentile).

Obviously it was necessary to achieve the right balance between moving vehicle usability and its portability when closed. This has led to important decisions that affected the project in general. The choice of a bike “one size fits all”, as well as the the innovative folding system, didn’t allow the inclusion of an adjiustable handlebar, unlike for traditional bicycles, making it even more important to study an ergonomic and comfortable posture for a large number of users.

Regarding task b) the objective is to assess the usability of the bicycle in urban intermodal and intermodal transportation by other means. It was decided to include a sample of users, appropriately selected, of some folding bikes already on the market today, and to make empirical tests with them, in collaboration with ATAF, Florence urban transport company, and LPP, Ljubljana Transport Company, both partners of the project.

The analysis aims to achieve the following:

- deepening knowledge concerning potential users types;
- defining the context and the scenarios of use;
- identifying the problems you might encounter in bicycle use into an inter-modal system with public transport and private;

- defining the different needs between folding bicycle users and classic city bicycle users.

The method adopted is the direct user observation, developed with people appropriately selected and available to experiment with this type of transportation. Selected users are followed, interviewed and involved in weekly focus groups focused on bike usability.

The tests focus on different aspects of usability, such as time and ease of opening and closing operations, convenience in transportation by bus, tram, train and everything else, comfort perceived by the user in carrying the folded bike, in using to move, etc.

This phase of the research is more related to the User Centred Design. Its aim, in fact, is to create products tailored to the user needs, whose use is pleasant and satisfying, and able to guarantee the possibility of carrying out the actions required efficiently and effectively.

Postural dimensional requirements: Survey methodologies

In order to define dimensional and functional project requirements, two different activities were carried out in parallel, in order to collect and analyse data related to a selected users group.

The data were collected for cross-checks on the results obtained with the two different methods. The first task called ‘People and their bikes’ has resulted in a detection phase of the anthropometric data of “city bikes expert users”, or “bicycle as a transportation mean users”, appropriately selected according to five different ranges of height, and recognition of their bikes.

Selection is aimed, in particular, to obtain dimensional data of the relationship seat-handlebar-pedal proportional to the user’s stature. The analysis was conducted on a sample of 100 users. This phase, together with a questionnaire on the usability of the city bike subjected to the same users, has identified the postures adopted, more or less consciously, by users and considered by them comfortable, and ergonomic. For the collection of the user’s dimensional data to analysed bicycles we optimized a detection system with 3D digital photogrammetric techniques. The whole process has been automated as much as possible: the camera calibration procedure was fully automatic, as well as the orientation of images (thanks to the use of coded targets).

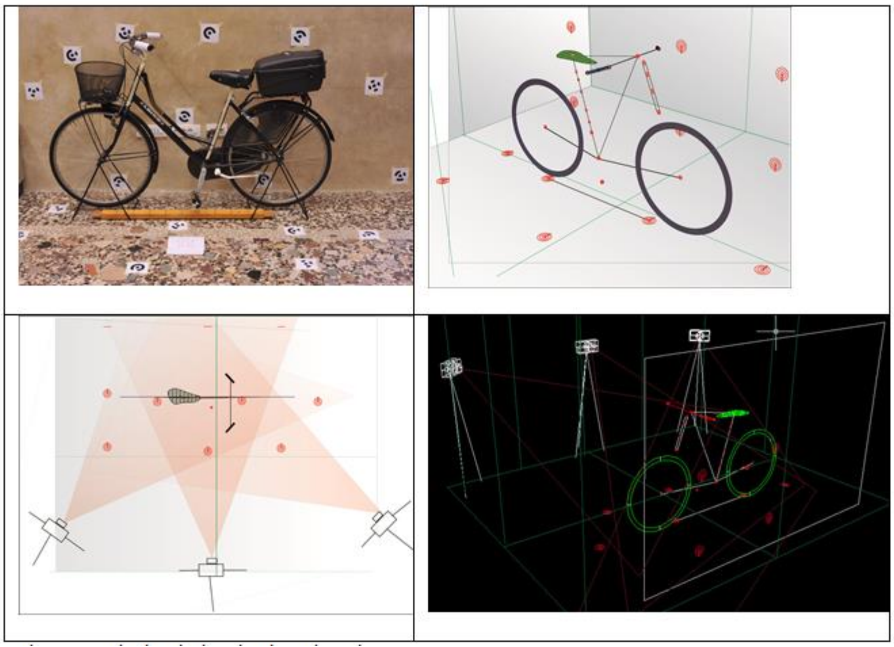


Figure 2-3. Bicyles during the detection phase, Figure 4-53D CAD processing of bicycles with digital photogrammetric system

The application of appropriate markers at strategic points has made it easier the phase of data graphic representation.

It was thus possible to represent point by point the geometry of each bike, each of which have been defined in three-dimensional coordinates. In particular, the points were found according to the relationship seat-handlebar-crank center, the center of the handlebar grips and the distance between the wheels. To verify the accuracy of acquisition method, were also made sample manual measurements. The collected data have high accuracy, with a tolerance of about ±10 mm.

A second research task, “Biomechanical analysis of posture and movement”, developed by the Occupational Medicine research group of the University of Florence is currently under completion.

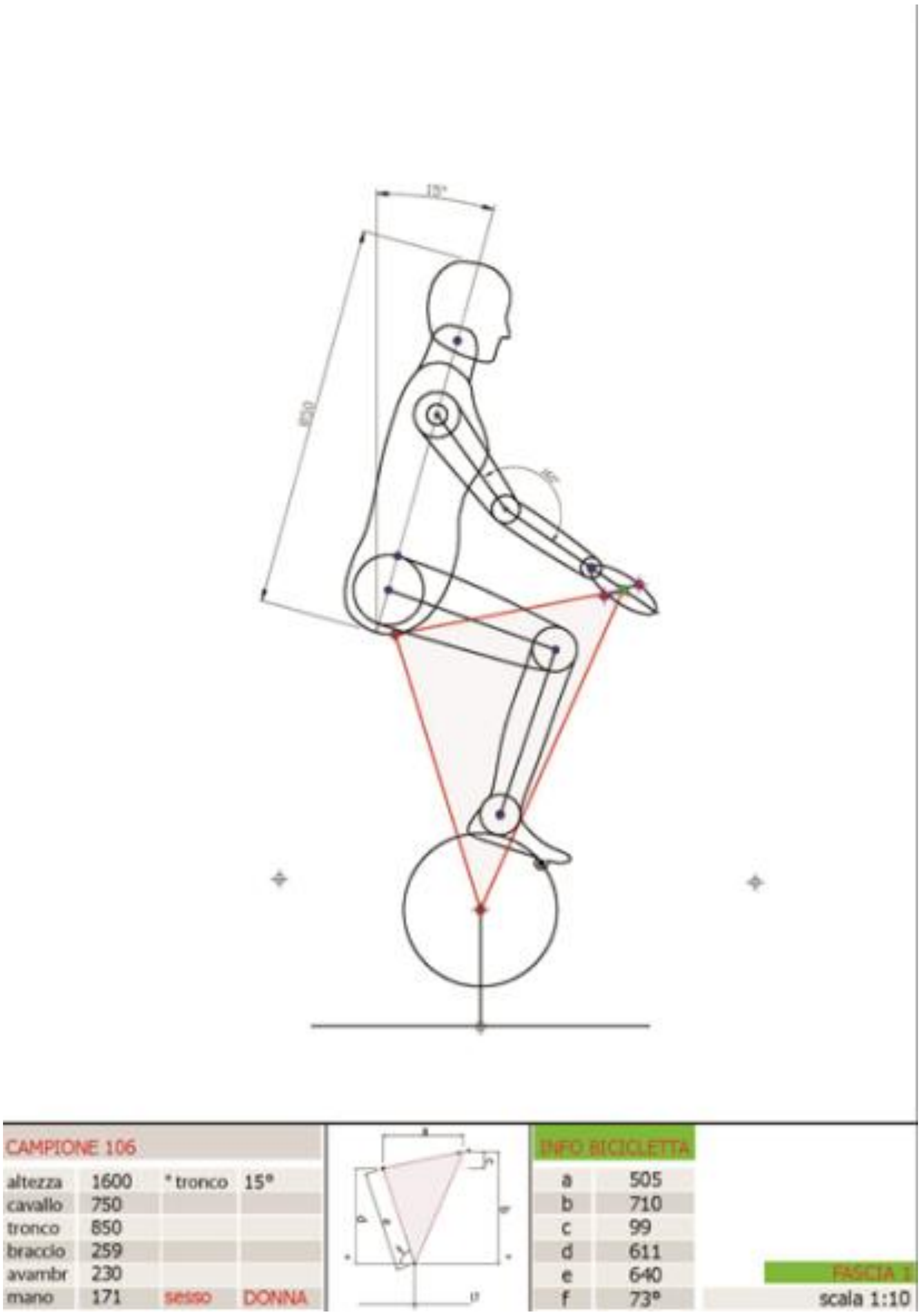
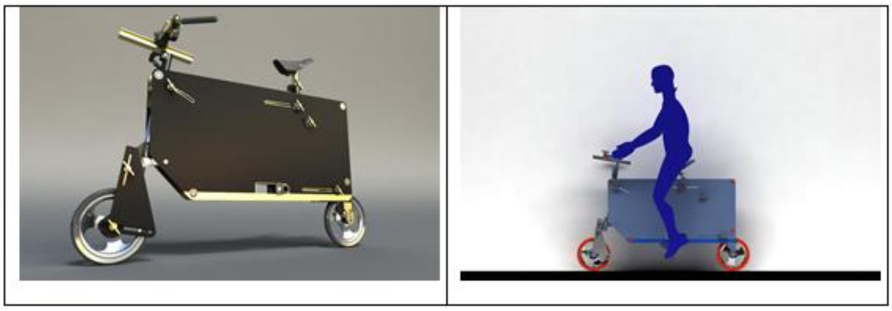


Figure 6. The data collected



Figures 7-8 Measuring model. This prototype has been used to test different postures. All the components are adjustable

.

Figures 9-10. Biomechanical survey of posture and movement

The data processed by our work group were compared and integrated with those developed by occupational medicine group, in order to define an integrate the evaluation of usability and biomechanics aspects.

Ergonomics and technology innovation: Design for light weight

The obvious thing to say about making something very lightweight is that it is necessary to put “less stuff” in it. In other words a lightweight object is also a more efficient object in the use of the materials.

There are several ways to reduce the material keeping constant the performance, depending on the relative importance of the structure’s stiffness and ultimate strength; i.e., if it is more important to control how much deformation a structure can have, or if the deformation is less important than the ability of the structure to absorb a lot of stress and rebound with no broken elements.

A bicycle falls in the first case. It must be designed for stiffness, as a “wobbly” frame feels insecure, makes the ride uncomfortable because the rider must use part of his or her energy to check continuously the posture, and the elastic work of the frame absorbs more rider energy dissipating it.

An ideal bicycle has a very stiff frame, moving to accurately chosen areas, outside a closed “hard cell”, the desired elastic elements for comfort and suspension. This happens for instance in all carbon-fiber modern frames.

The case of a “collapsing” bike frame is particularly difficult because the necessary joints break the material continuity of “monolithic” normal bike frames. The most common solution has been traditionally an “open” frame based on a large tube able to resist large bending and torcional stresses, that is the loading that any material has more difficulty dealing with. This translates directly into very heavy bicycles; as such tubes must necessarily have very robust sections.

Our project had the “light weight” as one of the two central requirements, and therefore we chose to follow a path common in bridge and large scale constructions but very seldom used in the bike industry: the space-frame, and to be more precise the pre-tensioned space frame (Frei & Bodo, 1996; Fuller, 1961; Heartney & Snelson, 2009; Motro, 2003).

A space-frame does not rely on torsion and bending to resist the external loads, but rather uses only “pure” tension and compression stresses, that is the stresses that any material has less problems with. The space-frame, and especially the type with pre-tensioned members – often termed, in certain cases, “tensegrity systems” – are in absolute terms the structures that have the least material per volume occupied or surface covered.

In our case the space-frame is particularly suited to realize an opening-closing structure, as a collection of rods and cables, all very similar in length, lends itself to be packed in a very dense configuration, with a minimum of “air” in between, obtaining therefore a very compact pack. Ultimately the bike frame is a sort of very stiff “cage” that offers the support to the cycling components as pedals, seat etc. but with a form that is quite different from the classic “diamond-frame”. In particular the seat-regulation system doesn’t have a “seat tube”, but can be designed to follow –within a certain range – a move on a line not bound to the frame. This helps in the design of a one-fits-many bicycle that must be adapted to bodies of different build.

Figure 11. Bike Intermodal: the space frame is a sort of tensegrety system.

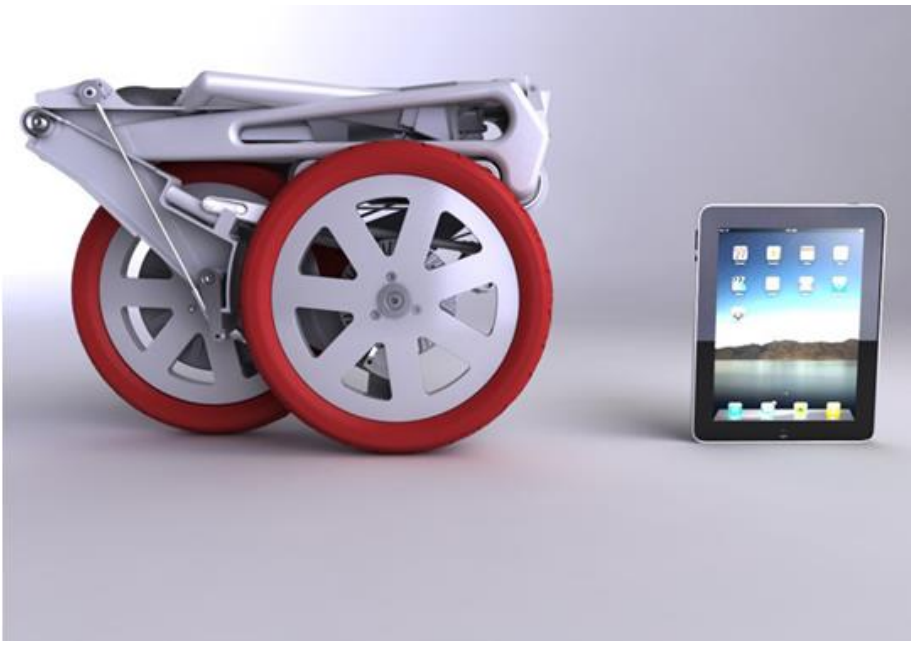


Figure 12. The folded bike compared with a tablet.

Conclusions

The ergonomic study recognizes the different "regional" riding styles, which are due to a number of factors, as industrial customary practices, social inertia, prevalent altimetric variability of the land and fashion, and points in the direction of a "corrected Dutch" posture as the most suitable for non-sport, daily urban commute of the type the users of an "intermodal bike" are expected to do.

"Corrected" refers to some early results of the field tests, pointing to a tendency to lean forward more as the subjects grow tired in the test effort. Discovering this tendency has been useful in the calibration of the "reference posture", given the peculiar frame and seat adjustment of Bike Intermodal.

The particular folding frame provides a type of seat height adjustment, where the length of the seat post moves along a direction slightly off-line with respect to the bottom bracket. This allows a different regulation of the seat-handlebar distance for tall and short statures, and allows adjusting for the differences emerged in the study.

Figure 13. Bike Intermodal compared with a folding bike on the current market

References

AA.VV. (2008). *Due ruote per il futuro*. *Prima Conferenza Nazionale della Bicicletta*. Portogruaro: Ediciclo.

Frei, O., & Bodo, R. (1996). *Finding Form: Towards an architecture of the minimal*. Stuttgart: Axel Menges.

Fuller, R.B. (1961). Tensegrity. *Portfolio and Art News Annual*, *4*, 112- 127.

Heartney, E., & Snelson, K. (2009). *Forces made visible*. Lenox, MA: Hard Press Editions.

Mapes, J. (2009). *Pedaling revolution: How cyclist are changing American cities*. Corvallis, OR: Oregon State University Press.
Motro, R. (2003). *Tensegrity: Structural systems for the future*. London: Hermes Science Publishing.
Pheasant, S. (1997). *Body space: Anthropometry, Ergonomics and the Design of Work*, (2nd ed.). London: Taylor & Francis.
Tilley, A.R. (Ed.) (2002). *The measure of man and woman: Human factors design*. Wiley & Sons Ltd.
Tosi, F. (2005). *Ergonomia, progetto, prodotto*. Milano: Franco Angeli.
Tosi, F. (2006). *Ergonomia e progetto*. Milano: Franco Angeli.

User requirements for supporting the accessible design process: Survey & user test results in the framework of VERITAS project

Leandro Guidotti[1], Caterina Calefato[1], Elisa Landini[1], Luca Minin[1], Romina Catani[2], & Monica Milani[2]
[1]RE:Lab srl, Reggio Emilia, Italy
[2]Indesit Company spa, Fabriano, Italy

Abstract

Older and disabled people are not just a tiny minority of the European population: although a lot of principles and standards for accessibility and universal design are available, adopting them during the design process is not always sufficient, since they do not provide developers with explicit guidelines. The aim of VERITAS EU-project is to develop an open framework for providing built-in accessibility support during all design stages. The application fields are: automotive, smart-living spaces, infotainment, healthcare and workplaces. In order to understand designers' needs, a State of Art analysis was undertaken regarding the existing approaches towards physical, cognitive, behavioural and psychological modelling of older and disabled people. A vast survey was carried out to define the industrial needs of designers and developers within the VERITAS project application sectors. It involved 217 questionnaires and 21 interviews. Survey findings point-out a gap in current design and development processes, indicating that accessibility is hardly considered in design due to the lack of knowledge and supportive tools. The VERITAS platform has been tested with designers of the smart-living space domain. Involved users were invited to design the HMI of a new oven and then to evaluate the usefulness, usability, and effectiveness of the VERITAS platform.

Introduction

Funded by European Commission in the Seventh Framework Programme for Accessible and Assistive ICT, VERITAS project (http://veritas-project.eu) aims to provide built-in accessibility support at all stages of realization chain of ICT and not ICT technologies in different fields of application (i.e. automotive, smart-living spaces, infotainment, healthcare and workplaces).

A survey among designers and developers from automotive, smart-living spaces, workplace, e-health and infotainment domains is presented in this paper. The survey was carried-out through interviews and also through web-based questionnaires administered to designers (Dilliman et al., 2001), aiming at better understanding the design and development process of products and services for people with disabilities as well as elderly. The collected data formed the basis to implement appropriate

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

tools and methodologies in order to increase awareness and needs for accessibility in the five selected areas in the framework of VERITAS project.

Survey: development and results

The survey involved designers of five different domains. In overall 217 designers were asked to fill-in questionnaires about their work, needs, competences and accessibility in design process. Also 21 face-to-face interviews were carried out (Table 1).

Table 12. Carried-out survey

	Automotive	Workplace	Smart living	Infotainment	Health care	Total
Questionnaire	47	40	44	40	46	**217**
Interviews	4	4	4	4	5	**21**
Total per area	**51**	**44**	**48**	**44**	**51**	**238**

Every survey was composed of seven groups of questions:

- *Personal data* **-** age, gender, educational background and field of work.
- *Job and professional knowledge* **-** details about working position, autonomy, team composition, skills and used tools.
- *Design process overall questions* **-** a scheme of a design process is shown and the participant has to describe how much it reflects the design process and the elements normally adopted by him/her. Then participants are asked to give details about the used tools and programming languages.
- *Accessibility* **-** items about usability, accessibility, simulation and virtual design environment.
- *Design, accessibility and your design process* **-** items about familiarity with accessibility guidelines or standards.
- *Accessibility and data collection* **-** items about how end-users may be involved in the design process and in which design phase.
- *Design process. VERITAS approach to design* **-** the traditional approach to design of application for disabled and old people is put in contrast to the design process proposed by VERITAS. Participants were asked which are the differences between the design approach used nowadays by them and the proposed one, which could be the best way to integrate VERITAS functionalities in the tool used by designers and finally in which design phase.

The gathered data deriving from participants' responses were analysed according to the domains in which designers were recruited for the survey. Each item was analysed independently. Descriptive statistical analyses were carried-out.

For each domain involved in VERITAS project (i.e. automotive, workplaces, smart-living spaces, infotainment, healthcare) survey results were analysed in order to highlight findings respect to the State of the Art in design process in the specific field of application. Then similarities and differences trough domains were pointed-out in order to fix designer requirements in the overall process and highlight lacks, restrictions and opportunities to include a technology aimed at improving accessible design for impaired and elderly people.

Sample details

Most of the participants were between 25 and 35 years old, postgraduate (sometimes PhD) and were engineers or work in the engineering field. The sample was equally spread among junior and senior designers and also with developers per each domain. Many participants had more than 15 years of experience. Participants were recruited within the R&D departments of partners involved in VERITAS project.

Main findings from survey

The most widely used computer coding and programming languages in all domains were: C; C++; C#; PHP, Java and JavaScript. The following table sums up survey results about participants' experience in accessibility, awareness of its meaning and the accessibility testing techniques adopted in their designs or products.

Table 13. Percentage of participants with experience in the different application areas

EXPERIENCE IN	Automotive	Workplace	Smart living	Infotainment	Health care
Usability	57.14%	63.83%	52.17%	80.00%	68.18%
Accessibility	37.14%	63.83%	47.83%	60.00%	68.18%
Simulation	77.14%	51.06%	30.43%	52.00%	45.45%
Virtual environment	62.86%	40.43%	26.09%	44.00%	18.18%

Table 14. Meaning of accessibility

MEANING	Automotive	Workplace	Smart living	Infotainment	Health care
Accessible web site or software applications	31.58%	46.81%	43.48%	65.38%	40.74%
Accessible buildings	36.84%	57.45%	65.22%	19.23%	33.33%
Assistive device	18.42%	34.04%	39.13%	30.77%	18.52%
Elevators	10.53%	12.77%	4.35%	3.85%	0%
Accessible for all	36.84%	59.57%	34.78%	23.08%	25.93%
Design for all	26.32%	44.68%	43.38%	11.54%	33.33%
Platform independent output device (i.e. mobiles, PDA)	10.53%	21.28%	17.39%	30.77%	11.11%
Accessibility of public transport	28.95%	29.79%	21.74%	15.38%	18.52%
Other	2.63%	4.26%	4.35%	0%	3.70%

Table 15. How the developed accessible applications are tested

	Automotive	Workplace	Smart living	Infotainment	Health care
With an accessibility assessment tool	5%	-	-	-	-
Involving the target users	12.5%	29.17%	34.78%	3.7%	48.18%
With expert evaluation	20%	37.5%	13.04%	14.81%	29.63%
With assistive technologies (i.e. screen readers)	-	-	13.04%	-	3.7%
With official guidelines	12.50%	31.5%	17.39%	29.63%	18.52%
With internal checklist	10%	16.67%	4.35%	3.7%	7.41%
Testing by yourself	22.5%	16.67%	13.04%	22.22%	14.81%
We often do not test accessibility	22.5%	14.58%	39.13%	29.63%	14.81%
Other	5%	-	-	-	3.7%

The survey highlighted that designers do not involve target users in the design process because of a lack of opportunities, resources or time. In fact the involvement of disabled or older people in some of the most crucial design phases as requirements collection and testing is rather low.

