# MEASURING THE IMPACT OF USABILITY ON DRIVING SAFETY

Thomas Vöhringer-Kuhnt, Manfred Thüring, Matthias Rötting, Thomas Herbig
Berlin University of Technology, Department of Psychology and Ergonomics
Franklinstr. 28/29
10587 Berlin
tvo@mms.tu-berlin.de

**ABSTRACT**: Ease-of-use and safety aspects are essential attributes for assessing the quality of in-vehicle information systems (IVIS). Although most accidents due to distraction are related to events outside the vehicle, a significant effect of secondary tasks on driver distraction has been proven. These influences are particularly relevant for liability issues which are of increasing importance in the automobile industry. As a consequence, there have been numerous attempts to provide evaluation routines for IVIS, but "industry-best-practice" methodologies, that are widely accepted, are still missing. In the field of software ergonomics, such tools are already at hand and at least first efforts in the automotive sector have been made in this direction. Based on these approaches, we developed a short usability questionnaire tailored to evaluate the human machine interface (HMI) of IVIS. This questionnaire addresses psychological aspects, such as intuitive use and hedonic qualities, as well as design principles for screen layout and dialogue figuration.

## 1 INTRODUCTION

Advanced in-vehicle information systems have become essential products of the automobile industry. Consequently, specific requirements for the design of the automotive human- machine interface (HMI) have to be met in order to combine all single systems and functions into a holistic concept of interaction. Beyond the pure functionality of in-vehicle information systems, usability and safety requirements have to be addressed. Guidelines for the design of cockpits and automotive systems have already been developed by the industry as well as by legal authorities. On the other hand, no standardized evaluation procedure is established, which is generally accepted by all parties involved.

### *1.1 The CarUSE Project*

The goal of the project ‚CarUSE' is to engineer a tool for describing and evaluating the automotive human-machine interface. It is intended to be used as a rapid screening test to give hints about strengths and weaknesses of a certain HMI regarding its influence on driving safety. Overall, four key aspects are addressed: First of all, ease-of-use, intuitive use and cognitive ergonomic criteria of the interface are relevant quality aspects with some impact on driving safety. Second, the time to complete a certain task influences driver distraction. This issue is based on benchmark criteria such as the 15-second rule [1,2] and similar suggestions regarding task completion time, which have found their way into guidelines and standard-like recommendations [3,4]. Although static time on task is no perfect predictor of the amount of attention the driver addresses to a secondary task while driving, this rule of thumb is appealing because it allows evaluation results to be assessed and communicated in a straightforward way. Cognitive ergonomic criteria and task completion time are directly related to the amount of attention the driver must pay to secondary tasks. Third, operating a driver information system requires visual scanning which leads to a certain eyes-off-the-road time. Even though lane keeping can be accomplished with peripheral vision only [5], especially the reaction-time to critical incidents depends on focal vision to driving-relevant areas [6]. In our project, we use a model developed by Horrey, Wickens and Consalus [7] to predict visual attention allocation to in-vehicle information systems (IVIS) and the outside world, respectively. Finally, IVIS have to meet certain normative criteria, which are defined by a number of ISO and DIN standards. For IVIS and automotive HMI, critical aspects concern the color, size and shape of symbols, letters and words conveying the relevant information. The latter aspects are considered in the CarUSE Toolbox by a checklist to assess the conformity with such documents.

### *1.2 Results of a preliminary study*

In an explorative study, a number of evaluation techniques such as questionnaires/checklists, subjective ratings about the degree of distraction, operating time, and objective driving data obtained with the 'Lane Change Test' (LCT) [8, 9] were investigated and compared between a single task (driving without operating the systems and vice versa) and dual task scenario (operating the systems while diving). Rather than evaluating in-vehicle information systems themselves, the study focused on the assessment of established evaluation methods. Moreover, it was tested to what extent the 15-second rule (the assumption that task-duration predicts perceived distraction) is an appropriate basis for the evaluation tool. This was accomplished by calculating correlations between the time spent to accomplish a simple and a complex task with two different driver information systems and objective and subjective judgments of distraction and mental demand. We found that task completion time alone is not sufficient to predict driver distraction and mental workload. Regarding task complexity we observed an interaction effect: for the complex task, the rated mental effort increased tremendously while driving; for the simple task, we only found a small difference in mental effort between the single and dual task scenario [10]. Therefore, we transferred classical approaches used in software ergonomic and usability evaluation procedures to the automotive sector in order to asses the cognitive ergonomic quality of IVIS. This work is based on former efforts by Stevens [11], Schmitz [12] and Pataki [13].

