Theater-system Technique and Model-based Attention Prediction for the Early Automotive HMI Design Evaluation

Sebastian Feuerstack
OFFIS – Institute for Information Technology
Oldenburg
Germany
feuerstack@offis.de

Bertram Wortelen
C.v.O. University
Oldenburg
Germany
bertram.wortelen1@uni-oldenburg.de

Carmen Kettwich
German Aerospace Center e.V.
Braunschweig
Germany
Carmen.Kettwich@dlr.de

Anna Schieben
German Aerospace Center e.V.
Braunschweig
Germany
Anna.Schieben@dlr.de

ABSTRACT
Automotive HMI design evaluation methods, such as usability assessments, attention and reaction time measurements require full working HMI prototypes to assess the usability and the subjects' performances in realistic situations. The theater-system technique and model-based prediction methods do not depend on functional HMI implementations and therefore promise HMI evaluation already in an early design phase.

We applied both methods to evaluate three HMI designs for an Urban Adaptive Cruise Control (ACC) System. In a qualitative study with twelve participants, we used a theater-system-based technique to let them experience the HMIs in realistic situations. Subjects clearly preferred the HMI variant, which offers the best understanding of the vehicle's automation. By following a model-based approach, we evaluated the impacts to the driver's visual attention distribution of the three HMI variants with six human factor experts and found significant attention changes for the front window and for the Urban ACC HMI.

ACM Classification Keywords
H.5.2. Information interfaces and presentation: Evaluation/methodology, prototyping; H.1.2. User/machine systems: Human factors; I.2.0. Artificial Intelligence: Cognitive simulation.

Author Keywords
Model-based attention prediction; Usability assessment.

INTRODUCTION
With the rising complexity of technical systems decades ago people realized, that the increased burden on the human operator to monitor and control the system calls out for optimized human-machine interfaces (HMI). Several methods for developing HMI in a user-centred way and assessing the quality of HMI were developed over the years. With this contribution we aim at highlighting two different techniques that are specifically targeted to the very early HMI design evaluation: The theater-system technique and model-based prediction. Both methods promise to gain qualitative and quantitative insights to HMI designs. Thus, they aim at reducing the chance of ending up with HMIs that are difficult to operate, require unnecessary monitoring efforts or introduce unwanted driver distractions.

The theater-system technique [9] works with two coupled, fixed-based driving simulators to allow the emulation of future driving assistance systems and automation. This setup enables users to experience HMI design variants for automated vehicle applications already before they are actually implemented. Model-based prediction methods using cognitive architectures [1, 7] can simulate human behavior based on psychological and physiological plausible models. They have been applied successfully to predict e.g. car drivers' attention distributions [3, 11]. Similar to eye-tracking studies that are often used to measure drivers' distraction, model-based prediction methods also end up with metrics such as estimated reaction times and percentage dwell times. But they also offer insights about causes for predicted changes of drivers' attention.

In the following we present the application of both design evaluation methods for three alternatives of an Urban Adaptive Cruise Control (ACC) HMI. We start by introducing the theatre-system technique and the model-based attention prediction for HMI evaluation followed by briefly introducing the exemplary Urban ACC HMI design variants that we evaluated. Finally, we present evaluation results that were gained by these methods and summarize the results and close by stating future work.

EARLY HMI EVALUATION METHODS
The theater-system technique is a method for agile designing and testing of system behavior and interaction concepts in early design phases. The technique allows doing a rapid prototyping of system behavior and multimodal interaction long before the complex software for such a prototype is build up. The theater-system technique is based on the Wiz-

Automotive'UI 16, October 24 - 26, 2016, Ann Arbor, MI, USA

ACM 978-1-4503-4533-0/16/10...$15.00
DOI: http://dx.doi.org/10.1145/3003715.3005466

ard-of-Oz technique (WoOz), where a human "wizard" hidden behind a curtain is emulating the functionality of a machine [4]. Originally, the technique was used for automatic speech or gesture recognition and picked up in other domains, e.g. for testing driver assistance systems in real vehicles [8]. The theater-system technique extends the WoOz technique in a way that there is no longer a hidden wizard but that the curtain between the user and a member of the design team (confederate) can also be open, and both user and confederate can play through different use cases as if they would play a role in a theater. Whereas the WoOz technique is primarily used for natural language processing, non-verbal behavior simulation [6], navigation and mobility tasks the theater-system can be used for user-related expectation assessment, early evaluation and design.