Consolidated requirements from survey

Designers and developers expressed the need for supporting tools for the design of accessible solutions with advanced features, applications and commands, since they are typically experienced and they work in engineering.

At least two levels of functions should be provided: basic functionalities that should be accessible by anyone and also advanced functions for engineers or skilled users involved in design process. It would be useful to have the chance to switch among several types of visualization (e.g. designer view, programmer or developer view, etc.) and to have resizable instrument boxes or windows. It would be useful to have the chance to create macros and commands. For instance expert users may set macros for testing applications and user interfaces using code directly, in order to allow less expert users to conduct tests using user-friendly or ad-hoc commands. Supporting tools must be able to generate code by creating or moving 2D and 3D graphical objects. Thus designers will use a graphical interface for prototyping and developers will switch to the coding view to test application.

Suggestions on possible guidelines, methods and standards for involving end-users may be given through pop-ups in the supporting tools because older people and impaired people are rarely involved in the design process. VERITAS tools should indicate when and why it would be better to involve older and impaired people in some phases of the design process, since designers and developers often don't know how to involve beneficiaries in the process phases.

Since designers and developers are hardly aware of accessibility problems and existing solutions, the supporting tools should allow also learning more about usability, accessibility guidelines, standards and problems involved in the design for impaired and elderly. Short guides, glossaries and introduction to Human Factors principles might be helpful tools to be included, as well as domain-specific use cases databases. It would be useful to have a bug report service and an online guideline facility, such as a wiki or/and a forum with a community. The access should be provided directly from the supporting tool. There should be an automatic download of updates of accessibility guidelines, ISO or other recommendations. The software should include the feature to create applications according to certain standards. A menu that allows user to choose the relevant standard by which developer wants to check or also a library with created code should be present. The design supporting tools should indicate in each stage which accessibility requirements have to be considered and how this can be achieved.

Designers and developers required also that tools have not to be more time consuming than traditional design tools. They should be seamlessly integrated in the tools they currently use and in their design process, with no or very little learning effort. Hence the applications developed in VERITAS project should immediately demonstrate clear advantages in using them during the design processes. In this sense, it will be useful to have a VERITAS plug-in, thus extending the scope of the adopted design tools, in order to minimize the learning effort. The supporting tools must be able to import files from other frameworks, such as Adobe framework applications (i.e. format such as .psd, .ai, etc.), CAD files (i.e. format .dwg, .dxf, etc.), files from Visual Studio framework (i.e. format vcproj, .sln, etc.). Designers must have the chance to select the coding language to make his application, choosing from the most used ones (e.g. Java, C, C++, C#, PHP, JavaScript).

Evaluation of the VERITAS platform

During VERITAS project a complete platform of tools for supporting designers and developers of user interfaces and applications in their working activities was finally developed according to project objectives. Tools were designed in order to introduce accessibility issues in design process and thus to guarantee impaired people and elderly usable user interfaces and application.

The tools developed by VERITAS project are conceived for 2D and 3D design environment and also immersive simulation was considered. VERITAS tools have been conceived as independent applications from the framework of applications normally used by designers and developers. During VERITAS platform design process, users requirements collected through the describe surveys have been considered. They provide as output relevant suggestions and data about how to face accessibility issues according to end-user models. VERITAS tools are listed below.

- User Model Generator.
- Model Platform.
- 3D Simulation Editor.
- GUI Simulation Editor.
- Avatar Editor.
- 3D Core Simulation Viewer.
- GUI Core Simulation Viewer.
- Interaction Manager.

First release of VERITAS tools was tested in order to get preliminary feedbacks from users (i.e. designers and developers) and then to allow a redesign and restyle of them. Tools were still prototypes but a preliminary qualitative evaluation of them was considered necessary to improve the following development.

Test methodology

One or more VERITAS tools were tested in pilot sites involved in the project according to domains of application. The aim of evaluation process carried out was to test functionality, usability and users' acceptance of tools in real working context.

Evaluation process consisted of two phases: preparation phase in which researchers establish contact with participants (i.e. designers and developers) and carried out a training session about VERITAS platform; then pilot phase in which tools were tested through direct experiencing by users and data collected through observation, audio-video recording, interviews and questionnaires and also through logging data from execution of defined tasks.

Test execution

This paper focuses on VERITAS tools evaluated by designers and developer working in the smart-living domain.

Test sessions were carried in two pilot site: in Italy at RE:Lab facilities and in Greece at CERTH/HIT facilities. In Italy designers from RE:Lab and from Indesit Company R&D departments were involved. Test sessions involved 11 participants (3 female; 8 male).

Partners involved in Italian test site evaluated tools relevant for designers and developers involved their R&D departments as listed below.

- User Model Generator, for the creation of impaired people or elderly users model.
- GUI Simulation Editor, for the creation of user interaction model with the designed application.
- GUI Core Simulation Viewer, for the simulation of the interaction experienced by impaired people or elderly.

Users were invited to design the HMI of a new oven and then to evaluate the usefulness, usability, and effectiveness of the VERITAS platform. They went through the overall design process focusing of potential adoption of the final target application (i.e. the oven HMI) by impaired people or elderly. They were asked to experience VERITAS tools in order to improve their typical design process. In order to reduce test session duration, test material (e.g. the oven HMI) was prepared.

Most relevant results and suggestions for improvement

Data from questionnaires (i.e. SUS and TAM questionnaires) reported users enjoyed their experience of using tools and they would generally like to use them more often as long as they become available. Someone expressed the belief that they would need to learn a lot before feeling confident using them. Users felt that the tested tools would increase their productivity and effectiveness, but also have some usability issues. The results from all pilot sites are included in this analysis. Data have been analysed and the full set of results is available in VERITAS project deliverables. Results from SUS questionnaires about tools is reported here as example.

Table 16. SUS questionnaire scores for VERITAS tools in smart-living domain

Users	*SUS Score*
Ver-GEN	54.6
Ver-AE	75.0
VerSed-3D	75.0
VerSim-3D	75.0
VerIM +CFL	65.0
VerSEd-GUI	18.0
VerSim-GUI	84.5

Lessons learned from preliminary evaluation of VERITAS tools mainly focus on test protocol and they aim at providing recommendations in order to improve test procedures and methodology during later evaluations. Several lessons learned and recommendations about the usability and functionality of tools are also summarized.

The preparation phase is an important part of the evaluation. However, based on the feedbacks received from interviews it is clear that participants presented some gaps in their understanding of the tools functionality. This finding brings up the need for even better preparation and possibly explanation of the tools' functionalities. This could be achieved by employing a better documentation that should detail features.

About testing protocol data gathering methods need to be improved in order to ensure data quality. For instance users need to be more encouraged in order to complete feedback forms and questionnaires. Log files need to be gathered consistently and the procedure to collect them need to be improved.

This evaluation provided feedbacks about users' perceptions of their experience of using the VERITAS tools. The overall feedback received with the questionnaires was quite positive and it briefly reported below.

- The majority of the participants revealed they would like to use tools more often once they become available.
- Participants appreciated the tools capability to help them improve their performance and noted they can be effective for their purpose.
- Tools were characterized as useful, providing a good experience to users.
- Usability, however, was not highly rated amongst participants initial perceptions. Tested prototypes were slightly difficult to use and users would need to learn a lot before efficiently using them. The most common need identified was to significantly improve the task workflow of the tools

Conclusions

An extensive survey was undertaken among designers and developers in the field of the automotive, smart-living spaces, workplace, infotainment and healthcare areas in the framework of VERITAS project in order to collect user requirements for the tools that the project is aiming to design. This resulted in a State of the Art analysis of currently existing models and simulation software for the different application areas, with focus on whether they were able to model and simulate disabilities, but also whether they were applicable across the application areas. Moreover the survey among developers and designers in the different application areas also revealed how accessibility is currently addressed and what tools are being used in this process.

Surveys revealed VERITAS project would be offering solutions that would fill a gap in current design and development processes. Accessibility is at present hardly considered in design process due to the lack of knowledge and supportive tools. VERITAS tools are expected to provide aid in the designing phases increasing awareness and knowledge, whether it is by tools themselves or by providing access to direct knowledge or sources.

This report provided also a brief summary of the results derived from the preliminary evaluation sessions of the tool prototypes carried out with developers and designers in the smart-living domain. Feedback obtained from participants using VERITAS tools in real context was reasonably positive with respect to their functionality; nevertheless, several important usability issues were identified and should be taken into consideration for the development of next releases. The quality of the collected data needs to be significantly improved for more in-depth and detailed findings.

New version of VERITAS tools will be released at the end of 2013 and new test sessions will be performed with users (i.e. designer and developers). Also some applications developed by users according to suggestions provided by tools about accessibility will be tested with impaired people and elderly.

References

Dillman, D.A. et al. (2001), Response rate and measurement differences in mixed mode surveys using mail, telephone, interactive voice response and the Internet. *Social Science Research*, Vol. 38, 1-18.

VERITAS Consortium (2013) Project FP7 n. 247765, *D1.1.1 UCD-based user requirements extraction*, available at http://veritas-project.eu/deliverables (consulted the 26th of September 2013)

VERITAS Consortium (2013) Project FP7 n. 247765, *D1.1.2 UCD design guidelines for applications development*, available at http://veritas-project.eu/deliverables (consulted the 26th of September 2013)

VERITAS Consortium (2013) Project FP7 n. 247765, *First Iteration of Pilot Tests with the Designers/Developers – A Summary Report of the Results,* Internal Report

Infrastructure redesign to improve vulnerable road users' safety

Tiziana C. Callari[1], Gabriella Tirello[2], Cristiano Occelli[1], & Alessandra Re[1]
[1]Department of Psychology, University of Turin
[2]Autostrada Torino-Savona S.p.A.
Italy

Abstract

The research reported the case of a footbridge in the Turin province (Italy) that surveyed increasing accidents involving vulnerable road users who chose not to use it when crossing the main road. The reasons behind this choice were investigated and, embracing an ergonomic and systems approach, possible redesign solutions were suggested. The results showed how the different components of the footbridge system (environment, physical design, users, etc.) interact to determine the vulnerable road users' behaviour. Requirements for infrastructure redesign were defined and discussed with the local institutional stakeholders to be implemented in the on-going local Plan of Urban Regeneration and Environmental Improvement.

Introduction

When addressing issues concerning road safety, recommendations usually mention to enforce and improve national and international legislation, especially in matters related to speed, drink–driving and use of helmets (for riders of motorized two-wheelers), seat-belts and child car restraints, and severity and promptness of punishment for violations (WHO, 2013). At the same time, other remedies to contrast the phenomenon are mentioned, like information, education, and police surveillance. Despite these measures, according to the World Health Organization (WHO) in 2012 the world vulnerable road users accounted for the 43% of road deaths (WHO, 2013). In 2011 in Italy over 21,000 injured and 589 deaths - out of which 368 were people over 65 years - were registered among the vulnerable road users (ISTAT, 2011). The phenomenon is of such international noteworthy relevance that the United Nations dedicated the 6th to the 12th May 2013 as "Road Safety Week" to highlight pedestrian safety across 70 different countries. In addition, the European Road Safety Programme 2011-2020 (European Commission, 2010) has envisaged halving the number of road deaths by 2020 and strengthening the actions in safety of vehicles, infrastructure, and users.

At a scientific level different approaches to road safety have been developed. These approaches move from an individual perspective - where accidents are attributed to human errors - to a systems theory - where all the stakeholders and components to a

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

road safety programme are taken into account to suggest countermeasures and realist solutions to be implemented (Larsson, Dekker, & Tingvall, 2010). To this extent, the proposed research took place in Italy, in a Turin province area assessed at high risk of road accidents. Despite there being a usable footbridge, the area has witnessed increasing accidents involving vulnerable road users who chose not to use it when crossing the main road. *Why do pedestrians and cyclists decide not to use the footbridge?* The answer to the question may be addressed under different perspectives. An individual and road-user approach would state that if the vulnerable road users had made proper use of the footbridge and the car drivers had driven respecting the speed limits, road accidents would not have happened. On the other hand, systems approach researchers would state that if the different components of the target road system had been properly designed, these would have limited the chance of human errors.

The present study intends to demonstrate how user-centred and systems design approaches can be used to illustrate the designers how road system components interact with the final users and suggest verified (re)design solutions. In the aim to design and implement "*good patterns of interchange between artificial and human systems*" (Re, 2012, pag. 54), the Human Factors and Ergonomics (HF/E) discipline has displayed a strong interest in safety issues. This approach is useful to address road safety issues because it is focused on adapting the system to human being needs, and highlighting interactions and users' centrality in design. The main objective is to improve the interactions between the road users, the environment, the designers and the management with the final aim to enhance road safety.

Therefore, the present research had the following specific objectives:

(1) To illustrate evidence-based design representations of the footbridge and road system, and point out the current flawed infrastructural aspects under discussion;
(2) To describe the vulnerable road users' behaviors, particularly in relation to the decision to use (or not) the footbridge of SP363 road Villastellone;
(3) To inform designers of future infrastructure system with possible redesign solutions.

Road safety approaches

All the countermeasures aimed at changing the behaviour of the road user are certainly adequate to produce some positive effects. Nevertheless these recommendations betray a direct connection to a traditional approach to road safety that proved not to be more effective in preventing road accidents and injuries (Wegman, 2002). Such an approach is often called the road-user approach (Larsson et al., 2010), and it states, at its basis, that road users are responsible for the safety of the road transport system. This view influences the road safety issues' interpretation in terms of human error, which is considered the cause of most road crashes (WHO, 2013; Brookhuis et al., 2011). In such a view accidents are to be prevented through behavioral changes and coping skills: in this approach it is the user that has to adapt herself to the road transport system. At the end of the nineties, a more system-oriented view gained visibility, in terms of the road system that has to be adapted to the human being and her limitations and imperfections. It became evident that accidents are rarely due to a single factor, but they are rather a resultant of the variability of different components. For this reason improvements in road safety can be achieved only if all the elements are addressed (Larsson et al., 2010; Elvik et al. 2009; Ghazwan, 2007). This led to the introduction of the Road Transport System (RTS) concept, outlining a road system composed of three main elements in interaction: the user, the vehicle and the infrastructure (the environment). This different focus highlighted factors that could be beyond road's users control, such as poor design or failure in the performance of road infrastructures: errors in road design can now be considered as one of the main elements that could cause accidents. Human errors are considered as part of the system, something that designers and managers have to take into account delivering a system that should accommodate potential mistakes (Brookhuis et al., 2011; IRF, 2003). The prevention of human errors is made through information and education, but also providing road users with surroundings in which the chance of human errors is limited (WHO, 2013; Racioppi et al., 2004). Humans make errors because the system is not adequately designed to meet their capabilities and needs. To provide an example, Vision Zero represents the Swedish holistic and systemic approach to road safety grounded on these premises. It states that the system must be adapted to the mental and physical conditions and limitations of the human being, placing a great responsibility on road builders and managers (WHO, 2009). Therefore the countermeasures adopted by policies have been integrated with different elements: development of road infrastructure design to mitigate the consequences of road crashes, measures to reduce the transfer of energy on impact, changes to the road network (motorways, bypasses, etc,), improved levels of road maintenance, installing safety barriers and obstacle protection, etc.

Nevertheless, these approaches could fail in preventing and addressing accidents causes and consequences. The main difficulty is to integrate all the different RTS components in the same intervention. First of all, the various elements of a RTS could have evolved in separate ways, due to historical reasons. For example, in the past, roads were designed primarily to meet social and economic needs; even now we sometimes have to face road networks neither designed nor adapted to vehicles, because two elements (the road and the vehicles) developed independently of each other. Furthermore, in the past the road user was hardly ever considered; when considered, they were above all drivers. Another common issue is the fact that, even though a systemic approach is taken in account, improvements put in place were not integrated in a whole. Safety measures regarding all the RTS components can easily become useless, if they are developed in isolation from each other. In the last few years, road authorities have tried to solve these issues and improve the quality and the performance of the road network, by allocating more and more responsibility to road designers, encouraging involvement of experts in human sciences and aiming to adopt multidisciplinary approaches.

The research setting: the SP393 road Villastellone

In 2012 the Piedmont Region published the ten-year balance of road accidents report (Regione Piemonte - Centro monitoraggio regionale della sicurezza stradale, 2012), in which it is shown that in 2011 the number of victims among the vulnerable road users has not significantly changed with respect to 2001. In particular, the data referred to in the two recent years of 2010-2011 (see Table 1) deaths among the vulnerable road users show only a slight increment of 5%. The pedestrians involved in mortal accidents were from 54 in 2010 to 61 in 2011 with an increment of 19%, while the motorcycles' deaths registered an increment of 23%. Only the cyclists show a reduction in mortal accidents, but they register an increment of 8% of injured cyclists. A concerning aspect is registered among the youths in-between 22-29 years old, whose deaths in road accidents increased by 44%.

Table 1. Accidents in Piedmont between 2001 and 2011 - injuries and deaths of vulnerable road users (Regione Piemonte - Centro monitoraggio regionale della sicurezza stradale, 2012)

	Deaths				Injured			
Vulnerable users	**Mean 2001-2010**	**2010**	**2011**	**Variation**	**Mean 2001-2010**	**2010**	**2011**	**Variation**
Pedestrians	63	54	61	**13%**	1.562	1.663	1.629	**-2%**
Cyclists	27	27	15	**-44%**	844	947	1.027	**8%**
Motorcyclists	68	52	64	**23%**	1.863	1.789	1.849	**3%**
TOTAL	**158**	**133**	**140**	**5%**	**4.269**	**4.399**	**4.505**	**2%**
Users by age	**Mean 2001-2010**	**2010**	**2011**	**Variation**	**Mean 2001-2010**	**2010**	**2011**	**Variation**
Kids (11-13 yrs old)	3	1	0	**-**	246	249	249	**0%**
Teens (14-17 yrs old)	14	9	5	**-44%**	1.060	947	867	**-8%**
New-licenced (18-21 yrs old)	30	26	17	**-35%**	2.388	2.106	1.951	**-7%**
Youths (22-29 yrs old)	79	32	46	**44%**	4.647	3.610	3.380	**-6%**
Adults (30-69 yrs old)	213	166	159	**-4%**	11.212	10.889	10.460	**-4%**
Older people (over 70 yrs old)	85	89	89	**0%**	1.348	1.494	1.594	**7%**

In the list of the Turin province roads declared, SP393 road Villastellone presents a high degree of accidents. This is a road with high traffic level that goes through the southern area of Turin connecting the towns of Moncalieri and Carmagnola, through Villastellone, Cambiano and Carignano. A road safety assessment with regard to SP393 road Villastellone was presented in a Safety Analysis Document, edited by the Province of Turin (Provincia di Torino, 2012) in line with the National Plan of Road Safety which suggests the Road Safety Audit (a preliminary safety analysis applied in the design phase) and Road safety Review procedures (applied to the operative infrastructures). The SP393 road Villastellone has been ranked at the 20th place out of 679 Piedmont Region roads regarding Risk Index[9]. When comparing these data with those at national level, the SP393 road Villastellone reports a high risk assessment, as it presents an average of 14.5 deaths × 100 km / year, despite the National Plan of Road Safety for this kind of roads has fixed the target at 0.9 deaths / year. The Severity Index is equal to 152.8 (number of deaths on the sum of deaths and injuries × thousand) compared to the regional 47.7 and provincial 67.1.

In the present study, the problem space refered to the roads intersection at km 10+800 of the SP393 road Villastellone, where cars and trucks intersect with pedestrians and cyclists who come from the nearby village of Villastellone, to reach

[9] The Risk Index RI is calculated by dividing the specific rate of road accidents equal to 1.53 (considering the regional RI equal to 1.0 and the provincial RI equal to 0.74) and the Mortality Risk MR equal to 2.30 (considering the regional MR equal to 1.0 and the provincial MR equal to 1.46).

an area called Villasport, which includes many sport centres, leisure and business activities (see Figure 1). The number of accidents involving road vulnerable road users and vehicles is considered.

Figure 1. View to the footbridge and the road system in Villastellone

Method

The research design consisted of different souces of evidence able to account for the different system components of the SP363 road Vittastellone area under analysis, and support data triangulation to achieve data saturation and convergence. The data collection took over a 7-month period. The main stakeholders and final users of the road and footbridge system area were involved. All the participants gave informed consent.

Archival data of the traffic road area were collected. They consisted of statistics regarding the road traffic incidents involving road means and vulnerable road users (ISTAT, 2013); infrastructure maps; newspapers and journal articles of the province target-area. These data tracked over a three-year period (Jan. 2010- Dec. 2012). In addition, primary sources were collected and included original documents, observation data, interviews, questionnaires and focus groups.

Firstly, direct observations and pictures were taken to gather first data on the environmental and infractural conditions of the 500 m road area under analysis, with an aim to register the state-of-the art of the signs and markings present in compliance with the Italian regulatory standards; the intersections with other roads and with private passages, the sidewalks, and parking areas, the lighting system (number and position of the street lights).

To understand the vulnerable road users’ behaviours in their choice in using the footbridge to reach Villasport area, a multiple-choice-answers questionnaire was designed. The items regarded the frequency they attend the services of Villasport area per week, the time of the day they travelled, the means of transportation (motorbike, byke, by foot) used, and comments to possible alternatives to the use of the footbridge. The participants were N= 60 vulnerable road users (N= 30 male users, N= 30 female users; mean age= 32 yrs). In returning the questionnaires, the respondees were asked the availability to participate in group-interviews to expand on the reasons behind the choice (or not) to use the footbridge. Consequently, two groups-interviews were organised involving N= 9 adults (users of Villasport area, parents, random passer-by) and N= 4 teenagers (whose informed consent was signed by their parents). They were inquired about their behaviours in different situations and the Critical Incident Technique (Butterfield, 2005; Flanagan, 1954) was used to investigate *negative* past events related to the choice of not using the footbridge. The interviews lasted approx. 60 minutes; they were recorded and completely transcribed.

The empirical data were analysed considering the specific road system components involved (environment, infrastructure, etc) in interaction with the vulnerable road user. The content analysis (Krippendorff, 2004) was performed through the use of the NVivo software (Tom and Lyn Richards, 1981, cit. in Bazeley, 2007), with a good degree of inter-judge agreement (k-Cohen= .74).

Results

The data collected from the first exploratory observational phase resulted in an *infographic* map (see Figure 2) of the SP363 Villastellone road system. The *infographic* map highlighted the current critical and/or missing environmental and infrastructural elements to bring under attention. It highlighted that there was a lack of signs and markings in the critical intersections with roads and private passages (flaws N. 2, 5), the lighting system was insufficient, in particular in the parking areas and along the footbridge (flaws N. 1, 3, 4) and the pedestrian path of the footbridge resulted compromised in not-paved sections (flaws N. 6) that could affect the vulnerable road users' safety in particular in adverse climate conditions, like rain or snow.