## 2 QUANTIFYING COGNITIVE ERGONOMIC QUALITY OF IVIS

Usability engineering and human-machine interface design guidelines as well as recommendations for dialogue principles have reached a high level of maturity; their underlying principles are widely accepted. In the domain of desktop computer applications, plant and machinery engineering and human-computer interaction for office work with visual display terminals, all levels of requirements according to Hacker's system of ergonomic evaluation [14] are tackled: practicability (fulfillment of anthropometric requirements), lack of impairment and nuisance (avoidance of short-term stressors, long-term health problems and accidents), and personality enhancement (high degree of self-fulfillment). Even so, these aspects play a relevant role for the human-machine interfaces of IVIS. Regarding aspects of lower ergonomic-relevant levels such as perceptibility, evidence of conformity can be proven in a straight line: the size of symbols, text, and color combinations are determinable, therefore compliance with normative rules can be assessed directly. For higher levels of ergonomic criteria, which are associated with higher mental processes, this verification can not be done in such a straight forward way. First, there is no broadly accepted definition at hand, which mental restrictions are responsible for driver distraction during secondary task performance. Second, if critical parameters are indentified, their quantification can not be established directly. The latter aspect has been solved in software ergonomics by developing questionnaires which operationalize relevant dialogue principles. In the field of automotive HMI, the solution was to focus on quantifiable measures such as task duration (which lead to quasi-standards such as the 15-sec rule [1]) or the decline in performance during the driving task (which often is done in a standardized way by the lane change task [8]).

# 3 DEVELOPING A QUESTIONNAIRE FOR COGNITIVE ERGONOMIC ASPECTS

Our approach tries to adapt cognitive ergonomic criteria from the field of human-computer interaction to the automotive HMI by applying techniques of classical psychometric test-development. The goal was to develop an instrument which measures the cognitive ergonomic quality of an IVIS and its consequences on traffic safety. We proceeded in three steps:

**Step 1: Item specification**

To start with, a pool of items was generated which capture several crucial aspects of cognitive ergonomics such as learnability, error tolerance or controllability. All items were formulated as declarative sentences that can be rated on a five-point Likert scale. For 20% of the items, the polarity was reversed thus requiring a negation of the presented statement. The sequence of items was randomized and each item was evaluated by experts to ensure their comprehensibility. Based on these reviews, items were optimized and used to form a first version of the questionnaire.

**Step 2: Item selection**

The first version of the questionnaire was used in a pilot study in which four IVIS were evaluated by another group of experts. The IVIS were a portable navigation system, a smart phone including a navigation application, a radio navigation system and an interface-prototype which was intentionally bad designed. The systems were evaluated in the context of a driving experiment. After analyzing the psychometric characteristics of the items, only those with inter-item and item-scale correlations (which are indices for homogenity and discriminatory power) exceeding 0.4 and 0.5, respectively, were preserved.

The reduced questionnaire was applied in a study in which over 100 experts and students in the field of automotive engineering assessed the interface quality of seven portable navigation devices. In this phase, the systems were not used while driving, but participants first had to accomplish two reference tasks and then evaluate the systems in a single task scenario only. Based on the data of this experiment, another 63 items were excluded and 77 items were kept.

**Step 3: mutli-level scale development**

To identity the underlying dimensions of these items, we conducted a card-sorting experiment in which participants first sorted the items into groups and then named the groups. In the data analysis, a frequency distribution over groups was calculated for each item and its group membership was determined, i.e., it was allocated to the group in which its frequency was highest. Items with the same frequency in two groups were included in both. As a result of the card sorting experiment, 10 groups (dimensions) were identified: intuitive use, operability, clarity, coloring, degree of reliance, lettering, design aspects, feedback, learnability, and others (containing ambiguous items). Inter-item and item-scale correlations were then calculated for each group of items (dimension) separately. Those items with unsatisfying metrics in respect to inter-item and item-scale correlation were excluded from the item list. This resulted in a final questionnaire consisting of 35 items.