The SEEV (Saliency, Effort, Expectancy, and Value) model of human attention allocation proposed by Wickens et al. [10] predicts human attention prediction based on four influencing factors. It has been applied in several studies, e.g. with car drivers [3, 11]. It was recognized, that the task-related, top-down factors *Expectancy* and *Value* are the main drivers of attention [10]. Determining the SEEV model's expectancy and value parameters is difficult and requires expertise in cognitive modelling. Recent research proposed tools like the HEE [2] to ease the creation of SEEV models based on top down factors. The HEE guides Human Factor (HF) experts through a process to systematically derive the parameters and then automatically constructs a model of a human monitoring a system. This human model is then instantiated in a cognitive architecture that simulates the human monitoring a system in a psychological and physiological plausible way and ends up with predictions of attention distributions and reaction times over a set of monitored areas of interests. Model-based predictions end up with similar metrics as eye tracking, but also offer to explore the causes for a predicted behavior and take less time, because no eye tracking hardware is required and thud participants can be tested in parallel.

APPLICATION: HMI FOR AN URBAN ACC SYSTEM

Urban ACC in combination with vehicle-to-infrastructure (V2I) communication opens up new ways to increase traffic safety and traffic flow efficiency. With the help of V2I communication the vehicle equipped with ACC is now able to adapt its maneuvers with regard to the traffic light signal status and its future phase change. During ACC drive the driver should be able to understand the chosen maneuvers of the ACC even so the vehicle reacts to a forecast which might be in discrepancy with the current environmental situation (e.g. traffic light "green", vehicle is decelerating). The focus of this paper is on the evaluation of three different visual HMI variants for the Urban ACC application (see Figure 1).

The first HMI concept (HMI 1) has the lowest information content. This design concept only gives the driver information about the success/failure of the communication between the vehicle and the corresponding traffic light. This is illustrated through a traffic light icon with radio waves. The other two design concepts have higher information content but different approaches to support the driver in comprehending and monitoring the current automation maneuver of the ACC. The second design concept (HMI 2) depicts the actual traffic light signal status. Additionally a countdown of the remaining time of the ongoing status of the traffic light signal is shown. The third design concept (HMI 3) shows the status of the traffic light signal, when passing the corresponding intersection.

EVALUATION RESULTS

HMI evaluation with the theater-system technique

The three different HMI variants for the Urban ACC were tested in a usability assessment [5] with the theatre-system technique. The ACC functionality was emulated by a human confederate behind a curtain. Twelve participants (six male and six female) took part in the usability study. The participants were between 23 and 49 years old. They tested the Urban ACC with the three described HMI variants in an urban environment approaching three times at an intersection (3x3 test runs). The status of the traffic light (TL) signals varied as follows: a) Approaching a red (TL), arriving at a red TL; b) Approaching a red TL, shortly before arriving the TL switched to green; c) Approaching a green TL, arriving at a green TL. Participants were asked to evaluate the HMI variants by filling in a questionnaire (van der Laan scale) and were asked for comments and suggestions to