The results from the questionnaires and two groups-interviews produced user-centred rich material to illustrate to the designers the behaviours of the users, their motivations and the specific context of use. These were related to the users' characteristics (socio-demographic data); the means of transportation used (bike, motorbike, car, or none); the specific needs to reach Villasport area; the external environment conditions. In particular, some elements were crucial for a more complete understanding of the interactions occurring between the vulnerable road users and the road system components. Users declared that the pedestrian path of the footbridge is not completely paved, causing slippery spots under adverse climate conditions and too long to use when travelling to Villasport area – in fact the footbridge path is 352.81 m. long and 12.80 m. wide. In addition, one side of the footbridge ends near a grove of trees that is perceived by vulnerable users as particularly unsafe in dark lighting conditions. In fact, this area and along the footbridge street lacks sufficient lighting spots. The intersections of the main road with secondary roads and the footbridge connect poorly with main infrastructural safety elements: lack of marks and signals for the drivers informing the presence of vulnerable road users; absence of speed limiters; etc.

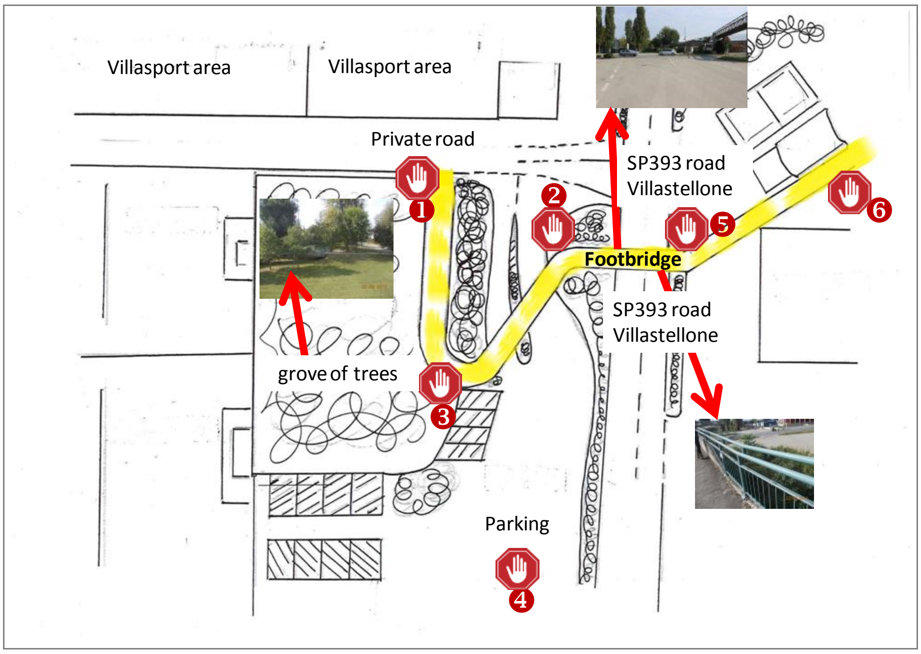


Figure 2. Infographic map of the SP363 road Villastellone of the area under analysis, with flawed infrastructural aspects marked.

In order to better illustrate to the designers how road system components interact with the final users and to share a sufficiently wide view on complex systems, the scenario-based design technique was used (Carroll, 2000; Rosson & Carroll, 2002). Scenarios are "*stories about people and their activities. [...] Scenarios highlight goals suggested by the appearance and behaviour of the system; what people try to do with the system;and what interpratations people make of what happens to them*" (Carrol, 2000, p. 46). Scenarios are a powerful instrument when addressing the complexity of systems and analysis, and deemed to be an effective mechanism to discover requirements for the scope of the present study. Scenarios may vary from brief stories to richly structured analysis, but are always based on the idea to present sequences of actions carried out by agents detecting possible omissions, threats and inconsistencies during their interaction with the system under analysis.

Five scenarios were designed highlighting the *setting* (the footbridge and the SP363 Villastellone area), *agents* who might be *involved* (the different range of vulnerable road users) with contextualised *needs* and *goals* (aiming to Villasport area, and deciding to use, or not, the footbridge). For example, in Scenario N. 1 the *agent* is an aged vulnerable user using the footbridge but due to lack of adequate infrastructure (path poorly paved) is at risk of falling and getting injured.

Scenario N. 1 - It is a cold rainy day of end November. Maria is 72 yrs old and needs to keep in shape. She has atrophy in her left foot therefore three days a week she has her morning gym practice. She has time, so safely decides to use the footbridge to reach Villasport area; she has her bag in one hand and in the other her gym bag. The pedestrian path is not completely paved, with barely avoidable puddles. Suddenly she loses control, and slides down.
Outcome: Loose of balance and fall with consequent injury/ies.

Scenarios help designers by providing (re)design requirements and solutions of future infrastructure system. To ensure the grounding of the shared requirements in vulnerable users' needs, the requirements were acquired *with the vulnerable users and stakeholder*, to be able to uproot tacit knowledge elicitation and modelling of users' needs. This implied a focus on requirements based on linguistic expressions rather than formal specification and validation (Sutcliffe & Maiden, 1993).

Scenario N. 2 - It is 2 p.m. and Marco, Giulio and Paolo –a group of teenage-friends needs to reach the Villasport area for swimming practice. They usually meet after lunch. To make it faster they decide to take the shortcut by the bowling green. It is daylight, they know the street, they are in a group and feel confident. At that time of the day, the traffic is consistent. The group is having fun and is distracted and thus does not perceive a vehicle arriving over the speed limit.
Outcome: Possible collision due to fast speed driving.

		Component
Requirement	Drivers shall be able to be informed of the presence of vulnerable road users.	Infrastructure
Requirement	Drivers shall be able to be informed of the obligation to reduce speed when approaching the footbridge area.	Infrastructure
Requirement	Vulnerable users and drivers shall be able to be surrounded by adequate lighting spots when approaching the footbridge area.	Infrastructure Environment
Requirement	Vulnerable users shall be able to feel a social safety perception when using the footbridge.	Infrastructure Users

A focus group with the local institutional stakeholders was then organised to present the study results. The possible redesign solutions of future infrastructure system were analysed and first suggestions were decided to be implemented in the on-going local Plan of Urban Regeneration and Environmental Improvement.

Discussion and conclusions

The study intended to address the road safety issue in a province assessed at high risk of accidents by user-centred and systems design approaches to illustrate the designers how road system components interact with the final users and suggest verified (re)design solutions. The data collection involved the main local institutional stakeholders and vulnerable road users and made use of different souces of evidence able to account for the different system components of the SP363 road Vittastellone area under analysis.

The results showed evidence-based design representations of Villastellone road system and provided possible redesign solutions which were discussed in a focus group organised with the local institutional stakeholders. In particular, the main infractructural flaws were immediately addressed. The lighting system was improved in the areas along the footbridge and car parking lot. New traffic circles and guard rails to separate the two driving lanes are under construction with the aim to slow the vehicles speed. To protect pedestrians and cyclists from the traffic road and promote the use of bikes instead of cars, new bicycle and pedestrian mobility paths separated by the main road are under construction.

These are concrete answers to address the infrastructural problems arisen from the research results. Still, the specific technical solutions have to be accompanied by an understanding of the different situations and factors occurring in a specific complex system, to be able to respond to the final users, in particular the vulnerable road users. Pedestrians still result among the most vulnerable road users. This is mostly because for years their needs have been neglected in favor of motorized transportation. To reduce injuries among vulnerable road users, an important role can be played by road infrastructure. The recent engineering innovations favours an understanding of needs and variability of all road users, adapting the system design to the capabilities and limitations of the human being for each user category and different use of the road.

References

Bazeley, P. (2007). *Qualitative data analysis with NVivo*. London: SAGE.

Brookhuis, K.A., Wiethoff, M., Bekiaris, E., & Gaitanidou, E. (2011). From accidents to measures. The general IN-SAFETY approach. In E. Bekiaris, M. Wiethoff, & E. Gaitanidou (Eds.). *Infrastructure and safety in a collaborative world. Road traffic safety*. Berlin: Springer Verlag.

Butterfield, L.D., Borgen, W.A., Amundson, N.E., & Maglio, A.S.T. (2005). Fifty years of the critical incident technique: 1954-2003 and beyond. *Qualitative Research, 5*, 475-497.

Carroll, J.M. (2000). *Making use: scenario-based design of human-computer interactions*. Cambridge, MA: MIT Press.

Elvik, R., Hoye, A., Vaa, T., Sorensen, M. (2009). *The handbook of road safety measures.* UK: Emerald Group Publishing Limited.

European Commission (2010). Towards a European road safety area: policy orientations on road safety 2011-2020. Brussels: European Commission.

Flanagan, J.C. (1954). The Critical Incident Technique. *Psychological Bulletin, 51*, 327-358.

Ghazwan, A. (2007). *Road Safety Development Index (RSDI). Theory, philosophy and practice.* Linköping, Sweden: LiUTryck.

International Road Federation (2003). *Safe mobility*. Geneva,Washington, DC: IRF.

Istituto Nazionale di Statistica (2013). *La rilevazione degli incidenti stradali con lesioni a persone. Anno 2011.* http://www.istat.it/it/archivio/87539

Krippendorff, K. (2004). *Content analysis: an introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Larsson, P., Dekker, S.W.A., & Tingvall, C. (2010). The need for a system theory approach to road safety. *Safety Science*, *48*, 1167-1174.

Provincia di Torino (2012). *Documento analisi di sicurezza*. Provincia Torino, Area Viabilità.

Racioppi F., Eriksson, L., Tingvall, C. & Villaveces, A. (2004). *Preventing road traffic injury: a public health perspective for Europe*. Copenhagen: WHO Regional Office for Europe.

Re, A. (2012). Interdisciplinary research and practice in ergonomics. *Journal of Biological research*, *LXXXV*, 53-56.

Regione Piemonte - Centro monitoraggio regionale della sicurezza stradale (2012). *L'incidentalità in Piemonte: bilancio 2001-2010 e situazione al 2011. Rapporto 2012*. Torino: IRES.

Rosson, M.B., & Carroll, J.M. (2002). *Usability engineering: scenario-based development of human-computer interaction*. San Diego, CA: Academic Press.

Sutcliffe, A.G., & Maiden, N.A.M. (1993). Bridging the Requirements Gap: Policies, Goals and Domains. *Proceeding of the 7th International Workshop of Software Specification and Design IWSSD* (52-55). CA, USA: IEEE Computer Society Press.

Wegman, F. (2002). *Review of Ireland's road safety strategy*. Leidschendam, The Netherlands: SWOV.

World Health Organization (2009). *European status report on road safety: towards safer roads and healthier transport choices*. Copenhagen: WHO Regional Office for Europe.

World Health Organization, (2013). *Global status report on road safety 2013. Supporting a decade of action.* Luxembourg: Department of Violence and Injury Prevention and Disability (VIP).

"Full monty" or low-key: two traffic management approaches and their effect on change perception

Ilse M. Harms[1] & Karel A. Brookhuis[2,3]
[1]Rijkswaterstaat Ministry of Infrastructure and the Environment
[2]University of Groningen
[3]Delft University of Technology
The Netherlands

Abstract

European road authorities differ in their views on whether variable-message signs (VMSs) can be in use continually or that they should only display information when necessary. This study compared the two approaches in terms of the ease with which drivers perceive changes in traffic management information on VMSs. In a 2x3 design, change detection for variable speed limits on VMSs was measured for information addition and information change under three conditions of information discriminability. Participants were shown videos of a single motorway – in order to create a more natural setting in which they are familiar with the road – while using an intentional approach in which participants were aware that something could change. The preliminary results are discussed in relation to detectability of, recollection of, and expectations about the new speed limit. This research gives more insight into change detection failure for the two traffic management approaches and possible countermeasures.

Introduction

On a vast majority of the many European motorways drivers may encounter one or more varieties of dynamic traffic management (DTM). Most of the DTM information is presented on variable-message signs (VMSs) and may include information regarding delays, route advice, road work warnings, graphic congestion displays, dynamic lane closures and variable speed limits. Road authorities like to use DTM as it enables them to influence real-time driver behaviour, by just changing the information for drivers to fit the present situation on the road and/or the road network. In this way road authorities can improve both traffic safety as well as traffic circulation. As congestions dissolve and road works stop at a certain moment, DTM information does not have to be displayed continually. When not in use for DTM, VMSs may therefore be left blank or display other information, for example road safety messages, information on the weather, and amber alert (Tay & de Barros, 2010). However, VMSs and their maintenance are costly. For this reason, among others, some road authorities might argue it is better to use VMSs continually while others feel that the VMSs should only display information when necessary for DTM purposes.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

One question is: will drivers still detect changes in DTM information if the VMSs are in use continually? Studies have shown that our ability to detect changes around us is limited (for a review see Simons & Levin, 1997) and that this visual limitation also applies to the driving domain (Velichkovsky et al., 2002; McCarley et al., 2004; Caird et al., 2005; Lee et al., 2007; Galpin et al., 2009; White & Caird, 2010; Martens, 2011; Koustanaï et al., 2012). This limitation in detecting changes which are in fact clearly visible is called change blindness. In semi-naturalistic studies in which drivers were familiarised with a route in a driving simulator, drivers failed to notice changes in driving related stimuli, such as fixed road signs, the priority at an intersection and variable speed limits on gantries (Martens & Fox, 2007; Charlton & Starkey, 2011; Harms, 2012). Even changes that are expected may still go unobserved (Rensink et al., 1997). Harms (2012) found that in her repeated measures driving simulator study, nine out of twenty-four participants expected that the speed limits on VMSs could change while only five of them actually detected the change.

Changes are regularly accompanied by other transient stimuli that catch attention. When transient motion signals are masked the changing object will be more difficult to detect. Such masking could occur due to a lapse of time, also known as gap-contingency. This occurs when items change during a temporal gap (Rensink, 2002). From the driver's seat of a car that speeds along a motorway, DTM information displayed on subsequent intermittent overhead gantries can be regarded as a continuous flow of information interrupted by lapses of time necessary to drive from one gantry to the next. Transient motion signals that accompany a change in the DTM information while driving to the next gantry, will be masked by the temporal gap that resulted from the lapse of time. This makes it more difficult to detect the change. Studies have also shown that it is more difficult to detect a change when already visible information changes slightly as compared to information that is added to a scene (for a review see Rensink, 2002). Therefore the current research focusses on the ease of detecting and identifying changes in information on VMSs which are left blank occasionally as compared to VMSs which are used continually and examines the effect of adding transient motion signals as a possible countermeasure for change detection failures.

Method

Experimental design

Using an intentional approach, in which the observer is instructed to fully expect a change and devotes all available resources to detecting it (cf. Simons & Mitroff, 2001), participants were shown fifteen short videos and one practice video. All videos represented a motorway equipped with three gantries displaying variable speed limits (VSLs) on electronic signs per driving lane, similar to Harms (2012). To familiarise participants with the motorway and its surroundings, the first video was displayed unchanged and viewed thirteen times. To prevent any interference from participants who might expect the change to happen in the last (15^{th}) video, the change was introduced in the 14^{th} video. In this video, the VSLs were changed from 100 km/h to 80 km/h on the second and the third gantries. Video 15 consisted of a recollection test.

In a 2x3 design, change detection was measured for information addition (IA) and information change (IC), under three conditions of information discriminability (see Table 1). In the IA condition, the VSL signs were always turned off, unless the speed limit changed; in the IC condition, the VSL signs were always turned on, hence depicting a speed limit on every VSL sign (see Figure 1). The three conditions of information discriminability varied in how the changed speed limit was displayed. In the Basic condition, it was displayed by solely depicting the speed limit itself; in the Flash condition, it was shown with alternating orange flashers; in the Wave condition, it appeared as if the depicted speed limit was moving in a wave-like manner (see Figure 2). To ensure that all participant groups were equally able to detect changes, participants were assigned to one of the six groups based on age and gender, as some studies reported an age-related decline in the detection of changes (Costello et al., 2010; Caird et al., 2005; Rizzo et al., 2009; Wascher et al., 2012).

Table 1. The 2x3 design

	Varieties of information discriminability		
	Basic	*Flash*	*Wave*
Information addition	Group 1	Group 2	Group 3
Information change	Group 4	Group 5	Group 6

Participants

One hundred an forty one participants completed the experiment, though only 133 participants did so successfully, since eight participants were excluded as they used a computer screen smaller than the video's resolution. Participants were recruited using invitations from both fellow participants and the researcher, including advertisements on the internet and an invitation on a birth announcement card. Participants' age varied from 19 to 75 years (M = 42.8 years, SD = 11.5), and both male and female Dutch drivers participated (86 males and 47 females). All participants possessed a driving licence and reported normal or corrected to normal eyesight. Participants were not paid for their participation. The groups of participants did not differ significantly for age, gender, education, years of driving licence possession and amount of kilometres driven in the past twelve months.

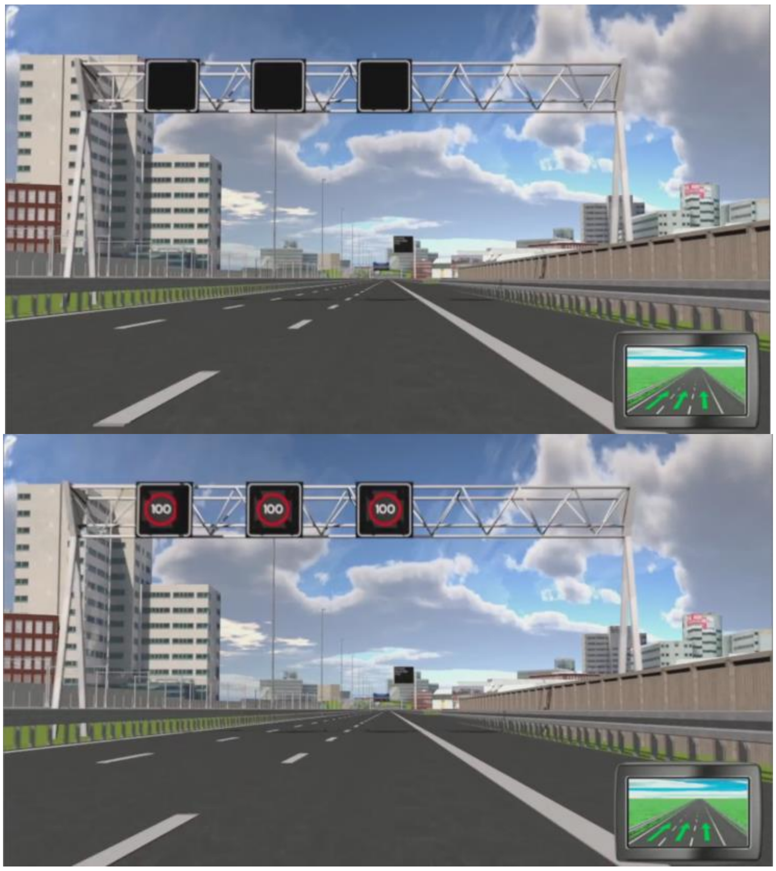


Figure 1. Screen shot of the second gantry in the information addition condition (top) and in the information change condition (bottom) in video 1 to 13.

Task

Participants were instructed they would see fifteen videos showing the same motorway from a driver’s perspective. They were asked to imagine driving the car. In videos 1 to 14, participants had to detect and identify all changes between the current and the previous videos. It was pointed out that the number of changes they would encounter would vary between participants and that it could range from several changes to none at all; in fact, all participants encountered two changes in video 14. To report a change, participants had to press the spacebar on the keyboard. In accordance with Crundall (2012), the screen subsequently turned black while the video was stopped. Participants then described the change they had seen or reported that they had either seen the change but failed to identify it, or had pressed the spacebar accidentally. In video 15, participants performed a recollection test in which several roadside elements – including all speed limits – were omitted. The video stopped at each deleted item and a yellow circle marked the spot where the item had been present in all previous videos. Participants had to report what had been visible at that spot in video 14 and were instructed to guess when necessary. They then reported how confident they were of their answer.

Figure 2. Screen shot of the second gantry under three conditions of information discriminability in video 14. From top to bottom: Basic, Flash, and Wave.

Measures

To detect whether participants noticed the two changes, both detection accuracy and reaction time were measured during videos 1 to 14. Simultaneously, change identification was measured by the descriptions of the detected changes. To ensure changes were attributed to elements of the road and its surroundings, no other traffic was present in the videos. The task in video 15 measured accurate recollection of the changed speed limits and a few other deleted items that had not changed. The deleted items in video 15 included all speed limits, a part of the signposting, route information on a variable message sign (VMS), and a billboard containing a traffic-related advertisement from a recent government campaign.

Finally, participants filled in a questionnaire concerning the omitted items they had encountered. The questions ranged from relatively general ("Did you notice anything special about the speed limits in video 1 to 14?"), to asking whether they had seen the change. This was done by suggesting that there had been two groups: one group for whom the speed limit on the gantries had always been 100 km/h and another group for whom the speed limit on the gantries had decreased from 100 km/h to 80 km/h. While in fact all participants belonged to the second group, they were asked to which group they thought they belonged and how confident they were of their decision. These questions were also asked for the other deleted items, always implying that there were two groups per omitted item and suggesting that it was possible to belong to a group that had encountered either several changes, one change, or no change at all in video 14. Participants were also asked whether they expected that the speed limits – as well as the other deleted items - could change.

Equipment

Participants accessed the experiment on their own computer through a website using a personal entry code. The website could be viewed on a Windows or Apple operating system using a mainstream web browser such as Internet Explorer, Firefox or Google Chrome. Internet access and a keyboard were required. The experiment was not suitable for smart phones or tablets.

The videos showed a forward view of the road ahead, and a navigation device was displayed in the lower right corner. The video resolution was 1136 pixels wide and 640 pixels high. Dedicated software automatically gathered the screen resolution of the computer participants used and recorded space bar hits during video viewings. These responses were sampled with a precision of $1/10^{th}$ of a second. The videos were equipped with car sounds, e.g. engine noise, to make the viewings more realistic. To ensure that participants could hear the audio signal, an audio test was embedded at the beginning of the experiment.

Procedure

Before the actual experiment started, participants received written instructions which explained their task and the procedure. The instruction also requested participants not to pause or to be assisted or disturbed during the experiment. Participants were told that the aim of the experiment was to gain more insight into the perception of changes in the road environment.

The experiment began with a short questionnaire containing questions on participant characteristics. To become familiar with the change detection and identification task, participants viewed one practice video which was self-paced. After viewing the fifteen 40-second videos, participants received a short follow-up questionnaire on the videos they had seen. The experiment took 20 minutes to complete.

Results

Change detection accuracy

Almost three out of four participants (70.5%) did not realise they viewed the same video thirteen times. Alleged changes that have been reported included adjusted travel times on the roadside VMS, the appearance of an emergency bay, the congestion indication on the roadside VMS changing from kilometres into travel time, adjustments in the delineation, new or moved road signs and buildings, and "the car is moving faster". After several video viewings the response rate declined (see Figure 3).