To complement the results of the card sorting experiment, we conducted a factor-analysis with these remaining items, based on the same data as in step 2. We detected six factors, explaining a total variance of 61,9 %. The test of data fit was not significant ($\chi 2 = 417{,}2$; $df = 400$; $p = 0.3$), i.e., the null hypotheses that the factorial model does not differ from the data could be kept. Since only 3 items loaded on factors 4, 5 and 6, the analysis was repeated, fixing the number of factors to be extracted to three.

Next, the three factors were interpreted according to the items with the highest factor loading ("leading variables"). With respect to factor 1, these were items such as "I always knew, which input I had to enter." and "I could use the system without thinking about it.". Therefore, this factor (or dimension) was named "intuitive use". The items loading on the second factor were related to the menue structure. Examples are "I have not achieved what I had intended." or "The system requested unnecessary inputs.". This dimension was termed "user guidance". The third factor concerned design aspects and was labeled accordingly. A reliability analysis with items loading on each factor revealed high internal consistencies of the identified dimensions: Chronbachs Alpha for the dimensions is 0.95 (intuitive use), 0.90 (user guidance), and 0.92 (design aspects), respectively.

# 4 VALIDATION OF THE RESULTS OF COGNITIVE ERGONOMICS ASSESSMENT

## *4.1 Method*

In order to compare the results of the cognitive ergonomic evaluation to results of system usage while driving, we conducted a simulator experiment. For the purpose of the study, our driving simulator (provided by the Volkswagen AG) was used. It is a VW Bora with the complete interior equipment. The goal was to generate an experimental environment in which participants feel the need to drive paying close attention to the happenings outside the vehicle while operating a navigation system. Besides the in-vehicle navigation system, type RNS MFD by Blaupunkt, which was placed in the center console, the vehicle was equipped with three other navigation systems. One system was a mobile device (smart phone) equipped with the TomTom navigation software. Another system was the Becker Traffic Assist 7977 and the third system was an intentionally bad designed prototype which violated all aspects of an ergonomic interface. These systems were mounted above the dashboard, below the rear view mirror, in the middle of the windshield. This position was chosen because it is a common place to mount mobile navigation systems.

55 participants between the ages of 20 and 71 years old (with a mean age of 32.5 years) were recruited for the study. 30 out of the 55 subjects were male and the rest were female. 19 subjects traveled less than 10,000 kilometers per year, 33 subjects traveled between 10,000 and 30,000 kilometers per year, and three subjects traveled between 30,000 and 50,000 kilometers per year. Twelve subjects did not own a navigation system, and ten out of the twelve usually don't use a navigation system at all. The other subjects owned (in some cases even two) navigation systems. The average usage of a navigation system ranged from daily to once or twice a year (13 participants used their navigation system once or twice a week, 13 subjects used their system once or twice a month). Among the subjects, the most used navigation system was TomTom. Three participants had to terminate their participation early due to nausea (simulator sickness) and were excluded from data analysis.

The driving task took place on a rural road environment. Participants were asked to follow a leading car. It traveled with irregular speed, with a distance of 50 m. Oncoming traffic and unexpected obstacles such as bicycle drivers and pedestrians were included in the scenery. Additionally, weather conditions were manipulated by adding fog to the scenery. Participants were requested to accomplish two secondary tasks with the IVIS while driving by a position-triggered cue: Changing the language setting from German to English, and entering a destination (Berlin Hardenbergstrasse 10). Additionally, participants completed a baseline drive without secondary task of about three minutes duration. As dependent measures we recorded driving performance in respect to the duration and frequency of driving errors (for details regarding driving error as dependent measures for IVIS evaluation refer to [15]), gaze data, task duration, subjective ratings of mental workload and a subjective risk estimation.