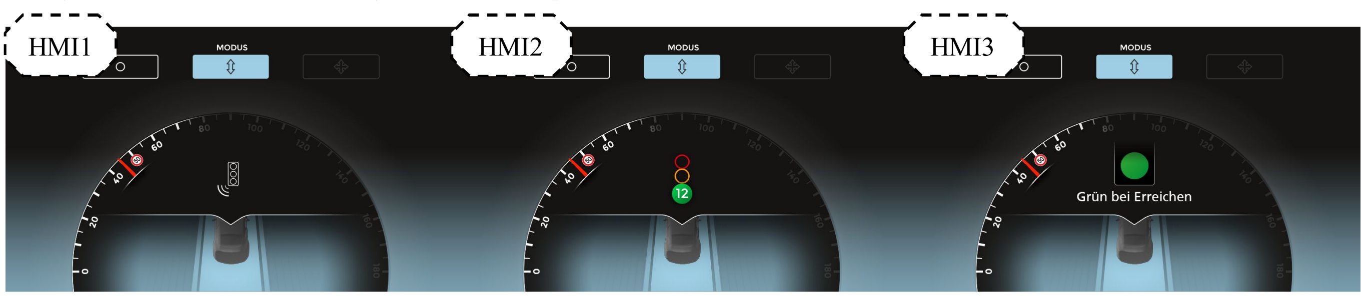

Figure 1. Display structures of the HMI; HMI1: Pictogram of a traffic light with radio waves; HMI2: Current traffic light signal status with countdown and HMI3: Traffic light signal status when arriving and passing the corresponding intersection

improve the existing design concepts in form of a semi-structured interview.

The results of the analysis of the satisfying and usefulness scale are shown in Figure 2. For each design concept the ratings of the nine semantic differentials were averaged over participants. The third HMI concept was assessed very positive in terms of satisfaction and usefulness. Participants consistently mentioned the large, well readable, clear and understandable design of the status of the traffic light, while passing the intersection. In addition the textual support beneath the traffic light icon was rated positively. Compared to the third HMI design concept the second HMI design concept was rated as less satisfying and useful. The clear and intuitive design as well as the color coding of the second HMI concept design was positive evaluated, whereas seven participants got distracted by the countdown of the current status of the traffic light. The first design concept received the lowest score on evaluated satisfaction and usefulness. The lack of information on the status of the traffic light, the absence of color coding of the traffic light icon as well as the lack of information about the performance of the ACC shortly before arriving at the traffic light was criticized by the participants [5].

HMI evaluation with Model-based Attention Prediction

For the model-based analysis of attention allocation we selected a specific situation that is relevant for Urban ACC systems: the car is approaching a traffic light, while the driver cannot see the traffic light, but a traffic sign announces the upcoming traffic light (see Figure 3). We used the Human Efficiency Evaluator (HEE) to create models of attention allocation for all three interface designs in the depicted situation. Six human factors experts used the HEE independently for all three HMI designs to identify the areas of interests (AOIs) (see Figure 3) and the *Expectancy* by ordering the AOIs based on how often they expect new information from them.

Four tasks that represent typical urban traffic driver tasks have been predefined for the study: “Observe Road Ahead”, “Observe Rear Traffic”, “Control Lateral Position”, and “Control Speed”. Due to the use of the Urban ACC “Control Speed” is mainly the supervision of the ACC function-

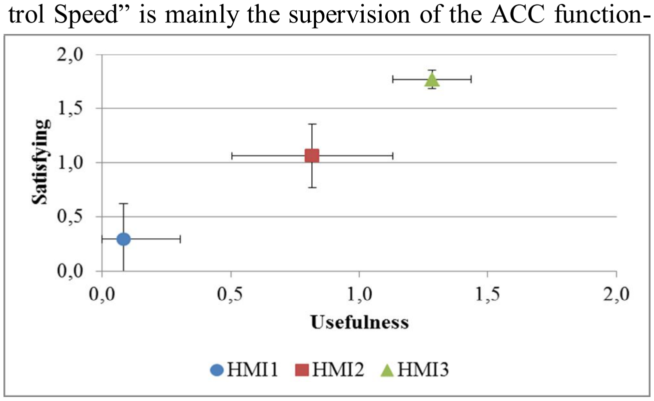

Figure 2. System acceptance assessed with help of usefulness and satisfying scale