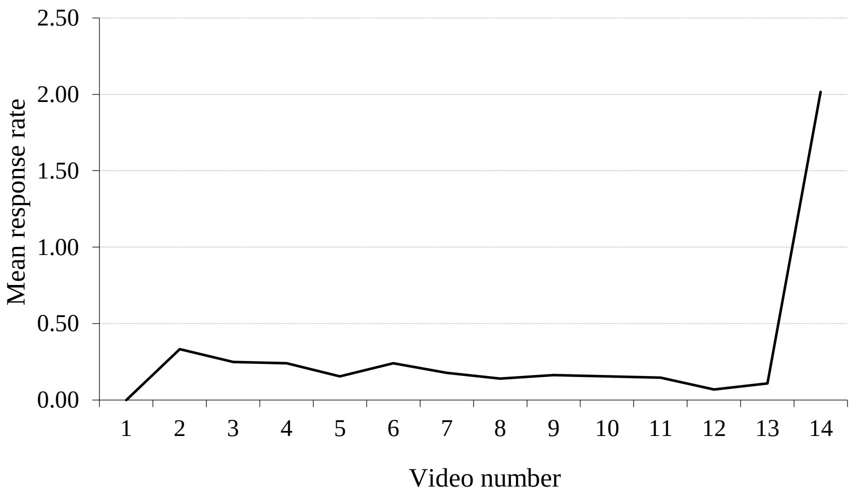


Figure 3. Graph of the mean response rate per video. The response rate for video 1 was zero, as video 1 could not be compared with a preceding video. Video 14 included the two changes.

In video 14, 85.3% of the participants accurately responded to both speed limit changes. Other participants either only responded to the first speed limit change (11.6%), the second speed limit change (2.3%), or did not respond to the changes at all (0.8% , n = 1). Due to a technical error four participants were unable to stop video 14, hence n = 129 for video 14.

Speed limit recollection

While viewing video 15, 34.6% of the participants accurately recollected the speed limits shown in video 14. Most participants (51.1%) falsely recollected all speed limits on VMSs to have changed to 80 km/h and 3.8% claimed that only the last speed limit had changed to 80 km/h. The recollected speed limit sequences of the remaining 10.5% of the participants varied widely (see Figure 4). There was no significant difference in correct speed limit recollections between both IA and IC and between the varieties of information discriminability. On average, the participants who answered correctly were confident of their answer (4.33 on a 5-point Likert scale). Participants who, incorrectly, recollected a sequence of 100 km/h, 80 km/h, 80 km/h and 80 km/h yielded the highest mean confidence rate, 4.41 on a 5-point Likert scale (see Table 2).

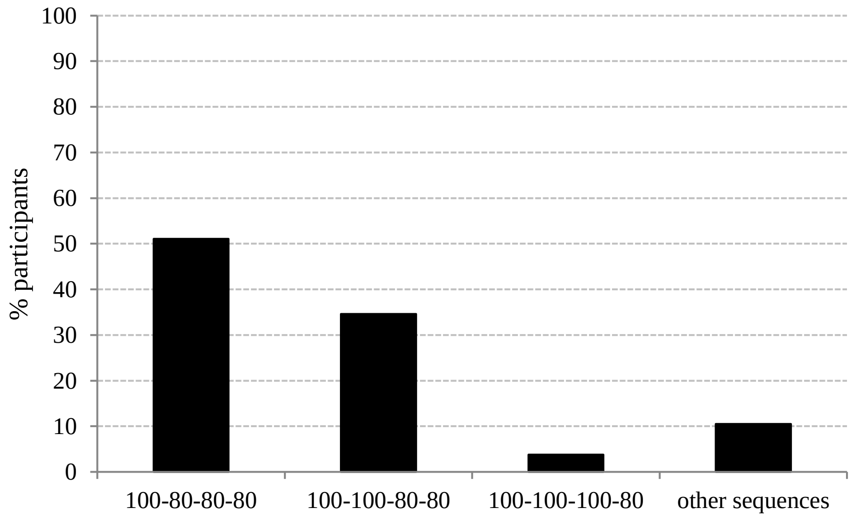


Figure 4. Recollected speed limit sequences in video 15. The correct sequence is 100 km/h, 100 km/h, 80 km/h, and 80 km/h.

Table 2. Confidence rates per speed limit sign.

Recollected sequence	*Speed limit 1*	*Speed limit 2*	*Speed limit 3*	*Speed limit 4*	*Mean confidence rate*
100-80-80-80	4.04	4.47	4.56	4.56	4.41 (*n* = 68)
100-100-80-80*	4.15	3.70	4.74	4.72	4.33 (*n* = 46)
100-100-100-80	3.60	3.80	3.60	4.00	3.75 (*n* = 5)
Other	3.93	4.00	4.21	4.36	4.13 (*n* = 14)

* correct speed limit sequence

When asked whether they had noticed anything special about the speed limits in video 1 to 14, 72.2% of the participants reported the changed speed limit. When a response to the changed speed limit was elicited, 97.7% of the participants accurately reported that they had been in the group for whom the speed limit on some gantries had been 80 km/h in video 14. The remaining 2.3%, reported being part of the group for whom video fourteens variable speed limits remained a 100 km/h. On average, these participants were less confident of their answers compared to the participants who had answered correctly (respectively 3.33 and 4.65 on a 5-point Likert scale).

Expectancy

Almost all participants (95.5%) had expected that the speed limits would change compared to 55.6% for the signposting, 91.0% for the route information on a VMS and 37.6% for the billboard. The participants who expected the variable speed limits would never change (3.0 % of n = 133), all responded accurately to the first change in video 14. Although they all recollected the speed limit sequence incorrectly, they correctly attributed themselves to the group whose speed limits had changed. The participants who did not react to the first change (3.1% of n = 129), did expect the speed limits to change.

Reaction time

The mean reaction time for IA did not differ significantly from IC for the first change in video 14. However, the variance in these reaction times is much larger for IC compared to IA, albeit not significantly larger (see Figure 5). Indicative, when the two largest outliers are excluded, Levene's Test of Equality of Variances shows that, although the mean reaction time is the same, the variance in reaction times for IA and IC differs significantly when testing with an α of 0.10, $F(1,121)$ = 2.889, p = 0.092.

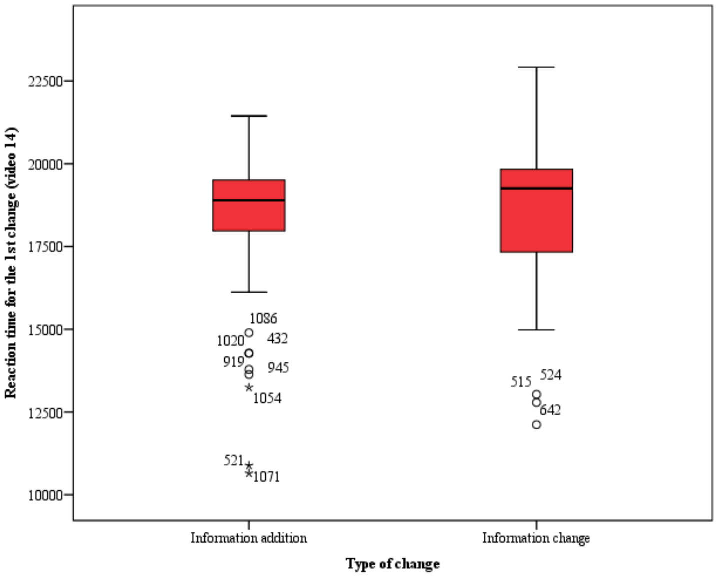


Figure 5. Boxplots of reaction times for the first change in video 14 for information addition (IA) and information change (IC). Reaction times are measured from the start of the video.

For the first change, the mean reaction time differed significantly for the three varieties of information discrimination, $F(2,122) = 3.829$, $p = 0.024$, with Basic yielding the fastest response and Flash the slowest. This difference seized to exist for detection of the second change. Levene's Test of Equality of Variances shows that, for the first change, the variance in the reaction times for Basic, Flash and Wave varies significantly as well, $F(2,122) = 3.440$, $p = 0.035$. Flash has the smallest variance (see Figure 6).

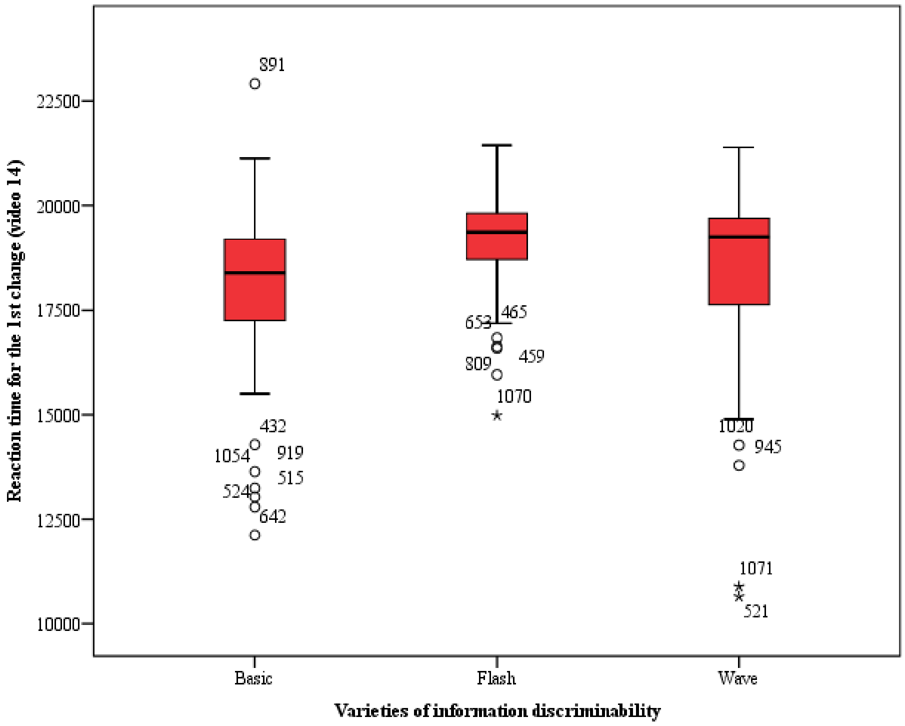


Figure 6. Boxplots of reaction times for the first change in video 14 for the three types of information discriminability. Reaction times are measured from the start of the video.

Conclusions and discussion

Although the study was designed to "force" participants to see the decreased speed limit, still 3.1% did not (timely) detect it; 65.4% did not recollect it correctly; and 2.3% reported afterwards that they had not seen the decreased speed limit at all. The conclusion must be that expecting a change does not necessarily result in change detection. This inability to detect and recollect clearly visible changes to VSLs is consistent with earlier findings (Harms, 2012) and stresses once more the limitations of perceiving information provision, even highly relevant information. This is specifically important for maintaining the correct speed on motorways as perceiving the relevant information is one of the necessary steps to be able to comply with speed limits in the first place.

It remains uncertain whether, without doubt, change detection for VSLs can improve by using either an IA or IC approach. This may be due to the fact that the amount of data was limited. However, the data does contain indications that using an IA approach may lead to less variance in reaction times when a VSL changes. To resolve this issue, more participants should partake in the experiment. Contrary to expectations, equipping the variable speed limits with flashers and therefore adding a transient motion signal to the changed VSL did not increase change detection. This finding is hard to explain. It must be kept in mind though, that reaction times were measured at $1/10^{th}$ of a second and differences were relatively limited. Further analysis with more participants may shed more light on this. However, the clear difference in variance between the three types of information discrimination points out that flashers do indeed decrease overall reaction time to changes in variable speed limits. Introducing a subtle wave to add a transient motion signal to a changed VSL does not appear to improve reaction times or reduce their variance. The wave may have been too subtle or the videos may not be the right instrument to test this.

Large differences in speed on motorways reduce road safety. Even small speed differences may encourage drivers to change lanes and overtake lead vehicles while changing lanes is a potential source of traffic accidents (Hegeman, 2008). Although more research is needed, it appears that with choosing an IC, or “full monty” traffic management approach, road authorities may introduce additional speed differences. Therefore drivers might be safer while driving on a motorway with VMSs that do not display information continually.

References

Caird, J.K., Edwards, C.J., Creaser, J.I., & Horrey, W.J. (2005). Older driver failures of attention at intersections: Using change blindness methods to assess turn decision accuracy. *Human Factors, 47,* 235-249.

Charlton, S.G. & Starkey, N.J. (2011). Driving without awareness: The effects of practice and automaticity on attention and driving. *Transportation Research Part F: Traffic Psychology and Behaviour, 14,* 456-471.

Costello, M. C., Madden, D. J., Mitroff, S. R., & Whiting, W. L. (2010). Age-Related Decline of Visual Processing Components in Change Detection. *Psychology and Aging, 25,* 357-368.

Crundall, D. (2012). Predicting change: a more robust measure of hazard perception? Paper presented at *5th International Conference on Traffic and Transport Psychology*.

Galpin, A., Underwood, G., & Crundall, D. (2009). Change blindness in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour, 12,* 179-185.

Harms, I.M. (2012). Change detection in variable speed limits: failed to look or looked but failed to see? Paper presented at *The 5th International Conference on Traffic and Transport Psychology, 29-31st August 2012, Groningen, The Netherlands*.

Hegeman, G. (2008). *Assisted Overtaking; An Assessment of Overtaking on Two-lane Rural Roads*. TRAIL Research School: PhD Thesis, Delft: Delft University of Technology.

Koustanaï, A., Van Elslande, P., & Bastien, C. (2012). Use of change blindness to measure different abilities to detect relevant changes in natural driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour, 15,* 233-242.

Lee, Y.-C., Lee, J.D., & Boyle, L.N. (2007). Visual Attention in Driving: The Effects of Cognitive Load and Visual Disruption. *Human Factors, 49,* 721-733.

Martens, M.H. (2011). Change detection in traffic: Where do we look and what do we perceive? *Transportation Research Part F: Traffic Psychology and Behaviour, 14,* 240-250.

Martens, M.H. & Fox, M.R.J. (2007). Do familiarity and expectations change perception? Drivers' glances and response to changes. *Transportation Research Part F: Traffic Psychology and Behaviour, 10,* 476-492.

McCarley, J.S., Vais, M.J., Pringle, H., Kramer, A.F., Irwin, D.E., & Strayer, D.L. (2004). Conversation disrupts change detection in complex traffic scenes. *Human Factors, 46,* 424-436.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53,* 245-277.

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: The Need for Attention to Perceive Changes in Scenes. *Psychological Science, 8,* 368-373.

Rizzo, M. D., Sparks, J. D., McEvoy, S. D., Viamonte, S. M., Kellison, I. L., & Vecera, S. P. (2009). Change blindness, aging, and cognition. *Journal of Clinical & Experimental Neuropsychology, 31,* 245-256.

Simons, D.J. & Levin, D.T. (1997). Change blindness. *Trends in Cognitive Sciences, 1,* 261-267.

Simons, D.J. & Mitroff, S.R. (2001). The role of expectations in change detection and attentional capture. In L.R.Harris and M.R.M. Jenkin (Eds.), *Vision and Attention* (pp. 189-207). New York: Springer Verlag.

Tay, R. & de Barros, A.G. (2010). Effectiveness of road safety messages on variable message signs. *Jiaotong Yunshu Xitong Gongcheng Yu Xinxi/Journal of Transportation Systems Engineering and Information Technology, 10,* 18-23.

Velichkovsky, B.M., Dornhoefer, S.M., Kopf, M., Helmert, J.R., & Joos, M. (2002). Change detection and occlusion modes in road-traffic scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour, 5,* 99-109.

Wascher, E., Schneider, D., Hoffmann, S., Beste, C., & Sänger, J. (2012). When compensation fails: Attentional deficits in healthy ageing caused by visual distraction. *Neuropsychologia, 50,* 3185-3192.

White, C.B. & Caird, J.K. (2010). The blind date: The effects of change blindness, passenger conversation and gender on looked-but-failed-to-see (LBFTS) errors. *Accident Analysis & Prevention, 42,* 1822-1830.

Driving an EV with no opportunity to charge at home - is this acceptable?

Franziska Bühler[1], Thomas Franke[1], Peter Cocron[1], Katja Schleinitz[1], Isabel Neumann[1], Moritz Ischebeck[2], & Josef F. Krems[1]
[1] Technische Universität Chemnitz
[2] Humboldt University of Berlin
Germany

Abstract

In most field studies investigating acceptance of electric vehicles (EVs), participants could charge conveniently at home. Given the reality that EVs are often labelled a 'perfect city car' and city residents often do not possess their own garage or carport where charging would be possible, it is of special interest to determine if EVs are acceptable to those who depend entirely on the public charging infrastructure and if the evaluation of acceptability changes over time of EV usage.

Eighteen EV drivers in a 6-month field study relied on public charging stations in Berlin. Data were assessed at three data collection points (before EV acquisition, after 3 and 6 months of usage) and compared with a matched sample of 18 home-charging EV users in a comparable study. Results show that EV-related attitudes and purchase intentions did not significantly differ between the 2 groups at any point of data collection. Except for the general perception of EVs, no significant interaction effects of experience and charging circumstances were detected. Overall, we conclude that EVs also seem to be acceptable for city residents without private charging facilities. The implications for EV market expansion are evident.

Introduction

Electric vehicles (EVs, in this paper, EV is defined as pure battery electric vehicle that is powered only by electricity) represent a promising solution to rising CO_2 emissions (King, 2010). The German government aims to have 1 million registered electric vehicles on the road by 2020 (Die Bundesregierung, 2010). To reach this goal it is critical to have a large market for EVs. Therefore, it is important to determine the acceptability of EVs among likely consumers, particularly typical city residents who do not have a home-based charging option.

Charging circumstances and acceptance of EVs

EVs present additional considerations (e.g. limited driving range, charging) compared to conventional cars and various prejudices exist (Burgess, King, Harris, & Lewis, 2013). Research indicates that long charging duration (e.g. Hidrue, Parsons, Kempton, & Gardner, 2011) and unsatisfying charging infrastructure (e.g.,

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Egbue & Long, 2012) are just a few of the barriers to purchase decisions regarding EVs.

In a 6-month EV field study in which drivers could charge at home, drivers perceived the EV positively, were satisfied with the EV and evaluated it as useful throughout the study (Bühler, Cocron, Neumann, Franke, & Krems, 2013). After experiencing the EV, general evaluation of EVs and intention to recommend an EV were even more positive and likely, respectively. At the same time, they perceived charging, including handling of the cable and infrastructure, as less of a barrier and the home-charging opportunity was even reported as 'great' advantage by some participants after gaining EV experience. In sum, those users who experienced charging as easy were not perturbed by the charging duration and even preferred it compared to refuelling a conventional car (Franke & Krems, in press). However, in a 7-day trial, British drivers relying on public charging stations and 'normal' household sockets for recharging their plug-in hybrid electric car (PHEV) or EV (Graham-Rowe et al., 2012) stated that charging is simpler than expected, but still needs planning or even changing lifestyles due to long charging times and lack of infrastructure. Waiting for the car to be fully charged was considered a major disadvantage, because it reduces flexibility. PHEV and EV drivers also mentioned safety concerns while charging in public (e.g., EV drivers are quite vulnerable, because strangers can easily unplug the charging cable), whereas charging at home was highly valued by some participants. Consistent with these findings, home-charging was found to be an important advantage of EVs in further studies (e.g., Jabeen, Olaru, Smith, Braunl, & Speidel, 2012). In sum, charging an EV seems particularly suitable if EV drivers can charge at home. Yet, the typical inner city-resident has no private parking space at home where an EV can be easily charged. To our knowledge, there are no published studies that directly compare EVs acceptance given different charging circumstances.

Perception and acceptance of EVs

Different variables, for instance, attitudes (e.g., Gärling & Johansson, 1999), intention to recommend (e.g., Jabeen et al., 2012) and purchase intentions (e.g., Carroll, 2010), have been used to assess acceptance of EVs. Following Schade and Schlag's definition (2003) that acceptance is a person's attitudinal and behavioral reaction after experiencing a product, attitudes and behavioral intentions need to be assessed in order to make conclusions about acceptance. In the present study, attitudes are defined as "predispositions to respond, or tendencies in terms of 'approach/avoidance' or 'favourable/unfavourable' " (p. 2, Van der Laan, Heino, & De Waard, 1997) towards a product. Apart from attitudes and behavioural intentions, some authors investigated general perception of EVs (e.g., Burgess et al., 2013) to draw conclusions about how EVs are evaluated by different groups of potential car buyers. In sum, general perception of EVs, attitudes, intention to purchase and to recommend should be investigated in order to gain a deeper insight into how EVs are evaluated and accepted.

Study objectives

The objective of the present study was to investigate *if there are significant differences in perception, attitudes, and behavioural intentions between EV users that can charge at home and EV users that mainly charge at public charging stations.*

On the basis of the reviewed literature, it can be concluded that home-charging is perceived as an advantage (e.g., Bühler et al., 2013) and some concerns exist regarding charging in the public (Graham-Rowe et al., 2012). Furthermore, Jabeen et al. (2012) argue that the perceived convenience of home-charging is a significant predictor of acceptance. Relying on public charging stations alone is probably less convenient, more challenging regarding planning and could introduce further problems (e.g., parking space with the charging station is occupied). Therefore, we hypothesize that drivers who rely on public charging perceive EVs less positively (H1a), show less positive attitudes (H1b), lower intention to recommend (H1c) and lower purchase intentions (H1d) than home-charging EV drivers.

In previous studies, it was shown that evaluation of EVs is more positive (e.g., Carroll, 2010) and intention to recommend is higher (e.g., Bühler et al., 2013) after experiencing an EV. Although purchase intentions mostly did not change after real-life experience, we want to investigate if that is also the case for drivers charging in public. Thus, our second research question is *if changes in perception and acceptance of EVs depend on charging circumstances.*

Drivers who could charge at home seem to highly value home-charging and perceive charging as less of a barrier after experiencing an EV (e.g., Bühler et al., 2013). As public-chargers do not have the advantage of home-charging and might experience additional challenges while charging in public, it is expected that the changes in perception (H2a), attitudes (H2b), as well as behavioural intentions to recommend (H2c) and purchase (H2d) are not as positive as for home-charging EV drivers.

Method

Study design

The present paper aimed to investigate the influence of charging condition (home-charging [HC] vs. public-charging [PC]) on EV perceptions. Therefore, data of two comparable samples selected out of two large scale field studies (first study: Cocron et al., 2011; Krems, Weinmann, Weber, Westermann, & Albayrak, 2013; second study: Krems et al., 2011) conducted in the Berlin metropolitan area were analyzed. In the first large scale study, EV drivers were equipped with home-charging stations. In the second study, 20 participants were required to charge in public within walking distance of their home or work.

In order to match samples regarding their daily mobility, and therefore their prospective charging needs, data were collected via travel diary (i.e., person-based records of all trips in one week; Franke & Krems, 2013) before EV delivery were analyzed. Out of the 20 public-charging-participants, 18 stated at T_1 that they used public charging stations for >85% of charging events (the remainder used the normal socket). Based on their median daily covered distances, 18 home-charging users were selected from the second usage phase of the MINI E 1.0 project, because they drove the EV from spring to the end of summer, like the public-charging participants. Finally, two samples with similar median daily covered distances were selected (Table 1): home-charging users (HC) and public-charging users (PC).