## *4.2 Data Analysis and results*

Regarding the driving performance, the standard deviation of lane position and the frequency of leaving the right lane was highest for the bad designed prototype. For 70% of all experimental runs, participants were not able to enter the address correctly with this system. The distribution of gaze allocation to the IVIS was higher for the prototype system than for the mobile navigation system (due to data recording problems the other two systems were excluded from this analysis). Mean task durations revealed similar results: They were highest for the prototype system (2 min 35 s), followed by RNS MFD (2 min 21 sec), the mobile navigation system (1 min 23 s) and the smart phone (1 min 13 s). This was also reflected in the subjective measures: the rated mental workload and the estimated criticality of operating the system while driving were lowest for the mobile navigation device, followed by the RNS MFD and the dummy system. Interestingly, they were highest for the smart phone, probably because participants had to hold and operate the system with one hand and the thumb. These results were compared to the assessment of cognitive ergonomic criteria by using our 35 item questionnaire. Hence, the parameter values for the three dimensions were calculated as a mean value of all items corresponding to the certain dimension. These dimension values were analyzed by means of an analysis of variance. We found significant differences an all dimensions between the systems (intuitive use: $F(6,85) = 14,5$, $p < .001$; user guidance: $F(6,91) = 21,9$, $p < .001$; design aspects: $F(6,88) = 18,5$, $p < .001$). Especially, the evaluation of the intentionally bad designed prototype is relevant in this context. It could serve as a negative reference for a certification procedure based on cognitive ergonomic criteria. Therefore, it was tested if the results for the prototype differ significantly from the other systems by using the statistical technique of contrast analysis. The expected differences could be proven for all three dimensions (intuitive use: $t85 = -7,6$, $p < .001$; user guidance $t91 = 9,7$, $p < 0.001$ and design aspects: $t88 = -10,3$, $p < .001$).

# 5 DISCUSSION

Regarding the analysis of the 77 item questionnaire version, we went back to a qualitative procedure (namely card sorting) before verifying the data structure with a quantitative technique. The most important aspect regarding the card sorting experiment was the exploration of relevant evaluation categories solely by experts in the field of human factors and IVIS engineering in an early stage of questionnaire development. Additionally, participants provided proficient labelings for the several dimensions. Due to this procedure, we could ensure not to oversee relevant subdimensions of the construct “cognitive ergonomics”.

Another relevant aspect is concerned with the concordance of the qualitative and quantitative data analysis procedures. It can be shown that the defined dimensions “intuitive use”, “degree of reliance” and “learnability”, which were identified in the card using experiment, are reflected for the majority of items in the first factor “intuitive use”. Similar conformities could be found for the qualitative aspects “learnability” and “operability”, which correspond to a great extend with the second factor “user guidance”. Items of the card sorting dimensions “lettering”, “design aspects”, “coloring” and in parts “operability” can be found in the third factor “design aspects”.

Regarding the participants of the data acquisition with the first revision of the questionnaire after the pilot study, it has to be stated that the population we drew our random sample from, was not absolutely identical with the intended target group of the final questionnaire. The latter consists of human factors and engineering personnel of original equipment manufacturers and component suppliers, who design and evaluate the human-machine interfaces of in-vehicle information systems. As a concession to severe time restrictions of such engineering professionals, we concentrated on young academics studying major subjects such as human-factors, psychology, engineering sciences and cognitive ergonomics. A second obstruction was the qualification and the training of the investigators. Data collection was administered during an university project; according to these circumstances some influences on data quality cannot be eliminated entirely.

Finally, the development of the questionnaire has been undertaken with German items only; up to now, no English version has been tested. After the final questionnaire is translated from German to English, additional estimations of reliability and validity have to be carried out. The questionnaire for estimating cognitive ergonomic criteria of in-vehicle information systems is part of an evaluation toolbox which also focuses on normative rules regarding colors and symbols as stated in the DIN EN ISO 15008 [16], and task duration based on a simplified KLM-GOMS analysis as proposed by the society of automotive engineers in the guideline J2365 [17].

# 6 ACKNOWLEDGMENTS

I thank Dr. Peter Rößger and Dr. Miklos Kiss for their support preparing this study and giving advice for the experimental design and interpretation of the results. The project CarUSE is conducted by the Chair of Human-Machine Systems of the Technische Universität Berlin in cooperation with Harman/Becker Automotive Systems GmbH and Volkswagen AG, and is sponsored by them.

# 7 REFERENCES

[1] Green, P. (1999). Estimating Compliance with the 15-Second Rule for Driver-Interface Usability and Safety. Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting. Santa Monica, CA. (CD-ROM).