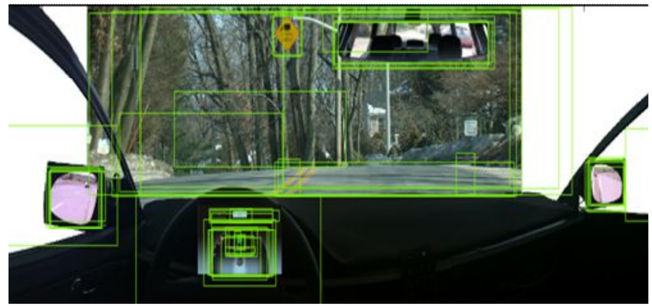

Figure 3. Sign that announces upcoming traffic light. Green boxes show areas of interest identified by HF experts.

ality. HF experts were required to order the tasks by importance via the HEE. Finally, by filling a relevance matrix, each AOI in each HMI variant scene was marked either as “required”, “helpful”, or “not relevant” for each of the four driving tasks. Based on this data the HEE uses the lowest ordinal heuristic proposed by [10] to calculate the SEEV model’s top-down parameters and simulates a driver’s monitoring behavior. The result is a prediction of the drivers’ attention distributions for each HMI variant.

Since the AOIs differed for each subject (c.f. Figure 3), we manually classified them with three raters into 15 classes with a high inter rater reliability (Fleiss’ $\kappa = 0.88$). The six subjects’ ratings for the tasks importance and expectancy rankings and also for the relevance matrix were highly concordant (Kendall's coefficient of concordance $W_t > 0.82$; $p < 0.01$) for the three AOIs that all subjects have been identified: The forward view, the Urban ACC, and the speedometer. Different to eye tracking that ends up in the same metric as our approach (percentage dwell time - PDT), a model-based approach allows to inspect the simulation models to learn about causes for the observed shifts in attention between the HMI variants. The described evaluation is a repeated measure design for which we found a significant difference of predicted attention distribution between the HMI variants for the Urban ACC display ($F(2,10) = 8.041$, $p = 0.00828$) (see left graph of figure 4). Subsequent t-tests indicated, that the mean PDT for HMI 1 ($M=0.069$,

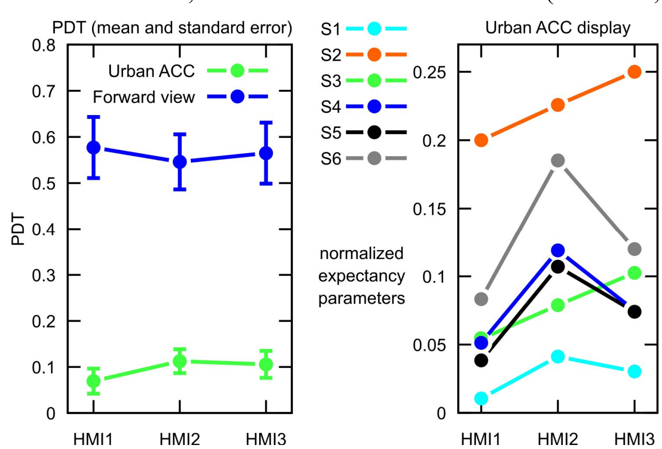

Figure 4. Left: percentage dwell time. Right: normalized expectancy parameters of Urban ACC display for all subjects (S1-S6)

σ=0.067) was significantly different to HMI 3 (M=0.106, σ=0.072, p=0.002) and also to HMI 2 (M=0.113, σ=0.063), but for HMI 2 only with marginal evidence (p=0.052). Attention is a limited resource. The data indicate that the increased amount of attention to the Urban ACC in HMI 2 and HMI 3 is mostly drawn from the forward view (F(2,6)=7.054, p=0.027).