In both studies, data were assessed three times: before receiving the EV (T_0), after 3 months (T_1) and when returning the EV after 6 months (T_2).

Participants

In sum, data of 36 participants (32 male, 4 female) were analyzed. These were on average 47.4 (*SD* = 11.2) years old. The majority of participants were highly educated (94% held a university degree). Participants lived in two adult (47%), three or more persons (42%), and single (11%) households. The majority of participants had a second car available during the study (one additional car: 49%; two: 23%; three or more: 9%). There were no significant differences between the two subsamples in age, gender, number of persons and cars per household (see Table 1). Participants in the PC group charged significantly less often than HC group (Table 1) and the remaining range when plugging in was significantly lower (Table 1).

Table 17. Socio-demographic, mobility and charging variables for the two subsamples: home-charging (HC) and public-charging users (PC).

variables		HC	PC	test of significant differences	*d*
Age (in years)	*M (SD)*	49.3 (10.5)	45.3 (11.8)	$t(34) = 1.08$, $p = .289$	0.19
Gender	Number	16 m, 2 f	16 m, 2 f	-	
Persons per household	*M (SD)*	3.2 (1.4)	2.5 (1.4)	$t(34) = 1.43$, $p = .163$	0.25
Number of cars	*M (SD)*	2.4 (0.7)	1.9 (1.0)	$t(32) = 0.67$, $p = .506$	0.12
Median of daily covered distances (in km) [1]	*M (SD)*	27.9 (9.9)	24.7 (16.4)	$t(34) = 0.72$, $p = .479$	0.12
Number of charging events per week[2]	*M (SD)*	3.4 (1.2)	2.0 (1.1)	$t(31) = 3.49$, $p = .001$	0.63
Remaining range when plugging in (in km)[2]	*M (SD)*	81.1 (34.4)	57.2 (28.7)	$t(31) = 2.17$, $p = .038$	0.39

Note. $N = 18$ for each group, [1] Data from travel diary (Franke & Krems, 2013a), [2] Data from charging diary (Franke & Krems, 2013b): $N_{HC} = 16$, $N_{PC} = 17$.

Scales and assessments

First, a scale regarding *General Perception* of EVs (Bühler et al., 2013) was used at all points of data collection (.64 ≤ Cronbach's α ≤ .82). The scale includes items that address different EV-related topics, such as the suitability of EVs for daily routines or the role of EVs in our transportation systems. A 6-point Likert scale from 1 (*completely disagree*) to 6 (*completely agree*) was applied for this scale, as well as for all intention items. Second, the *Van der Laan Acceptance Scale* (Van der Laan et al., 1997), an instrument that contains two dimensions (*Satisfaction* and *Usefulness*), was used to assess acceptance. Four of nine semantic differentials (ranging from -2 to 2) comprised the *Satisfaction* scale (.72 ≤ Cronbach's α ≤ .83). The other five items represent the *Usefulness* scale (.70 ≤ Cronbach's α ≤ .81). Third, one item was utilized to assess the *Willingness to Recommend* an EV (Bühler et al., 2013). Fourth, three items assessed purchase intentions. As in Bühler et al. (2013), one item assessed the *Willingness to Purchase* and two items assessed the *Willingness to Pay* (.62 ≤ Cronbach's α ≤ .72). The two *Willingness to Pay* items were anchored on realistic leasing rates (650€ per month) and purchase prices for EVs (1/3 more) that are comparable to the test vehicle in performance.

Test Vehicle

The test vehicle was a converted MINI Cooper, the MINI E (two-seater, 150 kW power, 220 Nm torque, top speed of 150 km/h, range of approximately 168 km on a single charge under 'normal driving conditions'). A lithium-ion battery pack stored the power and was rechargeable using 32 and 12 Ampere. Besides using the public charging stations that were available in Berlin, EV users in the first study could recharge at home using a "wallbox". Users in the second study were required to charge at public charging stations close to their permanent place of residence or work. An empty battery took approximately four hours (32 Ampere) to charge. For more details see Keinath and Schwalm (2013).

Results

For analyzing the data, mixed ANOVAs were calculated. Effect sizes were interpreted according to Cohen (1988).

General perception

EV drivers had a positive view of EVs (Figure 1). The PC group (M = 4.69, 95% CI [4.41, 4.98]) and the HC group (M = 4.69, 95% CI [4.46, 4.92]) perceived EVs positively and perception did not differ significantly between groups, $F(1, 28) = 0.00$, $p = .992$, $\eta^2_p = .000$.

As displayed in Figure 1, changes in the evaluation of EVs were different in the two groups. At T_0, the PC group showed lower mean scores in the beginning of the study than the HC users. After 3 months, HC users' perceptions did not differ significantly from T_0, whereas PC users viewed EVs more positively. At T_2, both groups showed their highest approval and had similar means. Results revealed a significant medium sized interaction effect of experience and charging condition, $F(2, 56) = 3.34$, $p = .042$, $\eta^2_p = .107$. As the increase for the PC group from T_0 (M = 4.23, 95% CI [3.84, 4.63]) to T_2 (M = 4.98, 95% CI [4.68, 5.29]) is higher than for the HC users (T_0: M = 4.69, 95% CI [4.19, 4.83]; T_2: M = 5.01, 95% CI [4.76, 5.26]), the results did not support hypothesis H2a.

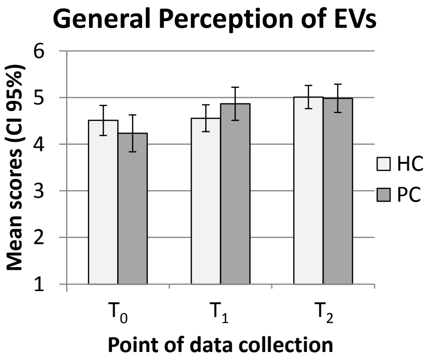


Figure 1. Results of General Perception of EVs

Note. $N_{HC} = 18$, $N_{PC} = 12$. *6-point Likert scale.*

Satisfaction and Usefulness

Participants were on average satisfied with the EV and evaluated the vehicle as useful (Figure 2). The *Satisfaction* scores of the PC group ($M = 1.31$, 95% *CI* [1.01, 1.55]) were lower than the scores of the HC group ($M = 1.31$, 95% *CI* [1.32, 1.78]), but this difference was not significant, $F(1, 33) = 2.81$, $p = .103$, $\eta^2_p = .08$. For *Usefulness* scores a similar pattern is displayed in Figure 2 (HC: $M = 1.31$, 95% *CI* [1.07, 1.55], PC: $M = 1.31$, 95% *CI* [0.90, 1.39]) and again, no significant effect was observed, $F(1, 33) = 1.37$, $p = .250$, $\eta^2_p = .04$. These results do not support hypothesis H1b.

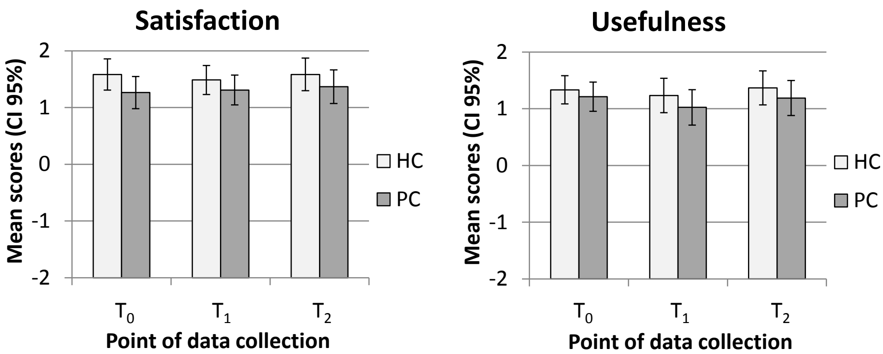


Figure 2. Results of the Van der Laan Acceptance Scale (Van der Laan et al., 1997).

Note. $N_{HC} = 18$, $N_{PC} = 17$. Semantic differentials from -2 to 2.

Inconsistent with hypothesis H2b, descriptive statistics displayed in Figure 2 did not appear to indicate an interaction effect. This visual impression was supported by the results of a mixed ANOVA; no significant interactions were found, *Satisfaction* – $F(2, 66) = 0.46$, $p = .634$, $\eta^2_p = .01$ and *Usefulness* – $F(2, 66) = 0.87$, $p = .868$, $\eta^2_p = .00$.

Intention to recommend

EV drivers were willing to recommend the EV (Figure 3). The PC group (M = 4.77, 95% CI [4.43, 5.01]) was on average less willing to recommend an EV than the HC group (M = 5.13, 95% CI [4.81, 5.45]), but this difference was not significant, $F(1, 33) = 2.55$, $p = .120$, $\eta^2_p = .08$, and therefore, hypothesis H1c was not supported. In both groups, the intention to recommend was higher after EV experience (Figure 3). The increase was similar and results revealed no significant interaction of experience x charging condition, $F(2, 66) = 0.60$, $p = .554$, $\eta^2_p = .02$.

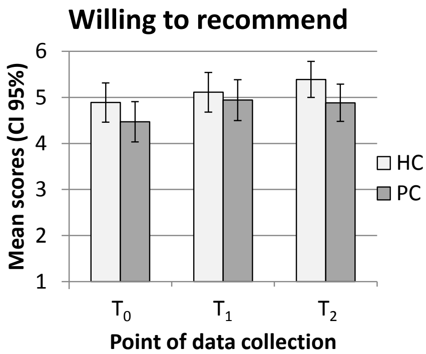


Figure 3. Results of Willingness to Recommend.

Note. N_{HC} = 18, N_{PC} = 17. 6-point Likert scale.

Purchase intentions

On average, participants were undecided as to whether they would purchase an EV after the study (Figure 4, left) and were not willing to pay the given prices for an EV (Figure 4, right).

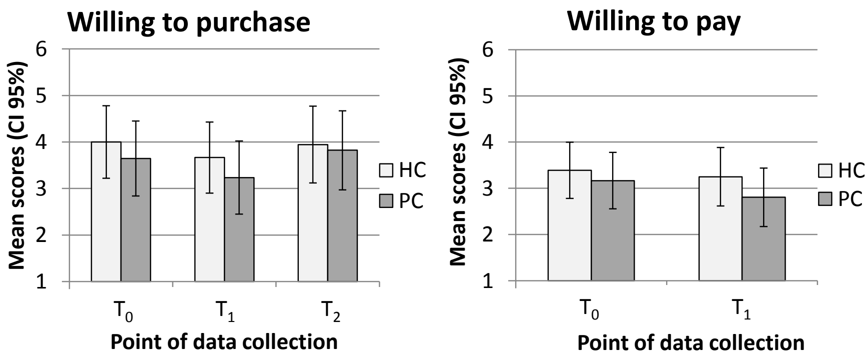


Figure 4. Results of Willingness to Purchase (left) and Willingness to Pay (right).

Note. N_{HC} = 18, N_{PC} = 17. 6-point Likert scale.

The HC group ($M = 3.87$, 95% *CI* [3.16, 4.58]) was on average more willing to purchase an EV than the PC group ($M = 3.57$, 95% *CI* [2.84, 4.30]), but this difference was not significant, $F(1, 33) = 0.49$, $p = .491$, $\eta^2_p = .01$. PC users ($M = 2.99$, 95% *CI* [2.44, 3.53]) were on average less willing to pay than users that could charge at home ($M = 3.32$, 95% *CI* [2.77, 3.87]). Again, results of the ANOVA revealed no significant effect of charging condition, $F(1, 34) = 1.03$, $p = .318$, $\eta^2_p = .03$. Overall, none of the results supported hypothesis H1d.

Hypothesis H2d was also not supported by the results. With increasing experience, there was no between-group difference on level of *Willingness to Purchase* an EV. Again, no significant interaction was found, $F(1.675, 55.275) = 0.38$, $p = .650$, $\eta^2_p = .01$. Although, the *Willingness to Pay* seems to decrease more for the PC group (Figure 4, right), the mixed ANOVA showed that there is no significant effect for the interaction of experience x charging condition, $F(1, 34) = 0.38$, $p = .539$, $\eta^2_p = .01$.

Discussion

In the present study, data from two comparable samples from two large-scale field studies were analyzed in order to investigate whether EVs are acceptable to users when they cannot charge at home and if there are differences in acceptance due to charging condition. As the typical inner city-resident has no private parking space at home where an EV can be conveniently charged, this question is highly important for predicting potential EV acceptance and expansion of market share.

Different indicators of acceptance (i.e., attitudes, intention to recommend, purchase intentions) were assessed and a significant effect of home charging versus public charging (charging conditions) was not observed. Although results were not significant, the strength of effects is worth discussing, especially when taking into account the small sample size. Charging condition only shows non-significant *medium* effects on satisfaction and the intention to recommend EVs. The rest of the effects were small. Therefore, the effect sizes do not indicate that there is a large difference between the potential acceptance of EVs for consumers who can charge at home and consumers who can only recharge their EVs at public charging stations. However, one difference between the groups was clear: home-chargers charge more often than public-chargers. At the same time, the remaining range in EVs of public-chargers is much lower than of EVs of home-chargers. One possible explanation is that public-chargers try harder to fully discharge the battery before recharging, because the public charging procedure requires more effort.

Our second research question was if changes in perception and acceptance are different for users who have to charge under different charging conditions. Results indicate that there is an interaction effect of experience and charging condition on the EV perception. However, this effect suggests that changes over time are stronger for participants who charge in public. Moreover, the observed effect was opposite to that which was hypothesized (H2a: for public-chargers changes in perception were expected to be not as positive as for home-charging EV drivers). These findings highlight another interesting point – when it comes to changing perceptions and prejudices regarding EVs, experiencing an EV that can only be charged in public might be even more important than experiencing an EV with private charging facilities. However, given the fact that there were many missing values and the number of analyzed participants was small, data should be carefully interpreted and this interaction effect should be further investigated. Furthermore, results regarding acceptance do not suggest an interaction effect of experience and charging conditions. Significant interaction effects were not observed for either attitudes or behavioural intentions. Additionally, effect sizes were small.

The results observed here may not be generalisable to the population of German car drivers as a whole, because early adopters were most likely overrepresented in the presented field studies (Rogers, 2010). Nevertheless, studies with early adopters provide important implications for the adoption process of new products.

Overall, the presented results are the first empirical evidence suggesting that access to home charging stations might not dramatically influence EV acceptance. Thus, the lack of an opportunity to charge at home should not reduce potential consumers' intention to purchase an EV. Yet, the sample in this study had relatively low daily mileages. Generalisability to populations with higher daily mileage requirements (e.g., commuters) should be studied in future research.

Acknowledgements

The two projects were funded by the German Federal Ministry for the Environment, Nature Conservation and Nuclear Safety. We thank our consortium partners Vattenfall Europe AG (Dr. C. F. Eckhardt and F. Schuth) and the BMW Group (Dr. M. Hajesch, Dr. A. Keinath, Dr. R. Vilimek and Dr. M. Schwalm) who gave us the opportunity to conduct our research.

References

Bühler, F., Cocron, P., Neumann, I., Franke, T., & Krems, J.F. (submitted). Is EV experience related to EV acceptance? Results from a German field study. *Manuscript submitted for publication*.

Burgess, M., King, N., Harris, M., & Lewis, E. (2013). Electric vehicle drivers' reported interactions with the public: Driving stereotype change? *Transportation Research Part F: Traffic Psychology and Behaviour, 17*, 33-44.

Carroll, S. (2010). *The Smart Move trial: description and initial results*. Retrieved 08th September , 2013, from http://www.cenex.co.uk/LinkClick.aspx?fileticket=yUKAcRDJtWg%3d&tabid=119&mid=695.

Cocron, P., Bühler, F., Neumann, I., Franke, T., Krems, J. F., Schwalm, M., & Keinath, A. (2011). Methods of evaluating electric vehicles from a user's perspective - The MINI E field trial in Berlin. *IET Intelligent Transport Systems, 5*(2), 127-133.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale: Lawrence Erlbaum Associates.

Die Bundesregierung (2010). *Etablierung der Nationalen Plattform Elektromobilität - Gemeinsame Erklärung von Bundesregierung und* *deutscher Industrie [Establishing the National Platform Electro-mobility – Joint statement of the federal government and industry]*. Retrieved 08th Sept, 2013, from http://www.bundesregierung.de/Content/DE/Artikel/2010/05/2010-05-03-elektromobilitaet-erklaerung.html.

Egbue, O., & Long, S. (2012). Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy Policy, 48*(0), 717-729.

Franke, T., & Krems, J.F. (2013a). What drives range preferences in electric vehicle users? *Transport Policy*, *30*, 56-62.

Franke, T., & Krems, J.F. (2013b). Understanding charging behaviour of electric vehicle users. *Transportation Research Part F: Traffic Psychology and Behaviour*, *21*, 75-89.

Gärling, A., & Johansson, A. (1999). *An EV in the family*. Göteborg: Department of Road and Traffic Planning. Chalmers University of Technology, Göteborg.

Graham-Rowe, E., Gardner, B., Abraham, C., Skippon, S., Dittmar, H., Hutchins, R., & Stannard, J. (2012). Mainstream consumers driving plug-in battery-electric and plug-in hybrid electric cars: A qualitative analysis of responses and evaluations. *Transportation Research Part A: Policy and Practice*, *46*, 140-153.

Hidrue, M.K., Parsons, G..R., Kempton, W., & Gardner, M.P. (2011). Willingness to pay for electric vehicles and their attributes. *Resource and Energy Economics*, *33*, 686-705.

Jabeen, F., Olaru, D., Smith, B., Braunl, T., & Speidel, S. (2012). Acceptability of electric vehicles: findings from a driver survey. *Proceedings of the ATRF (Australasian Transport Research Forum)*, Sep. 2012, Perth, Australia. Retrieved 8th Febr, 2013, from http://www.atrf.info/papers/2012/2012_Jabeen_Olaru_Smith_Braunl_Speidel.pdf.

Keinath, A. & Schwalm, M. (2013). MINI E 1.0: Der Berliner Feldversuch [MINI E: Berlin's field study]. In J. F. Krems, O. Weinmann, J. Weber, D. Westermann, and S. Albayrak (Eds.). (2013). *Elektromobilität in Metropolregionen: Die Feldstudie MINI E Berlin powered by Vattenfall [Emobility in metropolitan areas: the field study MINI E Berlin powered by Vattenfall]* (Vol. 12, pp. 34-47). Düsseldorf: VDI Verlag.

King, D. (Ed.). (2010). *Future of Mobility Roadmap - Ways to reduce emissions while keeping mobile*. Oxford: University of Oxford.

Krems, J. F., Weinmann, O., Weber, J., Westermann, D., & Albayrak, S. (Eds.). (2013). *Elektromobilität in Metropolregionen: Die Feldstudie MINI E Berlin powered by Vattenfall [Emobility in metropolitan areas: the field study MINI E Berlin powered by Vattenfall]* (Vol. 12). Düsseldorf: VDI Verlag.

Krems, J. F., Bartholdt, L., Cocron, P., Dielmann, B., Franke, T., Henning, M., et al. (2011). *MINI E powered by Vattenfall 2.0* Schlussbericht zum Forschungsvorhaben Verbundprojekt MINI E powered by Vattenfall 2.0. Chemnitz: Technische Universität Chemnitz.

Rogers, E. M. (2010). *Diffusion of innovations* (5th ed.). New York: Free Press.

Schade, J., & Schlag, B. (2003). Acceptability of urban transport pricing strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 6*, 45-61.

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.

ASSISTANT - Creating a Smartphone App to assist older people when travelling

Philip Barham
Transport & Travel Research Limited
UK

Abstract

ASSISTANT is a three-year project, funded jointly by the European Commission's Ambient Assisted Living Joint Programme (AAL JP) and the national funding agencies of Austria, Finland, France, Spain and the UK. The project's objective is to create a smartphone app that can help older people to travel independently by public transport, in comfort and with confidence. The target group for the commercial product that will be developed by the project is more mobile older people. This is in order to make the product appealing to the largest market sector possible, for when the app enters the market, in 2017. However, because the app will be highly personalised, and provide the user with the facility of audible and haptic, as well as visual, feedback, it will be appropriate for use by people who are blind and partially sighted, people who are deaf or hard of hearing, and people who have difficulty with oral communication. This paper describes progress made in the Concept Phase of the project, during ASSISTANT's first year, and outlines some of the ways in which the ASSISTANT app will be tailored to meet the needs of a variety of users, in a variety of public transport contexts.

Introduction

The theme of the Fourth Ambient Assisted Living Joint Programme is "ICT based solutions for advancement of older persons' mobility". The ASSISTANT Project's contribution to this objective is to produce an app which provides accessible support for older people when using public transport. This app will provide an online facility for trip planning, guidance during multi-step journeys on public transport vehicles – which will include assistance with transfers from one vehicle to another, and between different means of transport – and navigational guidance from the user's last stop on the public transport network, to his or her final destination.

As such, ASSISTANT addresses common, everyday concerns that many older people have with not being able to find the correct stop, with not knowing which vehicle to board, with failing to get off the vehicle at the correct location, and, ultimately, with getting lost. The app will seek to bridge this confidence gap during the planning and making of a journey by public transport, and will also provide a personalised "safety net" feature that will enable the user to readily summon help from a relative or carer of his or her choice, if necessary. Ultimately, the

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

contribution made by ASSISTANT will be to encourage the mobility, and thus the social participation, of older people in Europe, enabling them to freely access important goods and services, and fulfil their social and cultural needs, using more sustainable means of transport.

The main target group of the ASSISTANT project is more mobile older people, particularly those travelling to unfamiliar places or who might, say, be using public transport after losing the ability to drive. The decision to concentrate the project's design efforts on this sector of the population was taken for commercial reasons, to ensure that the project's end-product would be attractive to as large a market as possible. However, because the app will be highly personalised, and provide the user with the facility of audible and haptic, as well as visual, feedback, it will be appropriate for use by people who are blind and partially sighted, people who are deaf or hard of hearing, and people who have difficulty with oral communication. This high level of accessibility is made possible by the fact that smartphones already have the functionality of allowing people with a range of sensory impairments to customise their 'phone to suit their particular needs. If, for example, the smartphone user is blind, and has a screen reading device that enables him or her to read web pages using the smartphone, then this will enable this user to interact with the ASSISTANT app in the same way.

In terms of the business model to be adopted, ASSISTANT will be a software product that is purchased once, either through an online shop, or at a retail outlet. Technology Platform Providers will customise the app, whilst service operators will run the backend components of the system, ensuring the smooth running of all of the system components.

The ASSISTANT online application (app) - Technical characteristics

Whilst the design of the ASSISTANT app will have innovation at its core, its use will be designed for, and based on, tried and trusted devices, namely the home Personal Computer and the smartphone. This provides the system with a robust and reliable basis. The user's smartphone will be the medium through which information is presented, and the application running on a PC will enable route design. ASSISTANT will be of use even on well-known routes, since the haptic mode reminder of arrival at the desired destination or transfer point can allow the user to read or relax when travelling, removing the need to constantly check the progress of the vehicle.