[2] Green, P. (2000). Potential expansion of the 15-second rule. University of Michigan Transportion Research Institute, Human Factors Division. http://www-nrd.nhtsa.dot.gov/departments/nrd-13/driver-distraction/PDF/Green.PDF.

[3] AAM (2002). Statement of principles, criteria and verification procedures on driver interaction with advanced in-vehicle information and communication system. Washington: Alliance of Automobile Manufacturers.

[4] Society of Automotive Engineers (2002). SAE Recommended practice calculation of the time to complete in-vehicle navigation and route guidance tasks. SAE J2364. Warrendale: Society of Automotive Engineers.

[5] Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. Human Factors, 38(3), 442-451.

[6] Horrey, W., & Wickens, C. (2007). In-vehicle glance duration - Distributions, tails, and model of crash risk. Transportation Research Record (2018), 22-28.

[7] Horrey, W., Wickens, C., & Consalus, K. (2006). Modeling drivers' visual

attention allocation while interacting with in-vehicle technologies. Journal of Experimental Psychology - applied, 12(2), 67-78.

[8] Mattes, S. (2003). The Lane Change Test as a Tool for Driver Distraction Evaluation. In: H. Strasser, K. Kluth, H. Rausch & H. Bubb (Eds.), Quality of Work and Products in Enterprises of the Future (pp. 57-60). Stuttgart: Ergonomia Verlag.

[9] ISO/DIS 26022 (2007). Road vehicles—Ergonomic aspects of transport information and control systems — Simulated lane change test to assess driver distraction. Draft International Standard. Geneva: ISO.

[10] Vöhringer-Kuhnt, T. (2009). CarUSE – Developing a framework for IVIS evaluation. In: D. de Waard, J. Godthelp, F.L. Kooi, and K.A. Brookhuis (Eds.), Human Factors, Security and Safety (pp. 251-260). Maastricht, the Netherlands: Shaker Publishing.

[11] Stevens, A., Board, A., Allan, P., & Quimby, A. (1999). A Safety Checklist for the Assessment of In-Vehicle Information Systems. Crowthorne Berks, UK: TRL Library.

[12] Schmitz, M. (2002). Verfahren zur Bewertung der kognitiven Ergonomie von Bedien- und Anzeigekonzepten auf der Basis von CAR-E. Masterthesis, Westfälische-Wilhelms-Universität, Münster.

[13] Pataki, K., Schulze-Kissing, D., Mahlke, S., & Thüring, M. (2005). Anwendung von Usability-Maßen zur Nutzeneinschätzung von Fahrerassistenzsystemen. In K. Karrer, B. Gauss, B. & C. Steffens (Eds.), Beiträge zur Mensch-Maschine-Systemtechnik aus Forschung und Praxis (pp. 211 - 228). Düsseldorf: Symposion.

[14] Hacker, W. (1998). Allgemeine Arbeitpsychologie. Psychische Regulation von Ar-beitstätigkeiten. Bern: Huber.

[15] Glaser, R. & Waschulewski, H. (2005). INVENT – Forschungsprojekt Fahrerassistenzsysteme (FAS) Teilprojekt Fahrerverhalten und Mensch-Maschine-Interaktion (FVM) Arbeitspaket 3200 Validierung und Weiterentwicklung des Bewertungsverfahrens I-TSA (INVENT – Traffic Safety Assessment). Abschlussbericht. Tübingen: Universität Tübingen, Psychologisches Institut und MTO Psychologische Forschung und Beratung GmbH. Retrieved from http://homepages.uni-tuebingen.de/wilhelm.glaser/BeriExp08.pdf, accessed March 2009.

[16] DIN EN ISO 15008 (2009). Straßenfahrzeuge - Ergonomische Aspekte von Fahrerinformations- und Assistenzsystemen - Anforderungen und Bewertungsmethoden der visuellen Informationsdarstellung im Fahrzeug. Berlin: Beuth-Verlag.

[17] Society of Automotive Engineers (2002). Calculation of the Time to Complete In-Vehicle Navigation and Route Guidance Tasks. SAE J2365 . Warrendale: Society of Automotive Engineers.