Digging into the models revealed that the cause of the effect does neither originate from the task importance parameters nor the relevance matrix. It is caused by the subjects expecting a difference in the amount of expected information from the Urban ACC display. This is indicated by the expectancy that was calculated with the lowest ordinal heuristic. They are shown in the right graph of Figure 4. The figure does not only show, that the subjects expect a difference in amount of expected information, but also that their opinions differ. All subjects expect the least information from HMI 1, but subjects S2 and S3 expect the most information from HMI 3, while all other subjects expect it from HMI 2. Another indicator for the difference in opinions is that the standard deviation of expectancy parameters for the Urban ACC display (HMI 1=0.067, HMI 2=0.068, HMI 3=0.076) is by far higher than for any other AOI. Searching for such patterns can be extremely helpful information for the HMI designer, because the designer can see, whether his expectation about the HMI design matches with the expectations of other. This can help to early identify unexpected effects of the system.

CONCLUSION AND OUTLOOK

Different to other evaluation approaches, the theater-system technique and also model-based attention prediction can give meaningful insights about usability and the impact of an HMI design to car driver's attention distribution already at an early design phase. Whereas the former enables car drivers to experience HMI interfaces designs before they are actually being implemented, the latter benefits from the imagination and knowledge of human factor experts to model and discover design impacts on human's attention.

Combining both methods, e.g. by experiencing the HMI variants with the theater-system technique before the attention distribution is being modelled with the tool, promises comprehensive insights into a HMI design's impact on the driver's attention distribution and will be future work that we aim to validate with accompanying eye tracking studies.

REFERENCES

[1] Thierry Bellet, Pierre Mayenobe, Jean-Charles Bornard, Dominique Gruyer, and Bernard Claverie. 2012. A computational model of the car driver interfaced with a simulation platform for future virtual human centred design applications: COSMO-SIVIC. *Eng App Artif Intel*, 25, 7:1488–1504.

[2] Sebastian Feuerstack and Bertram Wortelen. 2016. AM-DCT: A visual attention modeling data capturing tool for investigating users' interface monitoring behavior. In *Proceedings of the working conference on Advanced visual interfaces (AVI 2016)*, Bari, Italy, June 7th-10th 2016.

[3] William J Horrey, Christopher D Wickens, and Kyle P Consalus. 2006. Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *J. of Exp. Psychol.-Appl.*, 12, 2:67.

[4] J. F. Kelley. 1984. An iterative design methodology for user-friendly natural-language office information applications. *ACM T. Off. Inf. Sys.*, 2:26–41.

[5] Carmen Kettwich, Raphael Haus, Gerald Temme, and Anna Schieben. 2016. Validation of a HMI concept indicating the status of the traffic light signal in the context of automated driving in urban environment. In *IEEE, Intelligent Vehicles Symposium (IV '16)*, Gothenburg, Sweden, 1374-1379.

[6] Laurel Riek.2012. Wizard of Oz Studies in HRI: A systematic Review and new Reporting Guidelines. *Journal of Human-Robot Interaction*, 1:119–136.

[7] Dario D. Salvucci. 2007. Integrated models of driver behavior. In *Integrated Models of Cognitive Systems*, Wayne D. Gray, (ed). Oxford University Press, 356-367.

[8] Gerrit Schmidt, Miklós Kiss, and Eckhard Babbel. 2008. Das Wizard of Oz Fahrzeug: ein Werkzeug für Rapid Prototyping und Usability Testing von zukünftigen Fahrerassistenzsystemen. In *Proceedings of 3. Tagung Aktive Sicherheit durch Fahrerassistenz*, Garching, Germany, April 7-8 2008.

[9] Anna Schieben, Matthias Heesen, Julian Schindler, Johann Kelsch, and Frank Flemisch. 2009. The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles. In *Proceedings of 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 43-46.

[10] Christopher D. Wickens, John Helleberg, Juliana Goh, Xidong Xu, and William J. Horrey. 2001. Pilot task management: Testing an attentional expected value model of visual scanning. Technical Report ARL-01-14/NASA-01-7, University of Illinois, Aviation Research Lab, Savoy, IL.

[11] Bertram Wortelen, Martin Baumann, Andreas Lüdtke. 2013. Dynamic simulation and prediction of drivers' attention distribution. *Transport. Res. F-Traf*, 21:278–294.