ASSISTANT will create a web server based application that will function by retrieving continually updated data in the form of maps and schedules. This will be part of a three-tier system architecture, consisting of a web based route design interface, a web server layer implementing the application and a database of system data. User-specific data will also be stored on the smartphone, which will have the capability to act as a Personal Navigation Device and mobile interface for the user. The route editor interface will enable users to update their profile data, and to create and edit a list of favourite locations to visit, and design pathways to them.

An important characteristic of the ASSISTANT product will be the high degree to which it can be personalised according to the user's ability levels, and contextualised according to the situation in which he or she is in, during any part of the journey. ASSISTANT will provide a high level of contextualisation by only providing information that is relevant to the user, at the right time and in the appropriate format. This is achieved by filtering available data, using pre-set information about the user's preferences, and then communicating these data to the user via audio, visual and haptic cues.

The product will provide further personalisation through its error trapping and remediation functionality, since it will be designed to be responsive to the user's physical, cognitive and mental capacities and preferences. The definition of an "error", in this context, is an unplanned situation caused by the user not getting off the bus at the correct stop, for example, or getting lost or anxious when searching for a bus stop. "Error trapping" describes the process of detecting that the user is not following the route intended, (Schank & Abelson, 1977). ASSISTANT's "design for failure" approach to design relates to both the possibility of human error, and the failure of system components, and so acknowledges the fragility of mobile devices, the brittleness of digital information and the limitations of location information provided by Global Positioning System (GPS) technology. The app will use location, system state, and user interaction behaviour as a basis for Artificial Intelligence based error detection. Once an error type has been detected, the user model will determine the appropriate mediation strategy. Such mediation will be based on user-specific replanning of routes, in the event of divergence from the planned route, using pre-set personalised strategies.

Human Factors considerations in relation to the Target Audience

A major effort will be made to ensure that the user interfaces, and functionality, of the ASSISTANT system are as suitable as possible for older people. An important component of ensuring accessibility of the product will be the filtering of data to enable the user to be provided with only the information that is necessary, given the context in which the information is provided, and the user's stated personal preferences. Furthermore, information will be presented in an accessible manner, being designed with an uncluttered screen, larger buttons, an intuitive representation of screen elements and a reduction in the number of menus and drop-down elements that are usually associated with smartphones. Maximum accessibility will be achieved through the involvement in the project of volunteers representing the device's target population group, at each stage of the product's development.

In terms of the type of information that the user will require, it is acknowledged that older people require the same information when travelling, as anybody else, although, in addition, the client will be provided with information on potential physical barriers to travel, on the accessibility of specific platforms and stations, and on staffing levels and the availability accessible toilets etc., as far as data availability allows.

The involvement of end-users in the design of the ASSISTANT app

A fundamental aspect of ASSISTANT's design approach has been to involve primary end-users in the process from the beginning of the project, engaging with older people in iterative evaluation trials in three cities: Vienna, San Sebastián and Paris. The same "panel" of end-users is to be involved in evaluations for each successive phase of the project, i.e. the Concept, Pilot and Prototype phases. The same group of volunteers, in each country, took part in a focus group, at the beginning of the project, designed to gain insights into their use of both public transport, and current assistive technologies, particularly mobile 'phones and smartphones. An important goal of this early phase of the research was to investigate the use of mobile communications technologies by older people in their daily, travel-related routines. It was important to establish the priorities, needs and acceptance factors of this diverse group of people, so that design of the app can be guided from a user's perspective, using a needs framework.

The sample of 30 people, over three countries, was structured so that at least 90% of the participants were public transport users – with half of the sample being frequent users – and at least 50% of the sample used a mobile 'phone. All participants were aged 55 or over, reflecting the envisaged market for the ASSISTANT app, and there was an even gender split.

The objective of the concept phase evaluations was to inform the redesign of the ASSISTANT user interface, with there being the opportunity for similar adjustments to be made after each feedback loop, up to the use of low fidelity prototypes in the third, and final, year of the project. These evaluations were carried out using a focus group format, with the aid of "life-size" mock-ups of user interfaces presented on a computer screen. Figure 1 provides examples of these user interface mock-ups.

Figure 1. Example of user interface mock-ups used during the Concept Phase.

These mock-ups enabled participants to interact with each interface in much the same way that they would interact with the touchscreen of a smartphone, with the clicking of the various "buttons" provided taking them to a different screen. Using this approach, it was possible for the participants to, for example, plan hypothetical journeys by public transport.

Findings from Qualitative Research and User Evaluations conducted to date

A common issue for users with less familiarity with computers, (not just older users), is at the intersection of two computational systems. A typical example of this is linking the use of a digital camera with a PC. The ideal solution to this issue is to prevent less experienced users from having to perform complex configuration tasks by providing a single application that provides all the functionality needed to connect the two devices. ASSISTANT will address this potential barrier to use of the project's app by making sure that both the PC and the smartphone are easy to use, with seamless data transfer and transparent user interfaces.

Another strong theme to emerge from the sample of potential users was people's concern for personal security when carrying and using an expensive item of equipment such as a smartphone. There was particular concern at the prospect of conspicuously using such a device on some urban public transport networks, especially at night. This represents encouraging feedback, in as much as a major selling point of the ASSISTANT system will be the facility for it to be used, with its options for tactile and/or audible output, whilst safely concealed in a pocket or a hand-bag.

Another reservation expressed by some participants was the requirement for the system's GPS function to be active continuously, as this gave rise to some concerns for privacy.

A general conclusion that was drawn from the evaluation of alternative user interface designs was that there was a preference for a simple, uncluttered screen. For the app's personal navigation function, the majority of participants expressed a preference for step by step, text instructions, although there were some who appreciated the facility to switch to a map display. Where a map is used, requests were made for a blinking "You Are Here" symbol indicating the user's position. Furthermore, there was a general preference for two levels of zoom, whereby an overview map is followed by more detailed instructions.

Negative feedback received from the evaluations included complaints about having to switch to using reading glasses for accessing on-screen information, and having to hold a smartphone device in the hand whilst travelling – but participants acknowledged that the facility to obtain audible instructions, using a Bluetooth headset, was a solution to both problems. However, the issue of not being able to read small icons and text without the use of reading glasses is one which is very relevant to the envisaged market sector at which ASSISTANT will be aimed, and so this is something that will be considered during the design of the visible user interface. Some participants complained that the touchscreen was too sensitive to light touches which might be involuntary, and so this is an additional design factor that will be taken into account.

There was also a fairly negative reaction to having photos of the vicinity of a bus stop or destination, accompanying a text instruction. Such illustration was seen as being superfluous and potentially confusing and, in cases where the illustrating photo was not of the stop itself, it was felt that it might be misleading.

There was a marked preference for the actual name of a stop to be specified, whenever it is identified as the stop where the user should get off the vehicle, (with compass directions never used); when the app indicates that the user should board a vehicle, it was felt that the app should always indicate to the user the name and number of the vehicle, and its direction of travel or ultimate destination. There was evidence to suggest that some people were not comfortable with instructions being given in terms of metres to a required vehicle or destination.

Issues raised with reference to the route editing function of the system, which will be designed to be executed using a home PC, included the legibility of text for street names on map displays. This highlighted the importance of text remaining of sufficient size to be legible once the user has changed the scale of the display. There was also evidence among the sample of participants that there was a preference for using an address expressed as text as a means of specifying a desired destination, as opposed to selecting a point on a map display. The least preferred mode of input was the use of coordinates. The participants found some of the terminology used in the mock-up of the route editor to be confusing – including words such as “font” – so care will be taken to use plain language in subsequent designs. Generally, there was an appreciation for the facility to adjust parameters such as font size, colour and contrast, as well as output such as volume and haptic cues. Another preference expressed was for alternative travel times to be expressed according to whether the user is a fast or slow walker.

The evaluations presented some very interesting design conundrums, which are currently the subject of intense discussion among the ASSISTANT Project partners. For example, there was some confusion surrounding buttons labelled “Help”, since it was not clear whether the button should be used to obtain fairly low-level information on how to perform tasks on the smartphone, or was for calling for help in an emergency. Further work is required on defining the optimum label for such functions, the challenge being one of devising labels that are unambiguous and describe a function precisely, whilst at the same time fitting comfortably in the space that is available on the screen. Feedback from participants suggested that any terms and/or icons used should be either familiar or intuitive, to avoid any learning curve that might be required with new constructs.

Next steps

The next step for the ASSISTANT Project will be to proceed, during 2014, to the Pilot Phase. Further development of the system will proceed, taking account of the findings of the qualitative questionnaire survey and the Concept Phase evaluations outlined above. System modules, including the route editor and the personal navigation facility, will be evaluated as stand-alone components, given that complete system integration is not due to be achieved before the Prototype Phase. Evaluations conducted during the Pilot Phase will feature more advanced tasks and more detailed user interfaces, and will also focus on ergonomic, rather than utility, aspects.

Acknowledgement

The ASSISTANT Project is part of the Ambient Assisted Living Joint Programme (AAL JP, No. AAL-2011-4-040), and so is funded jointly by the European Commission and the national funding agencies in participating countries. The consortium of project partners is led by the Project Coordinator, Fundación Tecnalia (Spain). Other project partners are Andago Ingeniería SL (Spain, 2012 only), Citruna Technologies Oy (Finland), E-Seniors (France), Fara Oy (Finland), Transport & Travel Research Ltd (UK), the University of Vienna (Austria) and VTT Technical Research Centre (Finland), with IN2 involved as a Sub Consultant.

References

Schank, R.C., & Abelson, R.P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.

“ecoDriver” HMI feedback solutions

Antonella Toffetti[1], Alessandro Iviglia[2], Claudio Arduino[1], & Mirella Soldati[1]
[1]Centro Ricerche Fiat (CRF)
Orbassano (Torino), Italy
[2]Fiat Group Automobiles
Torino, Italy

Abstract

The European Project *ecoDriver - Supporting the driver in conserving energy and reducing emissions*, was aimed at identifying an appropriate HMI to be given to the driver to suggest him/her how maintaining an eco-drive style. Two different testing activities involving forty users were held. The first activity, conducted in the CRF Usability Laboratory and involving twenty users aimed to evaluate adequacy and comprehensibility of different types of visual feedback, graphic layouts, kind of personalisation and rewarding solutions to be used in an eco-driving system. The first phase results were used to identify visual stimuli for the second part of the activity, conducted in the CRF Virtual Reality Driving Simulator. Twenty users were asked to drive for about an hour, using an eco-drive system which gave, when appropriate, visual advices and feedback or visual stimuli plus a haptic feedback on the gas pedal. Performance data regarding driving behaviour and fuel consumption were recorded during the driving sessions. Moreover, subjective evaluations were collected using several questionnaires, to identify the best feedback solutions in terms of usability, acceptability and perceived mental effort. Collected data (objective and subjective) allowed to identify HMI solutions that can lead to important contributions to the fulfilment of an eco-drive style and to a real benefit in reducing consumption, without making drive too annoying or unsafe.

Introduction

The ecoDriver project (http://www.ecodriver-project.eu) targets a 20% reduction of CO2 emissions and fuel consumption in road transport by encouraging the adoption of green driving behaviour, through delivering those most appropriate eco-driving advice or feedback for any given situation. ecoDriver aims to deliver drivers the most effective feedback on green driving, optimizing the driver-powertrain-environment feedback loop. Drivers will receive eco-driving recommendations and feedback adapted to them and to their vehicle characteristics.

It is an important project that addresses the need to consider the human element when encouraging “green” driving.

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Several HMIs, developed by different partners in ecoDriver, following common defined guidelines, are tested in 2014 in long term naturalistic or controlled studies, following FESTA methodology (FESTA Handbook, 2011) regarding Field Operational Tests, to understand aspects like perceived mental effort, usability and acceptability vs. on-field consumption. The CRF HMIs were defined following the User-Centred Design approach (ISO 9241-210:2010) through different phases of qualitative analysis to highlight users' eco-driving mental model, collection of creative ideas on innovative eco-HMI, parallel visual icons design, to arrive to the evaluation phases described in this paper.

Usability laboratory HMI evaluation

This laboratory study was aimed at identifying the types of visual interfaces that could guarantee most adequate solutions for displaying eco-driving information to the driver. A parallel design of several kinds of icons, representative of different events during the drive, visual advice and feedback on fulfilment of ecodrive behaviour, graphical layout, personalisation and rewarding options concerning an eco-drive system were developed (Jenness, Singer, Walrath and Lubar, 2009). Designed solutions were tested with users in the CRF Usability Laboratory.

Method

Participants
Twenty CRF employees (6 women, 14 men), not belonging to technical departments, took part in the experiment with ages between 29 and 60 years. The majority of the group (70%) had a university degree, 20% had a high school certificate and 10% a post-graduate degree. On average, they had their driving licence for 19.75 years, with an annual mileage of 17050 km/year. Participants drove almost every day in an urban context, less often on suburban road, highways, and country-side or mountain roads.

All participants were frequent users of mobile phones and personal computers while other devices (i.e. smartphones, MP3 players, portable or on-board navigators) were less used. The percentage of users that had rarely used an integrated navigation system (70%) were particularly high, lower were those of participants that rarely (25%) or never (50%) used a portable navigation system. The majority of participants (65%) never experienced driving with an eco-system and moreover, most of the users that had previous experience with eco-systems used them everyday (43%).

Stimuli
Several visual stimuli (Figure 1) representing information gdisplayed by an eco-driving system were shown to participants (Manser, Rakauskas, Graving, and Jenness, 2010). In particular:

- Fifteen visual stimuli on events during the drive (e.g. road signs), advice and feedback on eco-driving behaviour, given in different way (e.g. static icons, continuous bars…);

- Four types of feedback solutions to notice the driver about the compliance or not with an eco-driving behaviour;
- Three levels of an eco-system layout complexity with increasing number of information;
- Three solutions of display personalisation with different kinds and style of eco-information displayed on it

Moreover, a list of rewarding solution highlighted from previous studies was proposed.

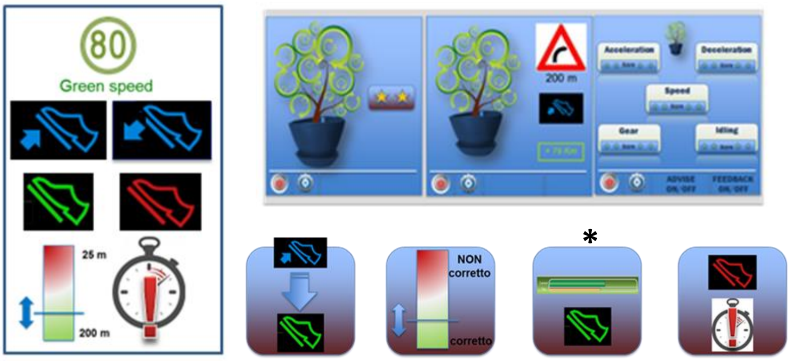


Figure 1. Examples of visual stimuli.

**Bar developed by TNO who gave kind permission to ecoDriver partners to test it.*

Experimental design

A within-subjects design was conducted. The different stimuli belonging to the same class (e.g. single visual stimuli, types of feedback, personalisation solutions…) were randomized in order to avoid learning and sequence effects.

Procedure and questionnaires

First participants were welcomed by the experimenter and briefed about the test they were invited to participate and then required to fill in a short socio-demographic questionnaire. After that the different stimuli were shown and participants were requested to evaluate those, using ad hoc developed questionnaires.

As a first step, participants were shown all the single images (15) with no explanations about their meaning and they were asked to explain their meanings to verify whether icons were self-explanatory or not. After this, all icons were explained and then participants evaluated their comprehensibility and the adequacy to represent the message (Campbell et al., 2004).

Three different types of eco-driving information layout with increasing level of complexity were shown to participants, who were requested to explain the meaning of the icons present in the different visualisation solutions. Then, each layout solution was explained and participants evaluated some their characteristics (e.g. comprehensibility, adequacy, utility, etc…).

Four feedback solutions (correct or wrong fulfillment of the action suggested by the interface for driving in an eco-mode) were shown and explained to participants, who had to indicate their preferred solution.

Immediately after that, participants evaluated different types of hypothetical “reward”, given eventually after the correct eco-driving behaviour. Participants had to choose three preferred rewards and answered questions about possibility of rewarding to enhance the use of an eco-system, to distract the driver if presented on the display and about willingness to have a reward.

Following that, participants were asked to evaluate the possibility to “customize the graphics” showing eco-information on the display. Three hypotheses were shown in which the customization was on number, type and style of eco-driving information. Participants judged different aspects such as the possibility to choose by themselves among the three different cluster configurations, the possible distraction induced, the adequacy and utility of the different displayed information. After the trial the experimenter thanked and said goodbye to participants.

Results

Participants considered most of the different types of visual icons presented as comprehensible and adequate. Only graphics which displayed a time related concept received negative judgements. Feedback by icons was preferred if displayed with other visual signs (e.g. labels, bars, time). The TNO bar was the preferred one among the tested bars types, because it displayed both current and average trip eco-scores.

The complete interface that showed eco-driving detailed information received the most positive votes, but from participants comments emerged that they would prefer to have a medium level of eco-information complexity, when the vehicle is in motion, to reduce distraction.

According to participants’ votes a reward can contribute to enhance the eco-driving behaviour [$t=3,1$, $p<.05$]. Anyway they had a neutral position on the possibility to have rewarding directly displayed on the on-board instrumentation.

In general, participants evaluated positively the possibility to have personalization features in their vehicles [$t=2,7$, $p<.05$]. They did not think eco-system would be distracting (this value is significantly different from the neutral point (3)) and they did not want to have a system in which visualization automatically changes while driving [$t=2,1$, $p<.05$] (Figure 2).

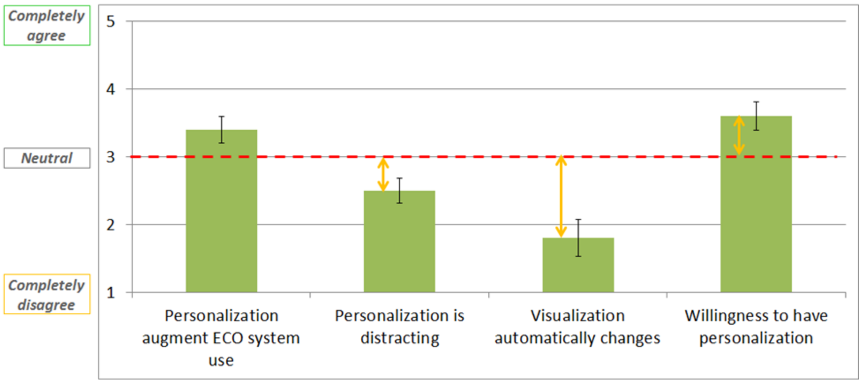


Figure 2. Evaluation of personalisation features.

From users' comments emerged that the personalization is appreciated if the selectable visual solution, shown when driving, is not too crowded.

HMI solutions highlighted as the best ones in the Usability Laboratory study were used to define the visual HMI adopted in the second study, in the CRF Virtual Reality Driving Simulator.

Virtual Reality Driving Simulator User Testing

The aims of the study were to identify the feedback modalities (*Visual* or *Visual & Haptic*) that guarantee an effective eco-drive performance, to identify and compare the consumption with eco-drive system off vs. eco-drive system on and to identify the eco-drive interface solutions that better guarantee an adequate compromise among consumption, perceived mental effort, usability and acceptability (Fricke & Schießl, 2011).

The experiment consisted of three driving simulator sessions, during which each participant drove in two different scenarios with and without the eco-drive system which gave, according to the experimental condition, visual feedforwards and *Visual* or *Visual&Haptic* (a counterforce on the gas pedal) feedback (ISO/TR 16352:2005) to the driver in relation to his/her driving behaviour.

Method
Participants
Twenty people were recruited for the experiment (mean age = 37 years), of which 75% was male and 5% female. Participants were recruited from CRF employees who had non-technical professions. Only participants with a low score on the motion sickness questionnaire were invited for the test in the Virtual Reality Simulator.

The large majority of the sample had a university degree (85%); 10% had a high school degree and 5% a post-graduate degree.

On the average, the participants held a valid driving licence for 18 years and reported they had driven more than 16.000 km per year. Participants drove almost every day in an urban context, less often on suburban roads, highways, and country-side or mountain roads (in decreasing frequency order).

All participants very often used mobile phones and personal computers and not very often smartphones, MP3 players, portable or on-board navigators. The percentage of users that never used an integrated navigation system is high (60%), as well as the percentage of participants that used rarely (40%) or never (30%) a nomadic one.

Moreover the majority of participants did not have experience in driving with eco-drive systems (55%), but mostly of the users that experienced those (45%) used them daily or several times a month.

Stimuli
Visual stimuli were of different types:

- Eco-drive task related: gear shift indicator, feedforward advices (curve, traffic light, roundabout, pedal release…), feedback (e.g. accelerator pedal release correct or not correct), green speed advice, eco-score behaviour indicator (e.g. eco-tree).
- Primary driving task related: speedometer
- Driving (speedometer) and eco-drive system stimuli were visualized on the cluster (Figure 3), located behind the steering wheel.
- Tertiary tasks related: during both baseline and experimental driving sessions, they were randomly administered by a recorded voice during the appearance of the feedforward advices and consisted in setting the volume or reading an SMS, which were displayed on the central head unit.

Figure 3: Information shown on the cluster.

The haptic stimulus was given by a counterforce pedal which produced a reverse force opposite to the force the driver applied to accelerate (Birrell, Youngand and Weldon, 2010). This feedback on the pedal communicated to the user it was preferable to remove the foot from the accelerator to fullfil the eco-behaviour. The pedal opposed to the driver a force between 10 and 30N.

Experimental design
The experimental design was a 2x2 within-subjects. The manipulated factors were eco-drive HMI (*Visual* and *Visual&Haptic* feedback) and driving scenario (*Urban* and *Extra-urban*). In order to avoid sequence and learning effects, these four driving sections were randomized and counterbalanced.

Procedure and questionnaires
At the very beginning participants were received in Virtual Reality Driving Simulator by an experimenter. They were thanked for their participation and introduced about the test, the driving simulator and the eco-drive system under evaluation. After that they filled in a socio-demographic and driving behaviour questionnaire.

A training driving session without any stimuli was performed by all the participants to guarantee each user the same level of basic expertise in driving the Virtual Reality Simulator.

Then participants performed a *Baseline* driving session, in which they had to drive both in urban and extra-urban scenarios without any eco-drive HMIs and at the end they filled in a questionnaire, in which, beyond evaluating their drives, the Rating Scale Mental Effort (Zijlstra, 1993) in the different encountered events (roundabout, traffic light…) was also administered.

After the *Baseline*, the driving test conditions with the two HMIs (*Visual* or *Visual&Haptic* feedback) started. Each participant drove again in *Urban* and *Extra-urban* tracks with fixed obstacles and difficulties. During these sessions, different visual feedforward and *Visual* feedback or *Visual&Haptic* feedback were used during pre-defined events in the scenario (curves, traffic jam, etc.).

At the end of each experimental session, participants filled in some questionnaires about their drive, their perceived mental effort (as after the *Baseline*) and their evaluation on different HMIs used in each session.

At the end of the driving tests, participants filled in a questionnaire to evaluate the system interfaces in detail and to indicate the preferred *Visual* or *Visual&Haptic* feedback.

Apparatus and Driving Scenario
The CRF simulator is based on a six degrees of freedom dynamic platform. Besides providing the classical motion-based driving simulators functionalities such as ride vibrations, chassis translation and rotation, motion cueing, it enables the visualization of the car interior, due to Virtual Reality techniques, such as the 3D

full immersive visualization with head motion tracking and visual compensation algorithms. The system includes an I-Space and a physical mock up mounted on the mobile platform. Other features of the CRF driving simulator are the highly realistic vehicle dynamic models and a flexible and configurable vehicular traffic model, making possible the generation of critical situations.

Two different driving scenarios were simulated: urban and extra-urban. Each driver drove on a simulated one-way road with two lanes. Partcipants were instructed to drive at 50km/h in the urban scenario and at 70km/h and 90km/h in the extra urban one and stay in the right lane as much as possible. The surrounding traffic was made up of other cars.

Figure 4. Picture of Urban (left) and Extra-urban (right) simulated scenario.

Results

Participants reacted in the correct way to the *Visual&Haptic* feedback more times than to the *Visual* one (p<.05). The 'correct' behaviour more frequently occurs with any type of feedback than in the baseline without the eco-driving system (p<.05).

Participants reacted faster when prompted by the interface. Moreover, they reacted faster with the *Visual&Haptic* interface than to the *Visual* one (p<.05).

Consumption data significantly decreased (p<.05) between the different conditions.

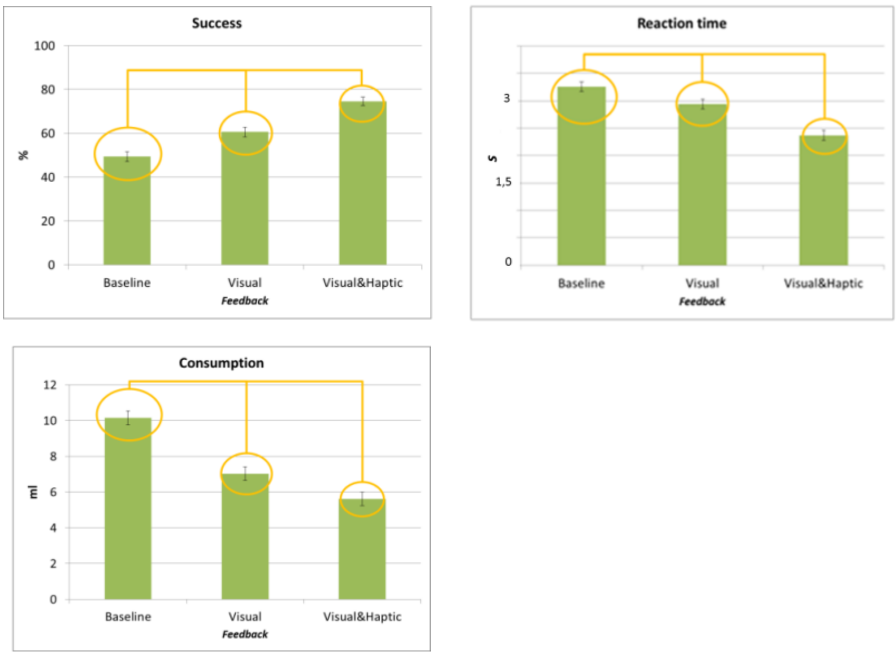


Figure 5. Objective performance with different interfaces [ANOVA statistical differences are reported; p<.05].

From these results we can assume that *Visual&Haptic* interface gives the driver the chance to react more promptly and to increase fuel consumption reduction (Figure 4). The use of both feedback and in particular the *Visual&Haptic* one seems to improve previous driver performances both in the *Urban* and *Extra-urban* environment.

Perceived driving performance with different interfaces

Comparing participants' evaluation of their driving performance in the three conditions, there is a significant improvement of *driving safety* and *driving accuracy* perception with the *Visual&Haptic* system in comparison with the *Baseline* condition [t=-2,17; $p<.05$][t=-2,22; $p<.05$]. No other differences are statistically significant between conditions. *Visual&Haptic* interface is also evaluated as significantly different from neutral (3) [t=3,5; $p<.05$] [t=2; $p<.05$] for both safety and accuracy dimensions.

Participants' perception of their driving seems to increase toward a positive perception with the *Visual&Haptic* system, while it remains neutral with the *Visual* feedback, which does not differ from the *Baseline* condition (Figure 6).

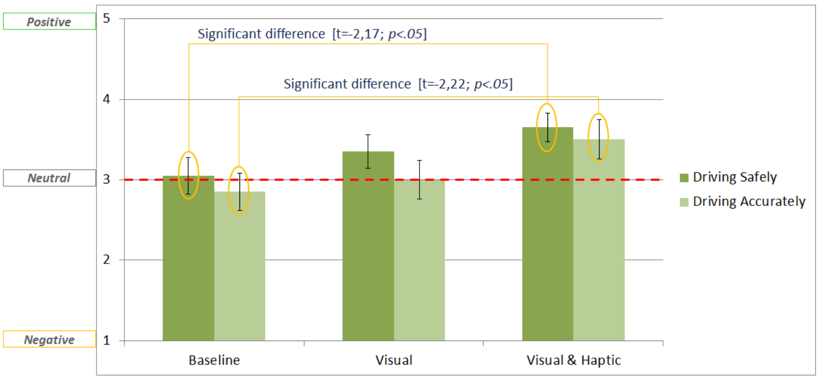


Figure 6. Participants' driving performance self-evaluation.

Participants self-perception while using the system (Visual/Visual&Haptic)

Mean levels of *anxiety*, *annoyance* and *curiosity* for the two different HMI configurations are very similar (Figure 7). No statistical significance has been found between the two conditions. Anyway, *anxiety* self-perception was statistically different from neutral point (3) for both interfaces [t=-2,1; $p<.05$] [t=-2,8; $p<.05$]. Other statistically significant data are those about *curiosity* towards the system, in both conditions [t=8,7; $p<.05$][t=10,2; $p<.05$] and *satisfaction*, with the *Visual* interface [$p<.05$]. In general, participants declare a positive perception while using both systems and they were, in particular, curious about them.

Figure 7. Users self-perception while driving.

Participants perception of systems impact on driving (Visual/Visual & Haptic)

Participants give evaluation not statistically different from neutral (3) on the idea that the system could make their driving more difficult. The evaluation is the same for both interfaces. They evaluated the systems positively in terms of making them drive more carefully: the evaluations of both systems are significantly different from neutral point (3) [t=3,45; $p<.05$] [t=3,55; $p<.05$]. The most interesting result is the significant difference between participants evaluation of the system ability to make them react more promptly to emergency situations: while the evaluation of the *Visual* system is not different from neutral, the evaluation of the *Visual&Haptic* interface is significantly more positive [t=-2,2; $p<.05$]. They feel their driving is more prompt with the combined *Visual&Haptic* feedback (Figure 8).

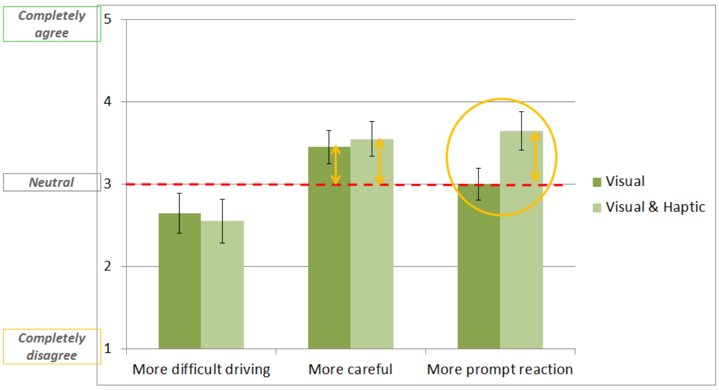


Figure 8. Participants perception of systems impact on driving.

Visual information evaluation

Participants' evaluation of the *Visual* interface in all conditions is positive (Figure *Figure*9). The main differences in evaluations are on *comprehensibility* of the system: there is a significant decrement in the evaluation between *Visual&Haptic* condition and the overall evaluation of *Visual* information [t=2,6; $p<.05$].

A similar result is relative to *raising alertness* [t=2,04; p=.055]: there is a tendency through a decrease in the evaluation between *Visual&Haptic* and final evaluation of the *Visual* system. No other differences are significant from a statistic point of view. The evaluation of the *Visual* interface, in general, seems not to be influenced by the presence of the haptic feedback.

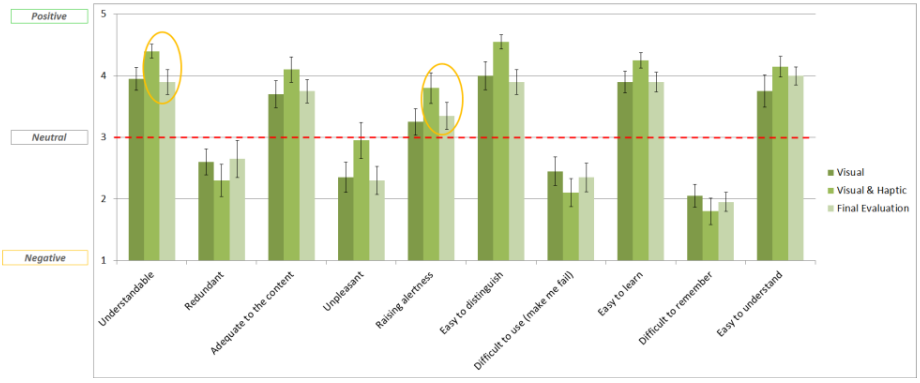


Figure 9. Visual information evaluation in experimental conditions and at the end of the test.

Visual information and haptic information evaluation in visual & haptic condition
During *Visual&Haptic* condition, participants evaluated visual interface as more *redundant* than haptic one, even if the score is near neutral point (3) [t=3,1; $p<.05$]. They evaluated the *Haptic* interface as significantly more *adequate to the content* than the *Visual* one [t=-2,4; $p<.05$]. They also indicated that the *haptic* feedback is significantly more *easy to distinguish* than *Visual* one [t=-2,7; $p<.05$]. In general, both interfaces are positively evaluated by participants, although the *haptic* one was evaluated more positively (Figure10).

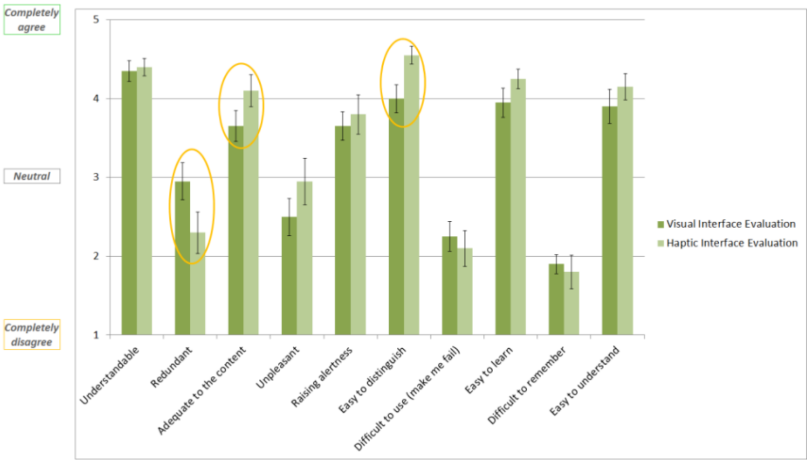


Figure 10. Visual and haptic information evaluation in Visual&Haptic condition.

Mental effort

Taking into account all different driving situations (curve, hills, etc.) participants, using the RSME scale, evaluated their mental effort significantly lower when they

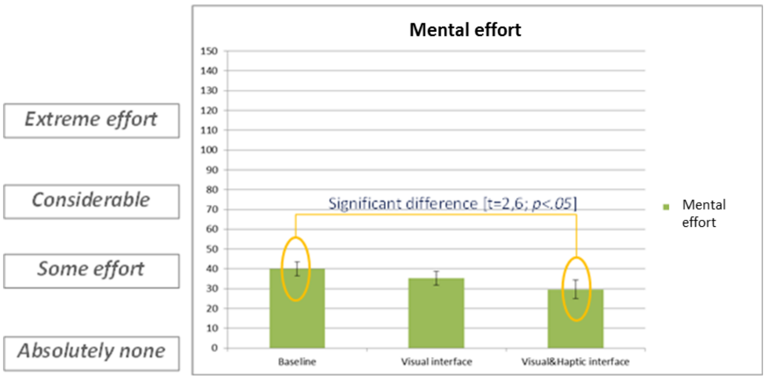


Figure 11. Subjective evaluation of mental effort after different driving.

drove using the *Visual&Haptic* interface, than without the system (*Baseline*) [t=2,6; p<.05] (Figure 11). No differences in the evaluation have been found between the *Baseline* and driving with the *Visual* interface, nor between the *Visual* interface and the *Visual&Haptic* one.

Willingness to have an eco-drive system, comparison between Pre and Post questionnaires

Before driving with the system, participants were not sure they would have liked to drive with an eco-driving system while, after tried it, they definitely wanted to (Figure 12).

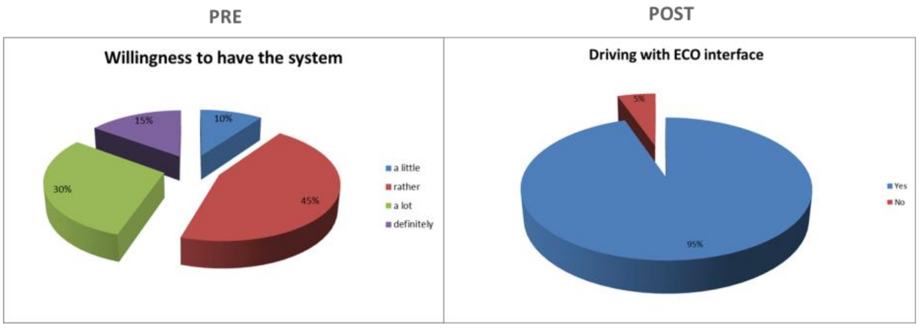


Figure 12. Pre and post willingness to have an eco-drive system.

Feedback evaluation

At the end of the test, participants were asked to choose between *Visual* only or *Visual&Haptic* interface. No differences emerged: 55% of participants preferred to drive with the *Visual* feedback only and 45% with the *Visual&Haptic* one.

Conclusion

The performance data results show that in particular *Visual&Haptic* feedback give the driver the possibility to react more promptly and to reduce fuel consumption respect to the baseline condition. This is coherent also with the subjective data on impact on driving and perceived mental effort. The use of both eco-interfaces and in particular the *Visual&Haptic* one seems to improve driver performances both in the *Urban* and *Extra-urban* scenario. Participants' perception of their drive improves between the *Baseline* condition and the driving with the *Visual&Haptic* interface, both on safety and on accuracy perception.

The eco-driving interface an in particular that with *Visual&Haptic* feedback reduces participants' perceived mental effort. Haptic feedback on the pedal is considered a good solution, even if too intrusive during the drive, as emerged from participant's comments. Then it is necessary to develop it further in order to reduce the perceived intrusiveness.

The visual feedback, important to understand the haptic one, was well evaluated by participants. All information are understandable and simple to follow, the visual stimuli are understandable and easy to be used during the driving. There is no a clear preference between only *Visual* and the *Visual&Haptic* type of feedback.

From a general overview of the obtained results arises a general positive evaluation of the tested eco-driving system HMI and participants declare their willingness to use it while driving. These results consitituted the starting point for designing the CRF ecoDriver HMI which is tested with a long term controlled on-real roads study in 2014.

References

Birrell, S.A., Youngand, M.S., Weldon, A.M., (2010). Delivering smart driving feedback through a haptic pedal, in Contemporary Ergonomics and Human Factors 2010, Chap.72, 431-439.

Campbell, J. L., Richman, J. B., Carney, C., and Lee, J. D. (2004). In-vehicle display icons and other information elements. Volume 1: Guidelines (Report No. FHWA-RD-03-065). McLean, VA: Federal Highway Administration.

FESTA Handbook (2011). Available from http://www.fot-net.eu/download/festa_handbook_rev4.pdf

Fricke, N., & Schießl, C. (2011). Encouraging environmentally friendly driving through driver assistance: the eCoMove project in PROCEEDINGS of the Sixth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.

International Organization for Standardization (2010). Ergonomics of human-system interaction Part 210: Human-centered design for interactive systems. Genève: ISO. (ISO 9241-210:2010).

International Organisation for Standardization (2005). Road vehicles - Ergonomic aspects of in-vehicle presentation for transport information and control systems - Warning systems. Genève: ISO. (ISO/TR 16352:2005).

Jenness, J.W., Singer, J., Walrath, J., & Lubar, E., (2009). Fuel economy driver interfaces: Design range and driver opinions. Task 1 and Task 2 Report. Washington, DC: National Highway Traffic Safety Administration (DOT HS 811 092).

Manser, M. P., Rakauskas, M., Graving, J., & Jenness, J. W., (2010). Fuel Economy Driver Interfaces: Develop Interface Recommendation. Report on Task 3. National Highway Traffic Safety Technical Report (DOT HS 811 319).

Zijlstra, F.R.H. (1993). *Efficiency in work behavior*. A design approach for modern tools. PhD thesis, Delft University of Technology. Delft, The Netherlands: Delft University Press.

Active Multisensory Perception tool: BUS experience and action comfort

Walter Gerbino[1], Carlo Fantoni[1,2], Carlo Nicolini[2], Robert Volcic[2], Fulvio Domini[2,3]
[1]University of Trieste, Italy
[2]Istituto Italiano di Tecnologia, Italy
[3]Brown University, Providence, RI, USA

Abstract

Beauty, Utility, and Simplicity (BUS) are interconnected goals of vehicle design, dependent on several factors including (1) the temporary mood induced in the user and (2) the quality of multisensory integration involved in the active perception of vehicle interiors. Within the *affective ergonomics* framework, the relationships between vehicle comfort, user's mood, and perceived BUS constitute a stimulating research field. Specifically, action comfort may affect the BUS experience. Here, we describe an advanced experimental environment, called the Active Multisensory Perception tool (AMPt), capable of supporting a variety of studies in this field. AMPt has been developed within the EcoAutobus INDUSTRIA 2015 project, by integrating the latest technologies in body motion tracking (Optotrak), positioning (Velmex actuators), and High-Res 3D visualization for virtual and augmented reality. We also report preliminary results on the usage of AMPt to measure how action comfort affects the perception of facial expressions of emotions. We induced opposite biases in the perception of facial expressions (leading to positive/negative evaluations of neutral faces) by adapting participants to comfortable/uncomfortable visually guided reaching for 3D objects at variable distances. Results are consistent with the idea that action comfort induces a positive mood in the user, which in turn enhances the quality of BUS global experience.

Introduction

A large body of research on object perception and representation refers to the processing of information within a given sensory modality and to its interaction with primitives, schemata, and other types of mental entities. Comparatively less research refers to the integration of information combined in the course of the action-perception cycle, despite the role classically attributed to action-specific information in the acquisition of knowledge about the world (Santos & Hood, 2009).

Extraretinal signals from egomotion are indeed used for the interpretation of different cues to depth (Wexler & van Boxtel, 2005). Tactile exploration contributes to the disambiguation of sensory information by reducing the alternation between

In D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, & A. Toffetti (Eds.) (2014). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

perceptual solutions equally compatible with the visual input (Conrad et al., 2012). Again, the efferent copy of motor commands involved in head movements (i.e., head translation velocity) is essential for the perception of structure from motion displays, by reducing the likelihood of tilt reversals (Wexler et al., 2001) and enhancing the sensitivity to planar surface orientation and motion (Caudek et al., 2011; Fantoni et al., 2010, 2012). The evidence, exemplified by these studies, supports the need of framing vehicle design within a theory of human perception that re-asserts the primacy of embodiment, development, and interaction in cognitive systems.

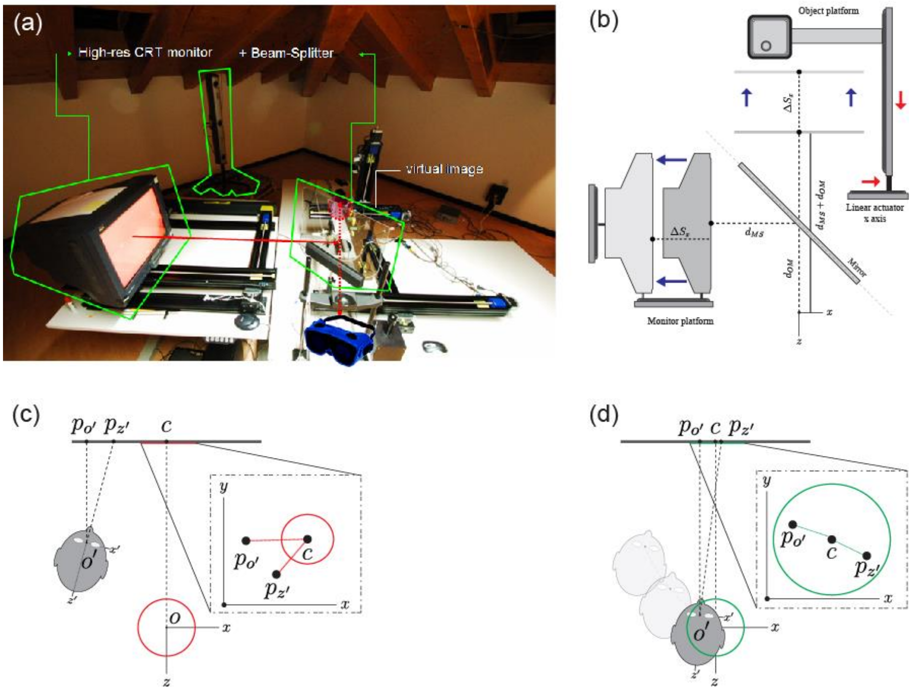


Figure 1. AMPt and observer's alignment procedure. (a) Frontal view of AMPt setting. Green bold lines surround the main devices composing AMPt: from left to right, a high resolution CRT monitor, a motion tracking system, a beam-splitter slanted 45° away from the screen and the observer's viewing direction. Stereo presentation is based on LCD shutter glasses. Two platforms with linear actuators are also integrated in AMPt, dedicated to the on-line positioning of screen and real objects, respectively. (b) A diagram of AMPt and viewing geometry. The 45° beam-splitter reflects the images from the CRT screen which is parallel to the z-axis of the reference frame. Dashed lines show the viewing direction for a standard observer facing the virtual image of the screen appearing at a distance that equals the mirror to screen distance (d_{MS}) plus the observer to screen distance (d_{OS}). The virtual image of the screen moves farther by Δ_S as the screen is moved along the linear actuator of the monitor platform parallel to the x-axis of the reference system. The object platform is placed behind the mirror and includes three linear actuators all aligned with the reference frame. (c-d) Alignment procedure. The world-centric reference frame is O(x,z); the head-centric reference frame is O'(x',z'). The black line is the virtual image of the screen (corresponding to the

projection plane) whose centre is in c. Point po' is the parallel projection of O' on the projection plane while pz' is the intersection of the optical axis (z') and the projection plane. In (d), the observer is centred and aligned with the reference frame: O' is indeed within the tolerance area for head position (green circle), while both po' and pz' are close enough to the screen centre since they are within the tolerance area for head orientation (green circle).

The goals of vehicle design are captured by the combination of Beauty, Utility, and Simplicity (BUS), a set of interconnected concepts consistent with the framework of *affective ergonomics*, which takes the effectiveness and pleasantness of actions performed on objects as determinants of our perceptual experience (Ho & Spence, 2013; Norman, 2004).

Thus, vehicle design and ergonomics can benefit from the development of innovative ways of studying the integration of information obtained during the active exploration of a multisensory environment. To overcome the limitations of traditional approaches based on the passive observer, an experimental setting should include the registration of relevant inputs from the active observer to provide him/her with the appropriate sensory stimulation. In order to fill this gap with an advanced solution, we developed the Active Multisensory Perception tool (AMPt), currently in place in three interconnected labs at the Istituto Italiano di Tecnologia, the University of Trieste, and Brown University.

In the present study we will: (1) describe AMPt (next section), and (2) report an application of AMPt relevant to the field of *affective ergonomics*, which draws on an embodied theory of cognition (Section "Scenario: Action comfort, perception of facial emotions, and BUS"). Our study shows that the perception of emotional faces in normal individuals can be biased by an innovative mood-induction procedure, based on the repetitive execution of reaches with different degrees of comfort.

The Active Multisensory Perception tool

AMPt integrates the latest technologies in the field of motion tracking (Optotrak 3020 Certus, Northern Digital Inc., Waterloo, Ontario, Canada), positioning system by multi-axes computer controlled actuators (Velmex Inc., Bloomfield, NY, USA), and 3D visualization in virtual and augmented reality with Open GL (Figure 1a). AMPt is used to optimally associate the visualization of 3D structures defined by different cues to depth (texture, stereo, motion, shading) to their tactile and proprioceptive counterparts, and to control in real time and to store: (1) motor responses for reaching and grasping (i.e., maximum grip aperture, velocity), used as indirect behavioural measures of underlying perceptual processes; and (2) behavioural responses based on advanced psychophysical techniques using forced choice, speeded classification and probing tasks.

AMPt superposes virtual images upon real objects, defining an augmented field of view and a systematically controllable interaction environment. AMPt also includes a specific algorithm allowing to replicate the very same optic information produced by the active exploration of a 3D visual object, and to display it to the same passive (stationary) observer. This innovative algorithm has been shown to be relevant to disentangle the contribution of retinal and extraretinal signals in the perception of 3D structures (Caudek et al., 2011; Fantoni et al., 2010, 2012).

Knowing where body segments are

Identifying the position and orientation of an active observer and his/her body parts in an arbitrary coordinate frame is mandatory in order to develop a virtual reality system with active head-tracking. We have implemented a method to track arbitrary points on a rigid-body that is based on the initial relation between these points and a set of reference markers. This solution is indispensable given that placing a physical marker on a particular body segment is frequently impossible due to both ergonomic and technical factors. This difficulty, known as the absolute orientation problem, was overcome by means of the Umeyama algorithm (Umeyama, 1991), which outputs the affine transformation (i.e., rotation and translation) that describes the isometry between two sets of points.

This approach has the advantage of describing the movements of each tracked body segment in 6 degrees of freedom (3 translations, 3 rotations), so that the user is completely free to move. Furthermore, our approach is numerically more stable and efficient than alternative approaches based on Euler angles, and it also avoids the problem of *gimbal lock*; i.e., the loss of rotational degrees of freedom, due to singularities. For the numerical and linear algebra computations an underlying C++ library called CNCSVision, based on the matrix library Eigen (eigen.tuxfamily.org), has been implemented that explicitly targets the needs of a head/hand tracking system.

In AMPt, the centre of projection (COP) is the nodal point of the observer's eye in a world-centred coordinate frame. In order to extract the COP, the user interocular distance is measured and then the nodal points are computed as the points that lay equidistant from the centre on the line connecting two markers on the opposite sides of a pair of goggles worn by the participant. Observer's head orientation is thus aligned to the world-centred reference frame by means of a first calibration phase (Figure 1c,d). In this phase the participants are asked to align the orientation of the sagittal plane of their head (extracted relatively to a triplet of markers attached to the back of their head) so to be perpendicular to the projection plane. Such a process is accomplished by means of an interactive procedure during which the participants must align the orthogonal projection of their cyclopean eye and a point representing their gaze direction to a central reference point displayed on the virtual projection plane (Figure 1c,d). This alignment procedure generates instantaneous measures of head rotation (pitch, roll, and yaw) and translation (in X, Y, Z), which are also available to the experimenter in real time on a secondary computer platform via a TCP/IP data transmission protocol.

Using a similar procedure, AMPt allows us to track parts of the body other than the head, such as the fingers and the wrist. The finger positions are easily computed by instructing the observer to lean one finger on a marker in a predefined position. Then, the finger tip position is extracted relatively to a triplet of markers attached to the distal phalanx. This method can be used to track single or multiple fingers to study both reach-to-point and reach-to-grasp movements (Volcic et al., in press).

The very same approach is applied in augmented reality conditions, where virtual stimuli are superimposed on real objects (Foster et al., 2011).

Combining High-Res 3D visualization with body motion

The aforementioned CNCSVision library is also used to generate static and dynamic visual stimuli and 3D shapes. To present visual stimuli in 3D, we use a frame interlacing technique in conjunction with liquid crystal FE-1 goggles (Cambridge Research Systems, Cambridge, UK) synchronized with the monitor's frame rate. OpenGL display lists as well as vertex buffer objects are employed in order to minimize the computational load of the CPU. All these technical solutions allow us to work at 100 Hz refresh rate.

The advantages of having a 100 Hz refresh rate are: (1) negligible time lag between user movements and visual feedback, (2) optimal integration with shutter glasses frame rate and almost total absence of flickering effects in artificially lit environments, (3) sufficiently large data collections for detailed and precise kinematics reconstruction and analysis.

The projection stage of our head tracking system has been modified with respect to the usual perspective projection technique implemented in OpenGL (*gluPerspective*) to avoid possible distortions due to oblique view of the active observers relative to the projection plane. Our implementation is based on the *generalized perspective projection* method (Kooima, 2009). We modified the projection matrix **P** used in OpenGL with a more general $\mathbf{P'} = \mathbf{P} \cdot \mathbf{M}^{\mathbf{T}} \cdot \mathbf{T}$, where **M** and **T** are respectively a rotation matrix and a translation matrix that describe the position of the projection plane in the world. This projection model generalizes to binocular viewing. Two COPs are made to correspond with the nodal points of the observer's eyes, leading to a correct stereoscopic vision during free head movements.

A "more real" virtual reality

Humans exploit a range of visual depth cues to estimate three-dimensional structure. Information from multiple cues is combined to provide the viewer with a unified percept of depth conveying an "illusion of reality" that strongly depends on cue congruency. Under natural viewing, ocular (vergence and accommodation of the eyes) and optic (texture, disparity, motion) information about depth are all consistent (Hoffman et al., 2008).

However, traditional devices for 3D visualization (virtual reality head-mounted displays as well as various types of stereoscopes) produce artefacts that are absent in natural viewing, due to a fundamental technological shortcoming: the distance of the point at which visual axes converge generally does not correspond to the distance of the projection screen. This situation creates a conflict between the viewing distance of the rendered 3D scene and the viewing distance specified by vergence and accommodation. Several studies have shown how this type of cue conflict reduces the realism of the 3D visualization and causes systematic distortions of 3D vision, in favour of a flattened 3D representation.

Our setup overcomes this shortcoming by allowing us to systematically adjust the distance of the projection screen during the experiments with submillimetre precision (Figure 1b). Displays are viewed through a high-quality front-silvered mirror placed in front of the observer's central viewing position and slanted 45° away from the screen and the viewing direction. The effective distance from the pupil to the centre of the projection screen is controlled by a system of linear actuators that physically move the monitor and/or the mirror so to reproduce the required focus distance (ranging from 35 to 250 cm).

Simulated 3D objects are thus visualized in the space behind the mirror where they can be superposed to real objects to provide haptic feedback when moving in and interacting with the virtual environment. Furthermore, the observer can freely move his/her body and head while watching 3D rendered displays.

Scenario: Action comfort, perception of facial emotions, and BUS optimization

In the real world, vision operates in harmony with body motion yielding the observer with a vivid impression of three-dimensionality. In laboratory conditions, because of technological drawbacks, researchers have often studied the static observer neglecting important aspects of the action-perception cycle. Instead, AMPt enables the experimenter to recreate and bring into the lab conditions characterized by a high degree of sensorimotor coherence, typical of real world conditions. A promising application of AMPt is the study of 3D perception during a systematically controlled interaction with environmental objects. In this section we describe how AMPt flexibility can be used to study BUS optimization in the context of *affective design* and, in particular, to evaluate artefacts capable of improving hedonic components of the user experience when information about vision and goal-directed action is combined.

Does action comfort affect the perception of facial expressions of emotions?

Emotions are known to be contagious as they can be evoked while viewing emotionally expressive faces (Wild et al., 2001). The categorical perception and representation of facial expressions of emotions depend on past experience (Pollak & Kistler, 2002), neutral faces (Klatzky et al., 2011), music (Jolij & Meurs, 2011), and mood (Caudek & Monni, 2013). Therefore, the observer's mood might bias the perception of emotionally expressive faces and modify the experienced quality of physical and social environments. In particular, we hypothesize that the temporary mood induced by the discomfort/comfort associated with goal-directed actions can bias the perceived expression of an objectively neutral face.

The expectation is thus that the more comfortable the action is, the higher will the probability be that a neutral face appears to express a positive emotion (e.g., happiness). Conversely, the more uncomfortable the action is, the higher will the probability be that a neutral face appears to express a negative emotion (e.g., anger). This suggests that action comfort might affect the BUS experience. Within a vehicle, comfortable artefacts at an easy-to-reach distance should induce a positive mood, which in turn would enhance the global BUS experience, as revealed by a positive bias in the perception of facial expressions. Such a finding would have relevant implications for the *affective ergonomics* framework, establishing a clear relationship between vehicle comfort and user's mood.

An experiment

Several studies showed that the postural comfort during hand directed actions can be objectively measured (Carello et al., 1989; Heft, 1993; Kolsch et al., 2003; Mark et al., 1997; Warren, 1984). A person is in a state of postural comfort if there is not, and likely will not arise, a (possibly unaware) desire or need for compensating motion of other body parts. During reaching, the lower is the amount of compensatory body movements not regarding the arm (such as shoulder or trunk) the larger will be the comfort of the action.

In our experiment action comfort was systematically manipulated during visually guided reaching movements under unrestrained body conditions. Perception of emotional facial expressions was used to measure the degree of negative/positive mood induced by action discomfort/comfort during motor adaptation. In two successive blocks of trials, participants performed comfortable and uncomfortable reaches toward a virtual random dot cylinder (1.5-cm wide, 6.5-cm high) glowing in the dark, at variable distance along their line of sight. Within each adaptation trial the participant started a right hand movement from a fixed, out of view, position shifted relative to the body midline by about 250 mm from the sagittal plane and 150 mm from the coronal plane. The tip of his/her index finger, marked by a red dot, was constantly visible from the moment the finger entered in the participant's visual field. Each successful reach was accompanied by a haptic feedback (Figure 2a). Through AMPt motor functions we controlled the position of a physical cylinder placed behind the mirror (completely occluded from the participant) so to align it perfectly with the virtual stimulus. For consistent vergence and accommodative information, the position of the monitor was adjusted on a trial-by-trial basis to equal the distance from the participant's eyes to the virtual/real object.

Each adaptation block lasted 50 reaches and the depth extent of each reach was randomly selected in a range below (0.65-0.75 of arm length, Comfortable block) or above (0.90-1.00 of arm length, Uncomfortable block) the individual preferred critical boundary for 1-df visually guided reaching (Mark et al., 1997), illustrated in Figure 2c. In a pre-experiment we established that the entire range of depths used to manipulate the reaching comfort (0.65-1.00 of the arm length) produced a sizable effect on the subjective estimate of action discomfort, with ratings on a 1-50 pain scale (adapted from Ellermeier et al., 1991) increasing linearly with reaching distance for all tested participants (average r^2= 0.89).

After each adaptation block, the participant was required to perform a facial expression classification task lasting 48 trials. On each trial, the participant was shown the picture of an actor/actress portraying an expression of anger/happiness and required to classify it as "happy" or "angry" (Figure 2b). Displays were randomly selected from a set of 48 standardized facial expressions resulting from the combination of 8 actors/actresses (4 males and 4 females, all Caucasian, selected from the Radboud University Nijmegen stimulus set) × 6 morph intensities along the continuum from 0.25 (rather happy) to 0.75 (rather angry). The ordering of the type of reaching was balanced across participants: half adapted first to comfortable and then to uncomfortable reaches, and the remaining half vice-versa.

Classification performance was calculated by fitting a psychometric curve to individual proportions of "angry" responses as a function of morph intensity, assuming a Gaussian model whose parameters were estimated using the constrained maximum likelihood and bootstrap inference method implemented by the *psignifit* software (Wichmann & Hill, 2001). Every psychometric curve generated a point of subjective neutrality (PSN) at the 50% threshold, representing the morph intensity perceived as a neutral expression. Panels d,e in Figure 2 illustrate the average proportions of "angry" responses as a function of morph intensity for comfortable (black) vs. uncomfortable (red) actions, for the two adaptation orderings: comfortable-uncomfortable (panel d) vs. uncomfortable-comfortable (panel e).

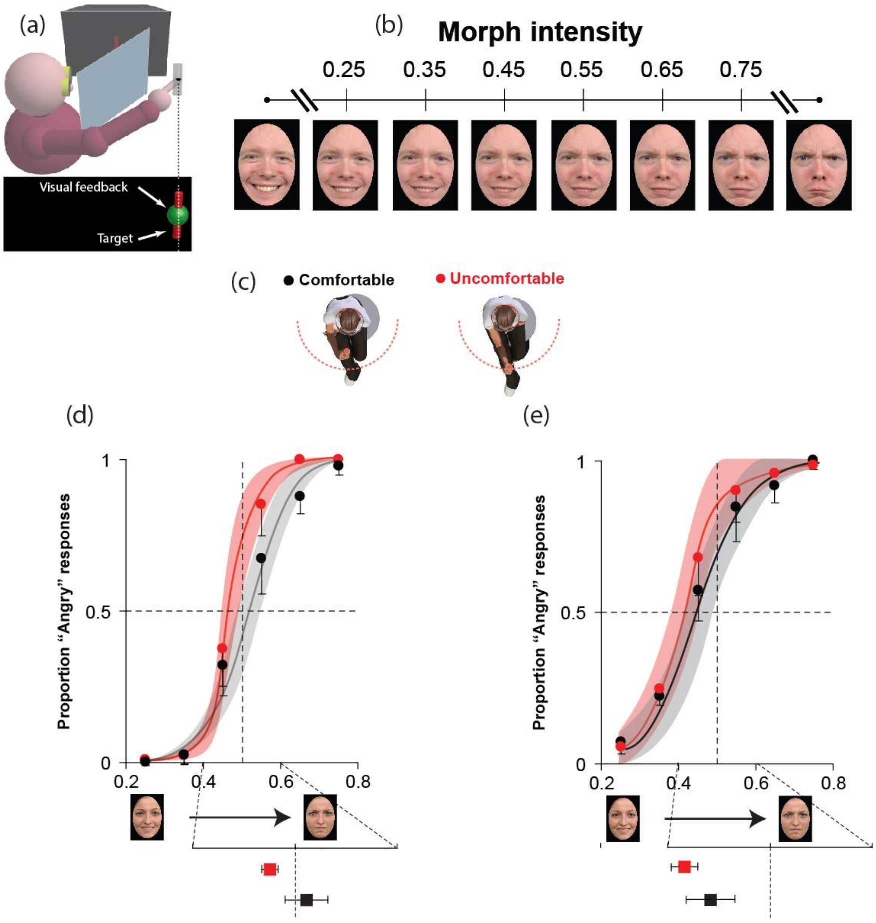


Figure 2. (a) The setting of the adaptation procedure. (b) A series of facial expressions morphed between happiness (left) and anger (right). (c) Two action types and the individual critical boundary for 1-df visually guided reaching, represented by the red dashed semicircle

with radius equal to 0.9 of arm length. (d-e). The two panels show the average proportions of "angry" responses [± 95% CI] as a function of morph intensity, for uncomfortable (red symbols) and comfortable (black symbols) reaching actions. Red and black curves are the average best fitting psychometric functions, with shaded bands indicating [± 95% CI]. Average PSNs [± 95% CI] are plotted on a zoomed scale (bottom). Data in panel (d) refer to the comfortable-uncomfortable order; data in panel (e) to the opposite order.

Average PSNs shown at the bottom of panels d,e of Figure 2 are in strong agreement with the expectation that classification performance is biased in opposite directions after comfortable (towards happiness) vs. uncomfortable (towards anger) adaptation phases. The likelihood of interpreting a facial expression as angry increased of about 7% after being adapted to an uncomfortable reaching rather than to a comfortable reaching (t_{40}= 3.4, p< 0.005). A similar difference in the perception of facial emotions was found in the two conditions of temporal ordering, no matter whether the comfortable reaching phase was administered at the beginning (7.5%; Figure 2d) or end (6.5%; Figure 2e) of the experiment. Furthermore, consistently with the induction of a negative mood by action discomfort, we found an overall increase in the tendency to interpret the facial expressions as "angry" (t_{40}= 3.0, p< 0.005), when the comfortable reaches were performed at the end ($\overline{PSN} = 0.43$; Figure 2e) rather than at the beginning ($\overline{PSN} = 0.49$; Figure 2d) of the session.

In summary, our results are compatible with the idea that the perception of emotion from facial stimuli depends on action comfort, thus providing new insight into the relationships between comfort, user's mood, and perceived BUS, which are central within the *affective ergonomics* framework and consistent with recent findings on the effect of posture on feelings and behaviour (Yap et al., 2013). A vehicle designed to support comfortable actions will likely induce a positive perception of environment; while the design of interior spaces that do not consider the limits of user's action will force users to perform uncomfortable actions, inducing a negative perception of the environment. Personalization of spaces might constitute a key tool for vehicle design oriented to the optimization of user's BUS experience.

References

Carello, C., Grosofsky, A., Reichel, F.D., Solomon, H.Y., & Turvey, M.T. (1989). Visually perceiving what is reachable. *Ecological Psychology, 1,* 27-54.

Caudek, C., & Monni, A. (2013). Do you remember your sad face? The roles of negative cognitive style and sad mood. *Memory*, [Epub ahead of print], doi: 10.1080/09658211.2013.765893.

Caudek, C., Fantoni, C., & Domini, F. (2011). Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow, *PLoS ONE, 6*: e18731, doi:10.1371/journal.pone.0018731.

Conrad, V., Vitello, M.P., Noppeney, U. Interactions between apparent motion rivalry in vision and touch. *Psychological Science*, *23*, 940–948.

Fantoni, C., & Domini, F. (2010). Systematic distortions of perceived planar surface motion in active vision. *Journal of Vision*, *10*, 1–20, doi:10.1167/10.5.12.

Fantoni, C., Caudek C., & Domini, F. (2012). Perceived surface slant is systematically biased in the actively-generated optic flow. *PLoS ONE*, *7*: e33911, doi:10.1371/journal.pone.0033911

Foster, R., Fantoni, C., Caudek, C., & Domini, F. (2011). Integration of disparity and velocity information for haptic and perceptual judgments of object depth. *Acta Psychologica, 136,* 300-310, doi:10.1016/j.actpsy.2010.12.003

Ho, C., & Spence, C. (2013). Affective multisensory driver interface design. *International Journal of Vehicle Noise and Vibration, 9*, 61-74.

Hoffman, D.M., Girshick, A.R., Akeley, K., & Banks, M.S. (2008). Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision*, *8*, 1-30.

Jolij, J., Meurs, M. (2011). Music alters visual perception. *PLoS ONE*: e18861, doi:10.1371/journal.pone.0018861.

Klatzky, R.L., Abramowicz, A., Hamilton, C., & Lederman, S.J. (2011). Irrelevant visual faces influence haptic identification of facial expressions of emotion. *Attention, Perception, & Psychophysics, 73*, 521–530.

Kölsch, M., Beall, A.C., & Turk, M. (2003). The postural comfort zone for reaching gestures. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *47*, 787-791.

Kooima, R. (2009). Generalized Perspective Projection, *http://csc.lsu.edu/~kooima/articles/genperspective/index.html*

Mark, L.S., Nemeth, K., Gardner, D., Dainoff, M.J., Paasche, J., Duffy, M., & Grandt, K. (1997). Postural dynamics and the preferred critical boundary for visually guided reaching. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1365-1379.

Norman, D. (2004). *Emotional Design: Why We Love (or Hate) Everyday Things*. New York: Basic Books.

Pollak, S.D., & Kistler, D.J. (2002). Early experience is associated with the development of categorical representations for facial expressions of emotion. *Proceedings of the National Academy of Sciences*, *99*, 9072–9076.

Santos, L.R. & Hood, B.M. (2009). Object representation as a central issue in cognitive science. In B.M. Hood & L.R. Santos (Eds.*) The Origins of Object Knowledge*. Oxford: Oxford University Press.

Umeyama, S. (1991). Least-Squares Estimation of Transformation Parameters Between Two Patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 376 – 380.

Volcic, R., Fantoni C., Caudek, C., Assad, J., & Domini, F. (in press) Visuomotor adaptation changes the processing of binocular disparities and tactile discrimination. *Journal of Neuroscience.*

Warren, W. H. (1984). Perceiving affordances: Visual guidance of stair climbing. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 683-703.

Wexler, M., & van Boxtel, J.J. (2005). Depth perception by the active observer. *Trends in Cognitive Science*, *9*, 431-438.

Wexler, M., Panerai, F., Lamouret, I., & Droulez, J. (2001). Self-motion and the perception of stationary objects. *Nature*, *409*, 85–88.

Wichmann, F.A., & Hill, N.J. (2001). The psychometric function: I. Fitting, sampling and goodness-of-fit. *Perception and Psychophysics*, *63*, 1293-1313.

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: quality, quantity, time course and gender differences. *Psychiatry Research, 102*, 109-124.

Yap, A.J., Wazlawek, A.S., Lucas, B.J., Cuddy, A.J.C., & Carney, D.R. (2013). The Ergonomics of dishonesty: The effect of incidental posture on stealing, cheating, and traffic violations. *Psychological Science* (published online, pp. 1-9), doi:10.1177/0956797613492425.